An international round-robin study for the analysis of particulate semi-volatile organics by thermal desorption gas chromatography mass spectrometry

Gianni Caravaggio a*, Chris Hebbern b, Luyi Ding c and Sabit Cakmak b

aNatural Resources Canada, CanmetENERGY, Ottawa, ON K1A 1M1, Canada; bHealth Canada, Population Studies Division, Environmental Health Science & Research Bureau, Ottawa, ON K1A 0K9, Canada; cEnvironment Canada, Air Quality Research Division, Ottawa, ON K1A 0H3, Canada

(Received 30 September 2014; accepted 4 May 2015)

Thermal desorption gas chromatography mass spectrometry (TD-GC/MS) is becoming more commonly used for the quantification and identification of organic compounds in particulate matter (PM), including ambient and source PM such as diesel particulate matter (DPM). It has been proven as an alternative to the traditional solvent extraction (SE) method and liquid injection gas chromatograph mass spectrometry (LI-GC/MS). However, little information is available on how different types of TD-GC/MS systems compare to each other for analysis of real-world PM samples or to direct LI-GC/MS for analysis of PM components in a test solution. To address this, CanmetENERGY Characterization Laboratory initiated a round robin with the participation of 10 laboratories worldwide. Three sample types were analysed: (i) a test solution with a suite of pure compounds commonly found in PM, analysed by TD-GC/MS and LI-GC/MS; (ii) a DPM sample, analysed by TD-GC/MS and SE; and (iii) an ambient PM sample, analysed by TD-GC/MS. The first part of the study showed good overall performance and comparability between the different TD-GC/MS systems and LI-GC/MS method for the analysis of PM components in a test solution, with some variability of results due to system types and parameters used, concentration of calibration standards, and whether or not an internal standards was used. The analysis of the DPM sample showed greater variability between laboratories and methods as many PM components were present near the detection limit and matrix effects particularly affected the TD-GC/MS analysis of heavier \( n \)-alkanes. In the last part of the study, for the analysis of an ambient PM sample by TD-GC/MS, the analysis of variance showed good comparison between labs for polycyclic aromatic hydrocarbons (94% non-significant), but slightly lower for \( n \)-alkanes (68%) and biomarkers (57%).

Keywords: thermal desorption; round robin; gas chromatography mass spectrometry; TD-GC/MS; ambient particulate matter; solvent extraction

1. Introduction

It has been recognised that elevated particulate matter (PM) concentrations in ambient air produce adverse health effects [1–3]. Identification of the various PM sources and assessment of their chemical composition are important steps in the management of air quality [4]. This is particularly important in the case of DPM as it contains known carcinogenic polyaromatic hydrocarbons [5]. The conventional approach to analyse PM chemicals typically involves using organic solvents to extract the particulate phase from sample filter media followed by concentration, clean-up, fractionation and quantification by gas chromatography equipped with a mass spectrometer (GC-MS). There are various solvent extraction (SE) techniques that are used such
as Sohxlet [6], sonication-assisted extraction [7], pressurised SE [8] and automated SE [9]. These extraction techniques use hazardous solvents, are relatively complex, time consuming and are at risk for the introduction of significant uncertainties in the final results as a result of the multiple sample preparation steps involved [10].

Thermal desorption gas chromatography mass spectrometry (TD-GC/MS) is becoming a commonly used alternative method to SE for the analysis of organic chemicals in PM [10–17]. PM samples collected onto solid sorbents, such as quartz filters, are desorbed thermally, focused in a cold trap, heated and transferred to a GC analytical column. The higher sensitivity of the technique compared to SE allows it to be used for much smaller sample quantities [18,19]. Chow et al., in their review paper, summarised recent advances and applications of TD-GC/MS to the determination of aerosol chemical compositions with the comments on the method limitations and the suggestions of future research needs. They also have provided readers with insight into the thermochemical properties of carbonaceous aerosols as to how thermal stability of organic compounds was related to their chemical structures [20].

Thermal desorption with time-of-flight mass spectrometry has been used for daily monitoring of semi-volatile aromatic and aliphatic hydrocarbons [21], while another study utilised TD-GC/MS for the analysis of volatile organic compounds [22]. Organic speciation of ambient PM$_{2.5}$ at the Golden airshed in British Columbia, Canada, has been studied using TD-GC/MS to identify target molecular markers such as polycyclic aromatic hydrocarbons (PAHs) and petroleum biomarkers [23]. High-recovery efficiencies of PAHs using TD-GC/MS were found by Bates et al. [24], while in another study, TD-GC/MS was found to have good recovery of analytes of interest in diesel source emissions, except for 5-ring and larger PAHs [25]. For the analysis of the National Institute of Standards and Technology (NIST) Standard Reference Material (SRM 1649a), urban dust, TD-GC/MS was found to have high sensitivity for PAHs (2–6 mg kg$^{-1}$) and good reproducibility over sample weights typically collected on short time scales [19]. Thermal desorption coupled with comprehensive two-dimensional gas chromatography and tandem mass spectrometry (TD-GC/GC-MS/MS) has resulted in dramatically improved sensitivity of the trace particulate samples for the determination of PAHs and their derivatives (oxygenated, nitrated and methylated PAHs) [26].

Ho et al. [14] evaluated the performance of in-port TD-GC/MS for the analysis of a large suite (132) of non-polar PM organic compounds. They showed that the in-port TD-GC/MS was highly reproducible achieving relative standard deviations (RSDs) <10% for the majority of non-polar compounds, from calibration standards and ambient samples. Furthermore, they determined the accuracy of their system for 15 PAHs to be within ±5% of the certified value of a NIST 1649 urban dust standard reference material. They also demonstrated that their in-port TD-GC/MS system was comparable to the SE-GC/MS method by comparing the results of 106 non-polar compounds analysed from 14 ambient samples. In a recent small-scale round robin [18], compared directly against SE, three TD-GC/MS were found be comparable for the analysis of a mixture of target analytes in a standard solution and for a diesel particulate matter (DPM) quartz filter sample. Detection limits were equal to or exceeded those obtained with SE.

There are many international laboratories that currently perform chemical analysis of PM samples using TD/GC/MS. These laboratories use TD-GC/MS systems with different configurations and parameters including different columns, GC/MS, calibration and internal standards (IS). However, there is currently little information on how the results of the different laboratories compare with each other for analysis of real-world ambient and source PM samples. The present study, on a much larger scale, compares results obtained from 10 international laboratories using their standard analytical TD-GC/MS methods as well as with liquid injection-GC/MS (LI-GC/MS) in a three-phase approach. The international community could benefit from the information of this study to ensure their data are comparable and for those that are pursuing to standardise
their systems and methodologies across various air networks jurisdictions and for the potential development of international test standards.

2. Experimental

2.1. Study design

Ten international laboratories were selected for the round robin that used TD-GC/MS on a regular basis for the analysis of PM chemicals. Five laboratories were from the US, three were from Canada, one from Italy and one from China. At the outset, labs were selected for their capacity to analyse PM. It should be noted that 7 out of 10 systems were from Gerstel; however, the testing protocol was not designed to evaluate TD-GC/MS systems from different manufacturers. Furthermore, the initiating laboratory did not provide calibration standards to each participant. These factors produced a highly randomised study that provided the opportunity to compare TD-GC/MS data from each laboratory in real-world applications by using their own calibration and IS.

The study was separated into three phases. In phase I, a test solution composed of a suite of pure compounds commonly found in PM was analysed using two methods: TD-GC/MS and LI-GC/MS. This stage of the analysis had two goals: to compare the results of the two methods and to provide information on the ability of various laboratories using a TD-GC/MS system to quantify PM chemicals in a solution without matrix interference. The components and concentrations of compounds in this test solution were unknown to all participating laboratories.

In phase II, a diesel PM sample was analysed by TD-GC/MS and SE methods. As SE has been shown to be comparable to TD-GC/MS, the SE data were used as a reference to calculate the relative accuracy (RA) for the TD-GC/MS results and to simplify comparisons between the data of participating laboratories. The RA, for each target analyte, was calculated by subtracting the mean of the TD data from the SE (when available) data and dividing by the SE data, expressed as a percentage. The second phase was used to compare TD-GC/MS systems of a real-world source sample with a matrix.

Finally, in phase III, an ambient PM filter sample was analysed by different TD-GC/MS systems, with the goal of evaluating the consistency of results from different laboratories for a real-world ambient sample.

2.2. Preparation of test solution, DPM sample and ambient PM sample

2.2.1. Test solution preparation

A 50 mL stock test solution was prepared in isooctane with 16 Environmental Protection Agency (EPA) target PAHs (Ultra Scientific PM611 mixture in methylene chloride), 16 n-alkanes (n-alkane, Ultra Scientific SFL-601 mixture in hexane), and 12 petroleum biomarkers (Chiron, S-4436-IO, Hopane/Sterane calibration mixture in isooctane) at concentrations between 0.3 and 1 μg/mL (supplementary Table A). The test solution was prepared by a person other than the analyst and analysed by LI-GC/MS at the CanmetENERGY Characterization Laboratory (initiating laboratory; 1 Haanel Dr., Canada, Ottawa, Ontario) to verify the concentrations of the compounds. Four repeat injections of the test solution were performed to confirm the true values. Following this verification step, approximately 2 mL aliquots of the test solution were transferred into 2 mL GC injection vials, sealed with Teflon crimp seals and stored in a freezer until sent to the participants. The composition and concentration of analytes in the test solution was unknown to the participants. This solution was analysed by nine participants by spiking onto blank filters and then run through their TD-GC/MS systems, and analysed by...
LI-GC/MS by three participants. The true value and the concentrations obtained by LI-GC/MS are given in supplementary Table A.

2.2.2. **DPM sample collection**

The DPM samples were collected using a similar procedure as described in Graham et al. [18]. The samples were collected from a 2004 Cummins ISM280 diesel engine operating in its certified configuration which included an oxidation catalyst and the use of commercial ultralow sulphur diesel fuel (<15 ppm sulphur). The engine was operated at a constant speed and load, and the total volume was collected using a constant volume sampling system. Eight different engine runs were performed, and for each run, four samples of the dilute exhaust were collected simultaneously using filters with URG-2000-30ENB cyclones for a total of 32 filters. All test runs were performed on the same day. The individual cyclones were fitted with 90 mm diameter filter packs with one of two configurations. Two of the filter packs contained a primary Teflon membrane filter (Pall Zefluor™ Membrane, 2 μm pore size) and a downstream quartz filter (Pall Tissuquartz™ Filters, 2500 QAT-UP pre-fired at 900° for 3 hours). In this case, the primary Teflon membrane filter was used to determine the dilute exhaust PM concentration by gravimetry. The average mass of the Teflon membrane filters after loading was approximately 1.00 g with a standard deviation of 5%, indicating loading homogeneity (supplementary Table B). The other two filter packs contained only a primary quartz filter. These quartz filters were not subjected to gravimetric analysis for two main reasons: first, quartz filters tend to lose fibres or break during sample collection; and second, it was difficult to achieve stable humidity equilibration with quartz filters and retain low organic carbon (OC) blank levels. As a result, there were no gravimetric data for these quartz filters, but it was assumed that they were loaded to the same degree as the Teflon membrane filters. Samples were collected at flow rates of 60 lpm for 20 min since this was the maximum stable flow that could be achieved with the Zefluor membrane filters. These cyclones obtain a PM$_{2.5}$ cut at 91 lpm; hence, the actual cut at 60 lpm was 3.5 μm. The entire quartz filters were kept in cleaned plastic Petri dishes and stored in a freezer until shipped. One filter was sent to each participant for TD-GC/MS analysis, and a second one was sent to the participants that also performed the SE procedure (supplementary Table B). In every case, samples were sent with the test solution to each laboratory in a polystyrene container containing ice packs. Nine laboratories performed TD-GC/MS analysis of the DPM filter sample. A minimum of three replicates analysis were done for each sample.

2.2.3. **Ambient PM filter sample collection**

The sample of ambient PM (with size less than 2.5 μm aerodynamic diameter, PM$_{2.5}$) was taken at a parking lot near the Ministry of Environment offices at 125 Resources Road in Toronto, Ontario, near Highway 401. The site was bordered by a large park to the east and south and by an area with considerable commercial activity and traffic to the west and north. The sample was collected using a CHEMVoL high-volume cascade impactor (Thermo Scientific BGI 900 lpm) equipped with a 170 mm quartz filter at a flow rate of 1318 lpm (0.022 m$^3$/s), for 72 hours. The impaction stage for removal of particles size larger than 2.5 μm was attached directly to the final juncture of the cascade impactor system. This configuration permitted the collection of a large amount of uniformly deposited fine particles on the quartz filter and allowed for multiple 25 mm diameter punches to be cut and distributed among laboratory participants. The sample filter was also subjected for organic and elemental carbon (EC) analysis to test the homogeneity of particulate deposition on the filter. OC and EC analyses were performed on six 10 mm punches.
taken from random locations on quartz filter sample to assess its homogeneity. The OC and EC were determined using a Sunset Labs Thermal Optical Transmission apparatus (TOT-Sunset Laboratory, Inc., Forest Grove, OR, USA). The temperature programme used was a combination of the NIOSH 5040 and Thermal optical reflectance approaches. A detailed description of the method used can be found in an article by Lee et al. [27]. The standard deviation was calculated from six OC/EC ratios and was below 10%.

In this study, the air flow rate of the sampler was higher than the manufacturer’s specification due to the use of quartz filter in the last stage instead of polyurethane foam. This permitted a larger amount of air to pass through the sampler, resulting in the collection of particles sized 2.07 μm and smaller. Before being used for sampling, the quartz filter was heated to 800°C for 2 hours and cooled to 105°C for approximately 18 hours. After sampling, the filter was wrapped in aluminium foil that had been baked for 2 hours at 250°C, sealed in a Whirl-pak® bag, and kept at 5°C prior to use. The sample filter was shipped to the CanmetENERGY Ottawa Characterization Laboratory in a polystyrene container packed with dry ice. On arrival, the filter was stored in a freezer until it was used. Filter punches (25 mm) were cut out in random locations from the larger quartz filter. These were then transferred and sealed into aluminium foil using cleaned tweezers, and one 25 mm punch was shipped and sent to participating laboratories in the same manner as previously described. The ambient filter samples were sent on a different day than the test solution and the DPM filter samples. Six laboratories performed TD-GC/MS analysis of the ambient filter sample.

2.3. Equipment and analysis

2.3.1. TD-GC/MS systems and analysis

Various TD-GC/MS systems (Tables 1 and 2) from 10 international laboratories were used to perform the study. Each participant used their own internal and calibration standards, and a calibration range suited for the analysis (Table 3). The initiating laboratory investigated the TD-GC/MS parameters prior to the round robin to obtain optimised peak separation and overall compound desorption while minimising analysis time. The low-focusing temperature was chosen to trap volatiles and ensure that breakthrough cross-contamination would not occur, as well as to maximise the recovery of the less-volatile compounds. These parameters were sent to the laboratories, and they were modified as needed by the participants to obtain optimum analysis conditions for their systems. The analysis of light PAH (naphthalene to fluorene) and light n-alkanes (C10–C16) was performed in this study to compare the largest possible amount of compounds between labs. Although these light compounds are known to partition in the gas and particle phase, the paper by Graham et al. [18] has shown good RA and precision by TD-GC/MS analysis. Each participant used their own punch size and shape to cut a sample from the diesel and ambient filters, and a minimum of three replicate analyses was performed for each filter. Tables 1–3 show the conditions, calibration and IS as well as the punch size and shape used by each lab for the TD-GC/MS analysis of the samples. Eight out of 10 laboratories reported their method detection limits (supplementary Table C).

2.3.2. Liquid injection GC/MS

Laboratories 1, 2 and 5 performed the LI-GC/MS analysis. The same parameters and GC types (equipped with liquid autosamplers) as the TD-GC/MS systems were used for the analysis (Table 1).
| Lab | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|-----|---|---|---|---|---|---|---|---|---|----|
| **Gas chromatograph (GC)** | Varian 1200 | Agilent 6890 | Agilent 6890 | Agilent 6890 | Agilent 6890 | Agilent 7800 | Agilent 6890 | Agilent 6890 | Agilent 6890 | Agilent 6890 |
| **Mass spectrometer (MS)** | Varian triple quadrupole | | | | | | | | | |
| **Column Type** | DB-5MS | DB-5MS | DB-SiKe | AT-5MS | Zeb anyhow Inferno | HP-5MS | DB-5MS | HP-5MS | HP-5MS | HP-5MS |
| **Length (m) × i.d. (mm) × film thickness (µm)** | 30 × 0.25 × 0.1 | 60 × 0.25 × 0.25 | 30 × 0.25 × 0.25 | 30 × 0.25 × 0.25 | 30 × 0.25 × 0.25 | 30 × 0.25 × 0.25 | 30 × 0.25 × 0.25 | 30 × 0.25 × 0.25 | 30 × 0.25 × 0.25 | 30 × 0.25 × 0.25 |
| **Flow rate (ml/min)** | 1 | 1 | 1 | 1 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 |
| **Initial temp. (°C)** | 75 | 60 | 75 | 65 | 65 | NA | 40 | 60 | 60 | 40 |
| **Ramp #1 (°C/min)** | 14 | 8 | 14 | 10 | 5.3 | NA | 6 | 5 | 5 | 6 |
| **Temp. #1 (°C)** | 300 | 210 | 200 | None | None | None | None | 150 | None | 120 |
| **Hold time #1 (min)** | 9.39 | 0 | 9.93 | None | None | None | None | 10 | None | None |
| **Ramp #2 (°C/min)** | 10 | 15 | 10 | None | None | None | None | 85 | None | 7 |
| **Temp. #2 (°C)** | 0 | 300 | 0 | None | None | None | None | None | None | None |
| **Hold time #2 (min)** | 0 | 30 | 0 | None | None | None | None | None | None | None |
| **Ramp #3 (°C/min)** | 0 | 10 | 0 | None | None | None | None | None | None | None |
| **Final temp. (°C)** | 320 | 320 | 320 | 300 | 310 | NA | 360 | 320 | 320 | 360 |
| **Final hold time (min)** | 0 | 0 | 0 | 26.5 | 7 | NA | 6 | 10 | None | 18 |
| **Notes:** NA, not available; SIM, single ion mode; TIC, total ion count. |

Table 1. TD-GC/MS systems and parameters.
Table 2. Thermal desorption parameters.

| Lab | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|     | Gerstel TDS | Gerstel TDS 2 | Markes TD | Injection port | TD | Gerstel TDU | Gerstel TDU | Gerstel TDU | Gerstel TDU | Injector port | TD | Gerstel TDU |
|     | Injection type | Automatic | Manual | NR | NR | Automatic | Manual | Automatic | NR | NR | NR | NR |
|     | Initial temp. (min) | 20 | 25 | 0 | 50 | 50 | 12 | 35 | 50 | 50 | 30 |
|     | Rate (°C/min) | 60 | 10 | NA | None | 60 | 60 | 25 | 60 | None | 25 |
|     | Final temp. (°C) | 325 | 300 | 360 | None | 320 | 320 | 320 | 320 | None | 320 |
|     | Hold time (min) | 5 | 5 | 25 | None | 5 | 6 | 5 | 6 | None | 3 |
|     | Mode | Spitless | Spitless | Split | Splitless | Spitless | Spitless | Spitless | Spitless | Spitless | Spitless |
|     | Focussing and desorption parameters |     |     |     |     |     |     |     |     |     |     |
| Conditions | Initial temp. (min) | −60 | −100 | 0 | None | −60 | −100 | −60 | −60 | None | −30 |
|     | Rate (°C/s) | 12 | 12 | NA | 0.46 | 12 | 12 | 12 | 12 | 11 | 12 |
|     | Final temp. (°C) | 325 | 300 | 210 | 275 | 325 | 320 | 310 | 325 | 275 | 340 |
|     | Hold time (min) | 0.5 | 75 | 17 | None | 5 | 3 | 2 | 5 | None | 10 |

Notes: NR, not reported; NA, not available.
Table 3. Calibration curve parameters and filter shape used.

| Lab | Standards | Points | PAHs | n-Alkanes | Biomarkers | Range (µg/mL) | IS PAHs | IS n-Alkanes | IS Biomarkers |
|-----|-----------|--------|------|-----------|------------|----------------|---------|--------------|--------------|
|     | Lab 1     | 2      | 3    | 4         | 5          | 6              | 7       | 8            | 9            | 10           |
|     | Internal  | Internal | Internal | Internal | Internal | Internal | Internal | Internal | Internal | Internal | Internal | Internal |
|     | PAHs      | 5       | 4    | 4         | 8          | 4              | 1       | 6            | 20          | 5            | 4          |
|     | n-Alkanes | 5       | 4    | 4         | 4          | 4              | 3       | 6            | 20          | 6            | 4          |
|     | Biomarkers| 5       | 4    | 5         | 4          | NA             | 6       | 6            | 20          | 6            | 4          |
|     | Bio.      | 0.25–5.0| 0.1–1.0| 0.04–2.668| 0.2–15     | 0.25–2.5       | 2       | 0.05–2.5     | 0.1–2.2     | 0.5–10       | 0.01–4.0   |
| IS  | PAHs      | D6-Naph, D8-Ac, D10-Phen, D12-Benzo(a)anth, D12-Chrys, D12-Benzo(k)fluor D12-Benzo(a)pyrene, D12-Pery, D12-Indeno(1,2,3-c-dipyr, D14-Dibenz(a,h)anth, D12-Benz(g,h,i)per | D10-Ac, D10-Pyr, D10-Benz[a]anth, D12-Cor | D10-Pyr, D10-Benz[a]anth, D12-Cor | n-DC24 | None | D10-Ace, D8-Naph, D10-Phen, D10-Ac, D10-Phen |
|     | n-Alkanes | n-DC15, 20, 24, 32, 36 | n-DC15, 20, 24, 30, 32, 36 | n-DC24 | n-DC12, 16, 20, 24, 30, 32, 36 | None | n-DC24 | n-DC12, 16, 20, 24, 30, 32, 36 | IS PAHs | IS n-Alkanes | IS Biomarkers |
|     | Biomarkers| D6-Cholestane | D4-Cholestane | n-DC24 | 5α(H)-Androstan | NA | 5α(H)-Androstan | IS PAHs | IS n-Alkanes | IS Biomarkers |
|     | Filter sample | 8 | 30 × 2 | 10 × 15 | NR | 8 | 8 | 16 | 9.75 | 6 |

Notes: IS, internal standard; NR, not reported.
2.3.3. **Solvent extraction GC/MS analysis**

Two laboratories performed the SE procedure for the DPM filter with the results to be used as a reference value. Each laboratory performed SE analysis of two separate DPM filters. Their procedures are described in detail in the supplementary information.

2.4. **Statistical analysis**

The average concentrations, RA and relative precision (RP) for the test solution, the DPM filter and the ambient filter sample for each compound class and method can be found in supplementary Tables D–G.

2.4.1. **Accuracy assessment**

The accuracy of the results from each laboratory in the analysis of the test solution, DPM sample and ambient PM sample was assessed through the calculation of $z$-scores. Following the International Union of Pure and Applied Chemistry guidelines [28], the $z$-score is a measure of the difference between the individual result and an assigned standard value, standardised by the standard deviation, which was defined here as a measure of proficiency. For example, if the measurement differs from the assigned value by an amount equal to twice the standard deviation, the measurement would have a $z$-score of 2. The $z$-score provides a standardised statistic that can be used to compare all results from the study. For the data from the test solution analysis, participants’ results have been converted to $z$-scores for accuracy assessment using the following equation:

$$ z = \frac{(x - x_a)}{\sigma_p} $$  \hspace{1cm} (1)

where $x$ is the individual result, $x_a$ is the assigned value which was set as the calculated concentration of the analyte and $\sigma_p$ is the standard deviation chosen for proficiency assessment, defined here as 25% of the assigned value (i.e. $\sigma_p = 0.25 \times x_a$). Therefore, a $z$-score of 1 represents a result that was 25% above the assigned value while a $z$-score of −1 would represent a result 25% below the assigned value. This percentage was chosen as it corresponds to that used by the NIST’s Intercomparison Program for Organic Contaminants in PM 2.5 Air Particulate Matter [29]. For the test solution, the known concentrations were used as the assigned values. A deviation of 25% from the assigned value was chosen for $\sigma_p$. While IUPAC does not recommend the classification of $z$-scores, for this study it was decided that $z$-scores of <2 were acceptable, between 2 and 3 as questionable, and >3 were unsatisfactory.

2.4.2. **Analysis of variance**

Analysis of variance (ANOVA) was used to assess intra-lab differences in results using a nested design. ANOVA is a statistical method used to compare multiple means and is more appropriate than a $t$-test when comparing multiple means and when testing for differences between groups. ANOVA reduces the chance of erroneously concluding there was a difference between labs or samples when there was no actual difference (type I errors) in comparison to multiple $t$-tests. Differences among laboratories were considered significant if $p \leq 0.05$. ANOVA was completed in R version 3.01.

2.4.3. **$t$-Test**

The $t$-test (two-tailed, assumed unequal variances, $\alpha = 0.05$) was used to determine whether there were significant differences between the results measured by the labs using the TD-GC/MS method and LI-GC/MS (labs 1, 2 and 5).
3. Results and discussion

3.1. Evaluation of TD-GC/MS and LI-GC/MS comparison, accuracy and repeatability by using test solution analysis results

3.1.1. Confirmation of test solution concentration values by LI-GC/MS

The concentration values as determined by LI-GC/MS injection for the PAHs, \textit{n}-alkanes and biomarkers and the true values are shown in supplementary Table A. For the PAHs, the true value was set at 0.60 µg/mL. The true values for the biomarkers varied and are listed individually in supplementary Table A. For both PAH and biomarkers, the deviations were less than 15% from the true value, and this was considered confirmation of the concentration of these compounds in the test solution. The concentrations of the \textit{n}-alkanes in the test solution analysed by LI-GC/MS are also shown in supplementary Table A. The values for the \textit{n}-alkanes from decane (C\textsubscript{10}H\textsubscript{22}) to octacosane (C\textsubscript{28}H\textsubscript{58}) were within 10% of the true concentration (1.00 µg/mL). However, the concentrations of five \textit{n}-alkanes larger than octacosane (C\textsubscript{28}H\textsubscript{58}) were between 19% and 39% lower than the true value of 1.00 µg/mL. The low concentration of these heavy \textit{n}-alkanes may be due to them slightly precipitating out during the preparation of the test solution. Provost et al. [30] showed that for relatively heavy \textit{n}-alkanes (C\textsubscript{23}, 25, 26, 28) their solubility decreased with increases in chain length and solvent had no major influence on solubility. Similarly, Ashbaugh [31] demonstrated that the solubility of C\textsubscript{24}, C\textsubscript{28}, C\textsubscript{32} and C\textsubscript{36} in decane decreased with increasing chain size as well as with lower temperatures and that C\textsubscript{32} and C\textsubscript{36} can precipitate near 20°C. Therefore, the concentrations of these larger \textit{n}-alkanes were considered unreliable and have been excluded from the analysis.

3.1.2. Evaluation of TD-GC/MS systems accuracy and repeatability

The test solution, the concentration of which was not known to the participants, was analysed by nine laboratories using TD-GC/MS and three laboratories using LI-GC/MS. The purpose of phase I was to estimate the accuracy and repeatability of the TD-GC/MS systems and to compare them to the LI-GC/MS methods for the analysis of compounds typically found in real-world samples. For the comparison of the results, we defined the accuracy of the method for each target analyte as the difference between the mean of the repeat analyses and the known true concentration. The RA for each target analyte was the accuracy divided by the target analyte concentration, expressed as a percentage. The RP was calculated by dividing the standard deviation by the mean, expressed as percentage. The overall method accuracy was also calculated, to be able to quickly summarise differences between methods. It was defined as the absolute value of the average of the relative accuracies of all analytes of each chemical class. Overall average method precision was similarly calculated to the overall method accuracy, as the average of the RPs of all analytes of each compound class.

The target analytes, average concentrations measured for the test solution, as well as average RA and RP for each analyte obtained by replicate measurements at each lab are shown in Table 4 and in the supplementary Tables D and E.

There were many factors that could influence the accuracy and precision of the TD-GC/MS results. These included the number of calibration standards, the supplier of the calibration standards, the supplier and different compounds used as IS, the type of calibration curves used for the analysis, the TD-GC/MS systems that also comprised different methods of transferring compounds from the sample filters onto the GC column and the injection methods (manual, automatic). Details of these parameters can be found in Tables 1–3.

Average RAs for the PAHs were similar for all labs that used TD-GC/MS, ranging from 9% to 46% (Table 4). However, a closer look at the individual results reveals some notable
differences between the true values that may be explained by the differences in analysis parameters. For example, as shown in supplementary Figure A, lab 3 had low RAs (−36% to −68%) (supplementary Table D) for the PAHs ranging from acenaphthylene to phenanthrene with associated large RP (32% to 11%, respectively, supplementary Table D). Lab 3 used a relatively high-focusing trap temperature (0°C) that may not allow adequate trapping of these small, high-volatility PAHs, resulting in their losses and associated high variability. Labs 4 and 9 that used in-port injection TD-GC/MS did not report naphthalene possibly due to the unreliability of the results caused by the losses that occur during the insertion of the filter punch into the warm liner prior to closing the injection port. Furthermore, labs 4 and 9 obtained large variability in the average RA values for the TD-GC/MS analysis of PAHs (1–92% and 6–176%, respectively, Table 4) with no apparent trends. However, they did obtain an overall RP of 14% and 9%, respectively (Table 4) that was reasonable for this type of analysis. The results of labs 4 and 9 suggested possible coelution of certain PAHs that may be due to the absence of a cryogenic focusing traps, resulting in poorer chromatographic peak shape [32]. Differences in concentrations of the calibration standards may be another factor, given that they had good

| Compound          | Lab | TD-GC/MS | LI-GC/MS |
|-------------------|-----|----------|----------|
| PAHs (16 target EPA) | 1   | 10 (0–37) | 10 (2–22) |
|                   | 2   | 15 (1–55) | 13 (2–41) |
|                   | 3   | 20 (3–68) | 6 (0–15)  |
|                   | 4   | 42 (1–92) | 4 (0–15)  |
|                   | 5   | 15 (6–27) | 9 (5–15)  |
|                   | 6   | 22 (9–50) | 9 (5–15)  |
|                   | 7   | 17 (2–48) | 5 (1–15)  |
|                   | 8   | 9 (0–17)  | 14 (2–32) |
|                   | 9   | 46 (6–176)| 9 (2–30)  |
| n-Alkanes (even n-C10 to n-C28) | 1   | 6 (4–8)  | 16 (0–28) |
|                   | 2   | 30 (23–43)| 24 (11–34)|
|                   | 3   | 21 (1–33) | 12 (1–32) |
|                   | 4   | 15 (5–26) | 4 (2–6)   |
|                   | 5   | 6 (0–24)  | 2 (1–5)   |
|                   | 6   | 33 (3–63) | 20 (15–29)|
|                   | 7   | 6 (0–26)  | 4 (1–17)  |
|                   | 8   | 13 (3–37) | 18 (3–33) |
|                   | 9   | 36 (21–43)| 3 (2–7)   |
| Biomarkers (12)   | 1   | 23 (0–64)| 18 (6–29)|
|                   | 2   | 86 (57–106)| 98 (22–185)|
|                   | 3   | 50 (2–141)| 16 (3–27) |
|                   | 4   | 15 (0–47) | 4 (1–8)   |
|                   | 5   | 10 (2–18) | 5 (4–6)   |
|                   | 6   | NR        | NR        |
|                   | 7   | 32 (18–51)| 2 (1–3)   |
|                   | 8   | 7 (0–12)  | 6 (1–20)  |
|                   | 9   | 53 (42–77)| 9 (5–14)  |

Notes: Results reported as average (min–max).

Average relative accuracy is calculated without regard for sign, i.e. using absolute values of target compound accuracies. NR, not reported.
repeatability as indicated by their relatively low RP. Lab 6 had slightly low RAs for the analysis of the heavier PAHs (RA <–24% from benzo(b)fluoranthene to benzo(ghi)perylene (supplementary Figure A). This lab used a single-point external calibration curve with no IS for the analysis of the PAHs. Consequently, there were no corrections for instrument drift or for losses that could occur due to incomplete transfer of heavier less-volatile compounds from the filter onto the GC column during the desorption step. The other laboratories obtained RAs and RPs for the PAHs that were consistent with this type of analysis, with most results that were within 25% of the true values.

Most labs obtained accurate results for the analysis of n-alkanes (average RAs from 6% to 36%) and good precision (average RPs from 2% to 20%, Table 4). Labs 3, 4 and 9 did not report the light n-alkanes (C10–C14). The calibration range and the lab’s decision to use or not to use IS appeared to be the main factors that affected the accuracy of the medium-range n-alkane (C18–C28) results. Lab 6 had low RA values for the n-alkanes from C18 to C28 (RA <–50%, supplementary Figure B), and it was the only one that did not use IS to correct for the losses of heavier n-alkanes that could occur due to incomplete transfer during the thermal desorption step. Insufficient desorption of heavier n-alkanes is a phenomenon that has been observed by a manufacturer of thermal desorption systems [33]. Lab 2 obtained high-average RAs for the TD-GC/MS n-alkane results (RA values 30 (23–43%)) that were similar to those obtained from the LI-GC/MS injection performed by this lab (Table 4), which suggested that there were differences in concentrations between the calibration standards used by lab 2 and the test solution. High-average RA results of n-alkanes were also observed for lab 9 (RA between 36 (21–43%), Table 4), which may also have been due to the differences in concentrations between calibration standards used and the test solution; but this cannot be confirmed as this lab did not perform LI-GC/MS.

It was observed that lab 7, which used only one mid-range IS (n-DC-24), obtained excellent average RAs between 0% and 26% (Table 4), which were better or similar to the other labs that used more than one deuterated n-alkane IS (labs 1, 2, 3, 5 and 8).

The biomarkers showed the largest average RA range of the three chemical classes: from 7% to 86% with an average RP range of 2–16% for all labs (Table 4). Lab 2 had the largest average RA among all participants, 86% (57–106%), but with small average RP of 5% (4–9%) (Table 4). Lab 2 also performed the LI-GC/MS analysis of the test solution and obtained average RA 98% (22–185%) and RP 5% (2–13%) (Table 4) that was similar to the TD-GC/MS method. For both techniques, the same IS, calibration standards and ranges were used for the analysis of the biomarkers. Furthermore, the concentrations of calibration standards were well within the concentration range of the test solution. These results suggested that the large deviation from true value for the biomarkers may be due to the difference between the actual biomarker concentrations of the calibration standards that were used by lab 2 and those found in the test solution. Lab 3 also had a large average RA range, 50% (21–41%) and relatively large RP at 16% (3–27%) (Table 4). In this lab, the calibration range was one order of magnitude smaller than the concentration of the biomarkers in the test solution (Table 3) and thus apparently not in the linear range of the instrument for these compounds. Furthermore, lab 3 reported that it did not have the full series of biomarkers in its calibration standards and these missing biomarkers were quantified in the test solution by using the response factors of the next closest biomarkers. Lab 9 also had unusually high-average RAs for the biomarker measurements, 53% (42–77%), which cannot be easily explained since this lab used one IS and the calibration curve was in the correct range. It is worth noting that lab 8, with two IS standards, obtained the best RA among all participants, suggesting that the accuracy of the biomarker analysis was improved by the number of IS used.
3.1.3. *z*-Score evaluation of TD-GC/MS and LI-GC/MS

Using the known concentrations as the assigned values, *z*-scores were calculated for each chemical class to assess laboratory proficiency (Figure 1). Figure 1 shows, for each of the nine laboratories and for each method (TD-GC/MS, LI-GC/MS), the median values of the *z*-scores of each chemical class, the values of the 25th (upper quartile) to 75th (lower quartile) percentile (box), and the minimum and maximum *z*-score value (whiskers). The round symbols are outlier data points, which are lower than the lower quartile, or higher than the upper quartile, by more than 1.5 times the interquartile range. For TD-GC/MS analysis of the *n*-alkanes, eight laboratories obtained *z*-scores within ±2, with median values all within 1 and −2. Lab 6 *z*-scores were less than 2 for the *n*-alkanes C18 and above, likely due to the lack of an IS necessary for correcting losses that can occur due to incomplete transfer of the heavier *n*-alkanes during the thermal desorption step, as explained earlier. Overall, Figure 1 shows low dispersion and good performance proficiency for the TD-GC/MS analysis of PAHs. The majority of the PAHs *z*-scores were between ±2 except labs 4 and 9 which obtained a few *z*-scores above 3. Lab 3 had some *z*-scores below 2 for the lighter PAHs (acenaphtylene and acenaphthene) that could be

Figure 1. *z*-Scores of TD-GC/MS and LI-GC/MS for the analysis of the test solution.
caused by losses of the compounds due to the relatively high-focusing temperature used in the participant’s thermal desorption system.

Except for labs 2 and 3, most labs obtained z-scores less than 2 showing adequate performance proficiency for the TD-GC/MS analysis of biomarkers. The higher z-scores of lab 3 were most likely due to the calibration standards outside the linear range of the biomarkers. The high biomarker z-scores for lab 2 appeared to be caused by the difference between the concentration in the calibration standards and in the test solution as indicated previously. Furthermore, lab 2 obtained similar high z-scores for the LI-GC/MS analysis of biomarkers. However, with the exclusion of lab 2 biomarker results, the LI-GC/MS z-scores were all below 2 showing good performance proficiency. Overall, the variability of TD-GC/MS of all compounds was in the same order of magnitude as the LI-GC/MS.

3.1.4. Comparison of TD-GC/MS and LI-GC/MS RA and repeatability

Overall average RA and RP is summarised for the TD-GC/MS and LI-GC/MS in Table 5. Three labs – 1, 2 and 5 – used direct LI-GC/MS to analyse the test solution. Average relative accuracies for TD-GC/MS were similar for PAHs 22% (12–46%) and n-alkanes 17% (10–25%) but slightly poorer for biomarkers 31% (22–44%). Average RA for LI-GC/MS was good for all chemical classes with the exception of a slightly poorer range for the biomarkers. The higher average RAs of the biomarkers were due to the high RAs of lab 2.

Average RP was below 7% for LI-GC/MS and below 10% for TD-GC/MS for all three chemical classes. LI-GC/MS analysis of the test solution had slightly better average RA than TD-GC/MS and, as expected, better precision than TD-GC/MS. When compared to the TD-GC/MS, the LI-GC/MS technique always uses an automated injector to deliver the solution onto the GC, which eliminates human injection as a source of error, has no interference from any compounds that may be desorbed from a filter media and has no desorption or cold trap losses inherent to the TD-GC/MS method. Furthermore, LI-GC/MS has no sample losses due to the SE stage or due to evaporation steps. While the use of internal and recovery standards for SE and TD-GC/MS should normalise for differences in injection quantities or other losses, they apparently cannot completely compensate for these sources of error.

3.1.5. Assessment of TD-GC/MS systems comparability using ANOVA analysis

To determine whether there was a statistically significant difference between laboratories using the same TD-GC/MS systems, a nested ANOVA design ($\alpha = 0.05$) was used (Table 6). While the power of the ANOVAs was limited by the low number of samples

| Compound                        | TD-GC/MS (9 labs) | LI-GC/MS (3 labs) |
|---------------------------------|-------------------|-------------------|
|                                 | Average RA (%)    | Average RP (%)    |
| PAHs (16 target EPA)            | 22 (12–46)        | 10 (2–19)         |
| n-Alkanes (even n-C10 to n-C28) | 17 (10–25)        | 15 (10–19)        |
| Biomarkers (12)                 | 31 (22–44)        | 30 (8–73)         |

Note: Results reported as average (min–max).
per laboratory, they can still provide valid information on differences between the methods for the different chemical classes. For the PAHs, 81% (only 3 out of 16 were different, i.e. acenaphthylene, anthracene and benzo.k.fluoranthene) of the data was not significantly different among labs. The ANOVA test also showed good comparison for the analysis of alkanes as 90% (1 out of 10 was different, i.e. dodecane) of the data show no significant differences between labs. It was found that 11 out of 12 biomarkers were significantly different, indicating that large differences did exist between the different laboratories. The statistical analysis thus suggested that different TD-GC/MS systems may not provide the same results for biomarkers. However, since the methods had good precision, the ANOVA could detect small differences in results. Despite the large differences between labs for the biomarker analysis by different TD-GC/MS systems, the z-scores, as shown in Section 3.1.2, were mostly below 2, indicating good performance for the analysis of these compounds and therefore, from a functional perspective, the degree of differences may be tolerable for this type of analysis.

### 3.1.6. t-Test analysis for estimating TD-GC/MS and LI-GC/MS comparability

For the laboratories that performed both TD-GC/MS and LI-GC/MS (labs 1, 2 and 5), their results from TD-GC/MS were compared against their own results from LI-GC/MS (Table 7). For labs 1 and 2, no significant differences were observed for more than 60% of any of the

| Compound        | TD-GC/MS (9 labs) | TD-GC/MS (9 labs) | TD-GC/MS (6 labs) |
|-----------------|-------------------|-------------------|-------------------|
|                 | Between labs (%)  | Between labs (%)  | Between labs (%)  |
| PAHs (16 target EPA) | 81 (3)            | 92 (1)            | 94 (1)            |
| n-Alkanes\(^a\)  | 90 (1)            | 85 (4)            | 68 (10)           |
| Biomarkers (12)   | 8 (11)            | 93 (1)            | 57 (6)            |

Notes: Values are percentage of non-significant differences ($\alpha = 0.05$), with the number of significantly different analytes in parentheses.

\(^a\)For test solution: even \(n\)-C10 to \(n\)-C28; for DPM and ambient air filter \(n\)-C-10 to \(n\)-C40.

Table 7. Significance test results comparing TD-GC/MS and LI-GC/MS.

| Lab | TD-GC/MS vs. LI-GC/MS |
|-----|-----------------------|
| PAHs (16 target EPA) | 10 of 15 |
| 2              | 13 of 16 |
| 5              | 7 of 16  |
| \(n\)-Alkanes (even \(n\)-C10 to \(n\)-C28) | 5 of 8 |
| 2              | 8 of 10 |
| 5              | 0 of 10 |
| Biomarkers (12)  | 7 of 11  |
| 2              | 2 of 7   |
| 5              | 2 of 12  |

Notes: the \(t\)-test (two-tailed, assumed unequal variances, $\alpha = 0.05$) was used to compare each of the TD-GC/MS methods to LI-GC/MS, where the same laboratory performed both methods.
analytes except for the biomarkers analysed by lab 2. The difference in biomarkers could be attributed mostly to differences in concentration of calibration standards (see Section 3.1.2). For lab 5, there were 7 out of 16 PAHs that were not significantly different; however, all of the \textit{n}-alkanes were significantly different and only 2 out of 12 biomarkers were not significantly different. The RP of lab 5 for \textit{n}-alkanes and biomarkers was very low (2\% and 5\%, respectively, Table 4); therefore, small differences were detected by the \textit{t}-test. Nevertheless, as indicated in Table 5, the relative average accuracy and precision for the TD-GC/MS and LI-GC/MS were similar and close to the true values. As shown in Figure 1, with z-scores mostly between 0 and 2, both methods performed well for the analysis of biomarkers.

### 3.2. Diesel particulate matter sample analysis

The point source sample collected from a diesel engine was analysed by nine laboratories using TD-GC/MS and two laboratories using SE methods. Lab 4 did not perform the analysis but another participant, lab 10, was added to the study. The concentrations of biomarkers and most of the PAHs in the diesel particulate sample were 20–200 times lower than the concentrations in the test solution, while the concentrations of most \textit{n}-alkanes were approximately 2–10 times higher in the diesel particulate sample than in the test solution.

The values obtained for the DPM filter by SE were used as an estimate of the true value, thus average RA was calculated compared to SE method as a reference. The results were much more variable than for the test solution, ranging from 34\% to 166\% for the PAHs, 43\% to 120\% for the \textit{n}-alkanes and 17\% to 394\% for the biomarkers (Table 8).

For the PAHs, a closer observation of the individual component values showed a better comparison between TD-GC/MS results. For example, phenanthrene, which has a concentration near 1.2 ng/cm$^2$ (ng/surface area of filter) that is above the method detection limit (ng on column, supplementary Table C), has the best average RA (27\%) of all PAHs with labs 1–8 that obtained RAs between 0\% and 36\% (supplementary Table F). Lab 9 had an RA of −53\% for phenanthrene that may be due to possible misidentification or difference in concentrations of calibration standards. Lab 10 also had a large RA (54\%) for phenanthrene (supplementary Table F). However, since lab 10 did not perform the analysis of the unknown test solution, the source of the difference was more difficult to identify. For the other PAHs with lower concentrations, the larger RA ranges were probably due to their low concentrations near the detection limit of the systems.

The average RA of the \textit{n}-alkanes ranged from 43\% to 120\% (Table 8). However, for the light \textit{n}-alkanes between C10 and C14 variability was expected since losses occurred both in SE and TD-GC/MS due to evaporation of these compounds from the filter before analysing. Furthermore, evaporative losses of the light \textit{n}-alkanes can occur for SE during solvent reduction, leading to poorer quantification accuracy [34]. For the reported data of \textit{n}-alkanes larger than C24, the individual RAs were mostly below −40\% (supplementary Table F). This suggests that there was an incomplete transfer of the heavier \textit{n}-alkanes to the GC columns during the thermal desorption step, which was not corrected by the IS. It has been shown that contaminants such as non-volatile and polar organics from heavily loaded filters can activate injection port surfaces and interact with the quantifiable compounds. The decrease was especially noted for the heavier compounds due to their longer residence time in the heated injection ports [35]. The RAs were particularly low for lab 3 (most lower than −85\%) for the \textit{n}-alkanes above C24. Lab 3 used a desorption temperature of 210°C that was apparently too low to properly desorb the heavier \textit{n}-alkanes from a DPM filter [36]. Additionally, the analysis of \textit{n}-alkanes was made more complex in a combustion source PM sample due to the presence of the unresolved complex mixture (UCM) that is composed of a mixture of compounds that cannot be fully
resolved by GC methods [37]. The UCM arises from the presence of other many compounds and those of similar saturated non-polar compounds to \(n\)-alkanes, i.e. branched alkanes, and alkylated cycloalkanes, which were also difficult to separate by mass spectrometry because of their similar fragmentation patterns and can interfere with identification of \(n\)-alkane [38]. For the biomarkers, data were only compared to lab 1 since lab 2 did not obtain any results for biomarkers as their concentrations were below the method detection limit. The average RA range was significantly larger than that of the PAHs or the \(n\)-alkanes; however, since the concentrations were small and close to the method detection limit, any small variation in their concentrations could contribute to large relative differences.

Average RP was better for PAHs than for \(n\)-alkanes or biomarkers, exceeding 24% in only one instance (lab 3, Table 8). Overall average RPs (Table 9) were 17% for PAHs, 24% for \(n\)-alkanes and 27% for biomarkers. Precision, plotted against mean concentration (Figure 2) for each analyte and lab, showed that for concentrations above 2 ng/cm\(^2\) the RSD remained below 20% for the majority of cases.

For the TD-GC/MS, the number of analytes with significantly differences between the labs were low (below 15%) for all chemical classes as measured from the ANOVA (Table 6),

Table 8. Average relative accuracy (using SE results as reference) and average relative precision for the analysis of a DPM filter sample by TD-GC/MS.

| TD-GC/MS | Lab | Average RA (%) | Average RP |
|----------|-----|----------------|------------|
| PAHs (16 target EPA) | 1 | 80 (36–131) | 18 (3–38) |
| | 2 | 34 (1–59) | 18 (4–47) |
| | 3 | 42 (3–61) | 24 (10–59) |
| | 5 | 166 (25–677) | 18 (7–61) |
| | 6 | 42 (4–67) | 5 (1–7) |
| | 7 | 51 (25–95) | 8 (1–26) |
| | 8 | 122 (0–248) | 16 (13–20) |
| | 9 | 64 (45–83) | 11 (3–19) |
| | 10 | 144 (53–260) | 6 (0–15) |
| \(n\)-Alkanes (even \(n\)-C10 to \(n\)-C28) | 1 | 98 (8–382) | 12 (1–28) |
| | 2 | 53 (3–87) | 28 (2–124) |
| | 3 | 64 (13–97) | 34 (7–58) |
| | 5 | 120 (8–932) | 9 (0–20) |
| | 6 | 54 (3–188) | 18 (5–79) |
| | 7 | 43 (0–178) | 6 (142) |
| | 8 | 89 (4–288) | 18 (2–99) |
| | 9 | 53 (7–92) | 4 (1–7) |
| | 10 | 71 (10–300) | 27 (0–100) |
| Biomarkers\(^a\) (12) | 1 | 99 (31–180) | 8 (8–9) |
| | 2 | 165 (142–188) | 14 (6–22) |
| | 3 | 44 (21–69) | 23 (9–40) |
| | 5 | 394 (339–415) | 37 (22–49) |
| | 6 | NA | NA |
| | 7 | 73 (49–97) | 16 (8–25) |
| | 8 | NA | NA |
| | 9 | NA | NA |
| | 10 | 16 (17–17) | 13 (13–13) |

Notes: Results reported as average (min–max).
\(^a\)Results compared only to lab 1.
NA, not analysed.
showing overall good comparison of TD-GC/MS results between all labs for these compounds analysed from a DPM filter.

### 3.3. Ambient PM filter sample analysis

The ambient PM sample, collected in a parking lot near commercial and vehicular emission sources in Toronto, was analysed by six laboratories using TD-GC/MS. Precision (%RSD) was generally good, below 20% when the mean concentration was above 2 ng/cm² (Figure 3). Without a consensus or reference value, average RA could not be calculated for the PM sample. However, additional comparison information could be gathered by observing the average relative standard deviation percentage (RSD%) calculated from the mean values obtained for the individual compounds of each laboratories (supplementary Table G). The RSD ranges for the PAHs were between 23% and 118%. The RSD of the PAHs of low molecular weight

| Compounds               | Average RA       | Average RP       |
|-------------------------|------------------|------------------|
| PAHs (16 target EPA)    | 170 (115–253)    | 17 (4–39)        |
| n-Alkanes (n-C10 to n-C40) | 108 (22–397)    | 24 (6–66)        |
| Biomarkers (12)         | 256 (207–316)    | 27 (17–40)       |

Note: Results reported as average (min–max).

![Figure 2](image)

Figure 2. Average relative standard deviation as a function of mean concentration for analysis of DPM filter sample for all methods (all compound classes).
(naphthalene to fluorene) were all above 47%, indicating large variability between labs. This was expected as the variability of these lighter compounds may be related to evaporative losses that occur during the handling of the filter (e.g. transportation, storage, cutting and transferring to TD tube) as well as the losses for volatiles indicated in the text earlier. Furthermore, the concentration of these compounds was near the detection limit, thus increasing variability between labs. The RSDs of the other PAHs (phenanthrene to benzo(ghi)perylene) were all below 51% and were hovering near 20–30%, with the exception of anthracene and dibenz(a,h)anthracene. In the latter case, the RSDs were 118% and 109%, respectively. These two have higher RSD than the rest due to the results of lab 8 that appear to be outliers. When the data of lab 8 were removed for anthracene and dibenz(a,h)anthracene, their RSD became 57% and 23%, respectively, in line with the RSD of the other PAHs. It was difficult to ascertain why lab 8 obtained such results as their analysis of the test solution showed good performance for all PAHs. The RSDs of the n-alkanes were more variable than the PAHs with an overall RSD range of 20% and 140%. The n-alkanes between C13 and C22 exhibited large variations in RSD that were between 47% and 140%. In particular, for these n-alkanes, the concentrations obtained from labs 8 and 9 were significantly higher than those found for the other labs (up to two orders of magnitude higher, e.g. result of C17 for lab 9). Such a large difference in concentration could be the result of an n-alkane contamination. For the n-alkanes between C22 and C40, the RSD% range decreased and was found to be between 2% and 59%, indicating less variability for these compounds, with the exception of C36 and C37 that had RSD of 87% and 107%, respectively. Once more, in this case, lab 9 obtained significantly higher concentration for C36 and C37 than the other labs, thus raising their RSD values. Matrix interference may have also played a role for the variability of the n-alkanes since their RSD range was higher than both the PAH and the biomarkers that had lower concentrations. The RSD range for the biomarkers was between 16% and 70% which was lower than that of the n-alkanes and indicated good comparison between laboratories for the analysis of biomarkers in the ambient PM filter sample.
Overall, differences in results may also be due to some of the factors enumerated in the previous sections (i.e. differences in concentration of calibration standards, possible misidentification due to similar branched \(n\)-alkanes in a matrix sample), but these factors were more difficult to identify with the ambient PM filter sample since there were no reference values. Despite the variability between results, the ANOVA (Table 6) showed good comparison between lab for PAHs (94% non-significant), but slightly lower for \(n\)-alkanes (68%) and biomarkers (57%).

4. Conclusions

This study compared, in the first part, different TD-GC/MS systems amongst themselves and against LI-GC/MS for the analysis of a sample test solution. The various TD-GC/MS showed good performance, compared well between each other and were similar to LI-GC/MS in terms of average RA for PAHs and for \(n\)-alkanes but slightly worse for biomarkers. RP of TD-GC/MS was slightly worse than LI-GC/MS. Differences in accuracy between systems and methods were partly attributed to differences in focusing trap temperatures, possible coelution of compounds, lack of IS and/or differences in concentration and range of calibration standards. The differences of calibration range and concentration of calibration standards used by each lab appeared to be the main issue for the large discrepancy in analysis of biomarkers between labs and methods. The analysis of a DPM sample showed more variability for the PAH and biomarkers due to the concentration of these compounds being near the limit of detection. Variability of \(n\)-alkanes was also high due to possible misidentification with similar branched alkanes or overestimating from coeluting peaks caused by sample with a matrix. Low RAs were observed for the TD-GC/MS analysis of the heavier \(n\)-alkanes, suggesting incomplete transfer of these heavier compounds to the GC column. The analysis of the ambient PM filter sample by the different TD-GC/MS systems showed good comparison between each other.

Overall this study demonstrates that different TD-GC/MS systems could provide similar results for the analysis of particulate organics from real-world point source and ambient PM samples. However, care must be taken to ensure complete desorption of heavier compounds when a sample containing a ‘dirtier’ matrix is analysed. Different types of liners with less-active site may be used to prevent analytes from losses and trapping. Optimised desorption parameters such as higher desorption temperatures and longer desorption times should be used to completely transfer analytes from sample filter to the GC column.

Future investigations should be designed in such a way that the same calibration and IS for all chemical classes would be distributed to all participating laboratories to assess the effects of the different analytical systems versus errors in calibrations.

Acknowledgements

We would like to generously thank all the participants who devoted their time and effort to this study. We would also like to thank all the reviewers for their time and input.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

We would like to thank the Government of Canada Program on Energy Research and Development (PERD) for financial support.
Supplemental data
Supplemental data for this article can be accessed at http://dx.doi.org/10.1080/03067319.2015.1055469

ORCID
Gianni Caravaggio http://orcid.org/0000-0002-1696-1290
Chris Hebbern http://orcid.org/0000-0003-0378-9038
Sabit Cakmak http://orcid.org/0000-0001-9921-2107

References
[1] C.A. Pope III, R.T.R. Burnett, M.M.J. Thun, E.E. Calle, D. Krewski, K. Ito and G.D. Thurston, J. Am. Med. Assoc. 287, 1132–1141 (2002).
[2] R.D. Brook, S. Rajagopalan, C.A. Pope III, J.R. Brook, A. Bhatnagar, A.V. Diez-Roux, F. Holguin, Y. Hong, R.V. Luepker, M.A. Mittleman, A. Peters, D. Siscovick, S.C. Smith, L. Whitser and J.D. Kaufman, Circulation 121, 2331 (2010). doi:10.1161/CIR.0b013e3181dbece1.
[3] K.L. Maier, F. Alessandrini, I. Beck-Speier, T.P.J. Hofer, S. Diabaté, E. Bitterle, T. Stöger, T. Jakob, H. Behrendt, M. Horsch, J. Beckers, A. Ziesenis, L. Hültner, M. Frankenberger, S. Krauss-Etschmann and H. Schulz, Inhal. Toxicol. 20, 319 (2008). doi:10.1080/08958370701866313.
[4] A.J. Friend, G.A. Ayoko, D. Jager, M. Wust, E.R. Jayaratne, M. Jamriska and L. Morawska, Environ. Chem. 10, 54 (2013).
[5] F. Liang, M. Lu, T.C. Keener, Z. Liu and S.-J. Khang, J. Environ. Monit. 7, 983 (2005).
[6] H. Li, C.D. Banner, G.G. Mason, R.N. Westerholm and J.J. Rafter, Atmos. Environ. 30, 3537–3543 (1996). doi:10.1016/1352-2310(96)00051-9.
[7] W.F. Rogge, L.M. Hildemann, M.A. Mazurek, G.R. Cass and B.R.T. Simoneit, Environ. Sci. Technol. 31, 2731 (1997). doi:10.1021/es9609563.
[8] H.A. Bamford, D.Z. Bezabeh, M.M. Schantz, S.A. Wise and J.E. Baker, Chemosphere 50, 575 (2003). doi:10.1016/S0045-6535(02)00667-7.
[9] P. Di Filippo, C. Riccardi, D. Pomata and F. Buiarelli, Atmos. Environ. 44, 2742 (2010). doi:10.1016/j.atmosenv.2010.04.035.
[10] F. Mercier, P. Glorennec, O. Blanchard and B. Le Bot, J. Chromatogr. A. 1254, 107 (2012). doi:10.1016/j.chroma.2012.07.025.
[11] A. Falkovich and Y. Rudich, Environ. Sci. Technol. 35, 2326 (2001). doi:10.1021/es000280i.
[12] E. Grandesso, P. Pérez Ballesta and K. Kowalewski, Talanta 105, 101 (2013). doi:10.1016/j.talanta.2012.11.047.
[13] M.D. Hays, N.D.D. Smith, J. Kinsey, Y. Dong and P. Kariher, J. Aerosol Sci. 34, 1061 (2003). doi:10.1016/S0021-8502(03)00080-6.
[14] S.S.H. Ho, J.Z. Yu, J.C. Chow, B. Zielinska, J.G. Watson, E.H.L. Sit and J.J. Schauer, J. Chromatogr. A 1200, 217 (2008). doi:10.1016/j.chroma.2008.05.056.
[15] J. Gil-Moltó, M. Varea, N. Galindo and J. Crespo, J. Chromatogr. A 1216, 1285 (2009). doi:10.1016/j.chroma.2008.12.080.
[16] R.J. Sheesley, J.J. Schauer, E. Garshick, F. Laden, T.J. Smith, A.P. Blicharz and J.T. Deminter, J. Expo. Sci. Environ. Epidemiol. 19, 172 (2009). doi:10.1038/jes.2008.11.
[17] B.L. Van Drooge, I. Nikolova and P.P. Ballesta, J. Chromatogr. A 1216, 4030 (2009). doi:10.1016/j.chroma.2009.02.043.
[18] L.A. Graham, A. Tong, G. Poole, L. Ding, F. Ke, D. Wang, G. Caravaggio, J.-P. Charland, P. MacDonald, A. Hall, Y. Cheng and J.R. Brook, Int. J. Environ. Anal. Chem. 90, 511 (2010). doi:10.1080/03067310903108352.
[19] D. Waterman, B. Horsfield, F. Leistner, K. Hall and S. Smith, Anal. Chem. 72, 3563 (2000). doi:10.1021/ac991372x.
[20] J.C. Chow, J.Z. Yu, J.G. Watson, S.S.H. Ho, T.L. Bohannan, M.D. Hays and K.K. Fung, J. Environ. Sci. Health, A 42, 1521 (2007). doi:10.1080/10934520701513365.
[21] J. Schnellekreis, M. Sklorz, A. Peters, J. Cyrys and R. Zimmermann, Atmos. Environ. 39, 7702 (2005).
[22] E. Kabir and K.-H. Kim, J. Chromatogr. Sci. 49, 731 (2011). doi:10.1093/chrsi/c49.9.731.
[23] L.C. Ding, F. Ke, D.K.W. Wang, T. Dann and C.C. Austin, Atmos. Environ. 43, 4894 (2009). doi:10.1016/j.atmosenv.2009.07.016.
[24] M. Bates, P. Bruno, M. Caputi, M. Caselli, G. De Gennaro and M. Tutino, Atmos. Environ. 42, 6144 (2008). doi:10.1016/j.atmosenv.2008.03.050.

[25] R.J. Lavrich and M.D. Hays, Anal. Chem. 79, 3635 (2007). doi:10.1021/ac0623282.

[26] A. Fushimi, S. Hashimoto, T. Ieda, N. Ochiai, Y. Takazawa, Y. Fujitani and K. Tanabe, J. Chromatogr. A 1252, 164 (2012). doi:10.1016/j.chroma.2012.06.068.

[27] P.K.H. Lee, J.R. Brook, E. Dabek-Zlotorzynska and S.A. Mabury, Environ. Sci. Technol. 37, 4831 (2003). doi:10.1021/es026473i.

[28] M. Thompson, S.L.R. Ellision and R. Wood, Pure Appl. Chem. 78, 145 (2006). doi:10.1351/pac200678010145.

[29] M.M. Schantz, D.L. Poster, J. Kucklick, S.A. Wise and S. McDow, J. Lewtas, Intercomparison Program for Organic Speciation in PM 2.5 Air Particulate Matter: Description and Results for Trial III, NIST Interagency/Internal Report (NISTIR) (National Institute of Standards and Technology, Galthesburg, MD, 2006), Vol. 7303.

[30] E. Provost, V. Chevallier, M. Bouroukba, D. Petitjean, M. Dirand, E.N. Supe, I. Chimiques, I. National, P. De Lorraine and N. Cedex, J. Chem. Eng. Data 43, 745 (1998). doi:10.1021/je980027m.

[31] H.S. Ashbaugh, A. Radulescu, R.K. Prud’honne, D. Schwahn, D. Richter and L.J. Fetters, Macromolecules 35, 7044 (2002). doi:10.1021/ma0204047.

[32] P.R. Loconto Trace Environmental Quantitative Analysis: Principles Techniques and Applications (CRC Press, Boca Raton, FL, 2006), p. 391.

[33] B. Kolahgar and E. Pfannkoch The Use of Different PTV Inlet Liner Types for Trapping Alkanes, Aromatics and Oxygenated Compounds During Thermal Desorption. Technical Note (Gerstel Inc., Baltimore, MD, 2002).

[34] E. Swartz, L. Stockburger and L.A. Gundel, Environ. Sci. Technol. 37, 597 (2003). doi:10.1021/es011128z.

[35] S.S.H. Ho, J.C. Chow, J.G. Watson, L.P. Ting Ng, Y. Kwok, K.F. Ho and J. Cao, Atmos. Environ. 45, 1491 (2011). doi:10.1016/j.atmosenv.2010.12.038.

[36] S.S.H. Ho and J.Z. Yu, J. Chromatogr. A 1059, 121 (2004). doi:10.1016/j.chroma.2004.10.013.

[37] A.W.H. Chan, G. Isaacman, K.R. Wilson, D.R. Worton, C.R. Ruehl, T. Nah, D.R. Gentner, T.R. Dallmann, T.W. Kirchstetter, R.A. Harley, J.B. Gilman, W.C. Kuster, J.A. De Gouw, J.H. Offenberg, T.E. Kleindienst, Y.H. Lin, C.L. Rubitschun, J.D. Surratt, P.L. Hayes, J.L. Jimenez and A.H. Goldstein, J. Geophys. Res. Atmos. 118, 6783 (2013). doi:10.1002/jgrd.50533.

[38] G.A. Caravaggio, J.-P. Charland, P. Macdonald and L. Graham, Environ. Sci. Technol. 41, 3697 (2007). doi:10.1021/es062233h.