Interactive comment on “Mapping and quantifying isomer sets of hydrocarbons (≥C₁₂) in diesel fuel, lubricating oil and diesel exhaust samples using GC×GC-ToFMS” by Mohammed S. Alam et al.

Anonymous Referee #1

Received and published: 4 February 2018

In this work, the authors Alam et al. present a novel method to characterize the composition of organic mixtures, especially those from fossil fuel combustion sources. This is an important topic for atmospheric sciences, as many of these mixtures can react in the atmosphere and form oxygenated compounds and organic aerosol. This work uses a similar approach as work by Zimmerman et al. and Goldstein et al., and combines low energy ionization with multidimensional chromatography to separate and identify aliphatic and aromatic isomers. This work is of clear interest to the field, and is within the scope of AMT. Overall, the work here is quite solid. I suggest a few more discussion points that would advance the understanding of the technique, so that the technique is more likely to be adopted by researchers in this field.

RESPONSE

We thank the anonymous reviewer’s comments and have addressed the comments raised below.

1. Ionization temperature plays an important role in the fragmentation of the molecule. The thermal energy can lead to substantial fragmentation, e.g. Isaacman et al. show that lowering ion source temperature from 300C to 150C reduces fragmentation temperature significantly. Amirav et al. (2008) uses a supersonic molecular beam to achieve cold electron impact ionization, and was able to reduce fragmentation substantially. Have the authors tried to lower ionization source temperature?

RESPONSE

We have not lowered the ionisation source temperature to reduce the fragmentation in this study. Rather we have focused on reducing the ionisation energy only to reduce fragmentation. Using the select-eV BenchTOF, the user can amend the ionisation between 10eV and 70eV and we have assessed the mass fragmentation patterns between 10-15eV and 70eV ionisation energies, without changing the ion source temperature. The transfer line and ion source temperatures used in this study are the recommended temperatures by the manufacturer (Markes International). We expect the fragmentation to reduce and to retain a larger fraction of the molecular parent ion if we were to reduce the ion source temperature, like that reported by Isaacman et al. However, this has not been incorporated into this study as it is independent of ion source temperature.

A sentence has been added to the end of the conclusion. See lines 540 – 547.

2. Is isomer-level differentiation/quantification necessary? There is some debate whether or not the isomers are different enough to warrant detailed characterization. I suggest the authors look at Tkacik et al. (2012) and Goodman-Rendall et al. (2016). In other words, it might be sufficient for process level understanding of hydrocarbon composition (e.g. volatility basis set approach or 2-product Odum model) to predict oxidation characteristics and volatility
changes. I think this is an important point: the authors will first need to identify the need to perform this level of chemical characterization before they will be able to convince the field that this technique is useful.

**RESPONSE**

*Volatile is not the only important characteristic of a VOC in determining its fate in the environment or its environmental impact.* We have included an additional few sentences in the introduction to include these important points raised by the reviewer. See lines 72 – 76 and 115 to 125.

3. Can the authors explain why the total ion current is so similar between different compound types? Ion sensitivity is a strong function of electron absorption cross section, and at lower energies, there is significant differences in cross section. Aromatic compounds, for example, would absorb much more effectively than aliphatic compounds, and cyclic compounds (six- and five-carbon ring) would also absorb more than cyclic compounds. I understand that the authors are just showing their results, but they seem a little counter-intuitive to me.

**RESPONSE**

*Chromatograms shown in Figures 1, 3 and S5 and S7A in this study are those performed with 70eV ionisation mass spectrometry, which we use for quantitative analysis. The 12eV spectra we use qualitatively (see Figure 2 and S7B). This may explain to the reviewer why the total ion current is so similar between the different compound types – because of the relatively large energy used for fragmentation.*

A sentence has been added to the manuscript to explain this. See lines 237 – 238.

All figures of chromatograms within the manuscript and supplementary information have included their respective ionisation energies of the mass spectrometer in their legends.

4. While this technique would be ideal for investigating fossil fuel sources (e.g. motor vehicles, oil extraction activities), how would this work for atmospheric mixture where there will be many other types of compounds and sources? In urban areas, the HOA factor from aerosol mass spectrometer studies accounts for around 20% of total OA, so the majority will be oxygenated. What happens when an urban aerosol sample is analyzed using this technique? Will the “contamination” by oxygenated compounds limit the ability to measure the hydrocarbons?

**RESPONSE**

*We have used this technique on atmospheric samples and have found that although there are an increased number of oxygenated species present, we can still adequately identify and quantify the hydrocarbons that are measured in this study. In the gas phase the majority of the aliphatic oxygenated compounds are present in specific bands above the monocyclic alkanes and below the monocyclic aromatics which can be identified in the atmospheric samples using additional CLIC expressions and fragmentation patterns (as outlined in the conclusion lines 537 – 540). A section has been incorporated into the conclusion to expand on this point made by the reviewer (see Conclusion, line 532 - 541)*

5. Given that this is an atmospheric journal, the writing should reflect the focus on atmospheric applications. For example, the discussion about other environmental mixtures in the conclusions may not be necessary and seem to detract from the main point.

**RESPONSE**

*This has been removed. We believe that this manuscript is now more focused on the atmospheric community with the further modifications that we have made. We have also*
modified the paper title to: “Mapping and quantifying isomer sets of hydrocarbons (≥C12) in diesel exhaust, lubricating oil and diesel fuel samples using GC×GC-ToFMS”

See lines 28 – 29, 44 – 45, 72 – 76, 115 – 125, 146 – 158 and 532 – 547

Minor comments:

1. Samples from the diesel exhaust are obtained from extraction using dichloromethane. What is the extraction efficiency for oxygenated compounds using DCM as a solvent?

**RESPONSE**
We have spiked a known concentration of an oxygenated standard (2-heptadecanone) onto a filter, extracted using DCM and analysed using the GCxGC and found that the extraction efficiency for that oxygenated compound was >92%. This gives us confidence that the extraction with DCM can obtain oxygenated compounds. The reason why we chose DCM as a solvent was to extract as many of the compounds present on the filter as possible, over the polar and non-polar range. We have performed multiple analyses with hexane, DCM and methanol as extraction solvents, and found that DCM performed the best in extracting the majority of the polar/non-polar compounds.

We have included a sentence in the supplementary information in Section S1. See lines 38 – 40.

2. Lines 231-233: How reproducible are the retention times in both dimensions? In other words, are the polygons drawn to separate different compound types consistent from one run to another, one column set to another (assuming the same type of columns), one day to another?

**RESPONSE**
Yes. The retention times are reproducible in both dimensions in separate runs. The polygons are all linked together in a large template that can also be linked to the internal standards. This means that it is easy to align the template (all polygons) in the event of slight shift in retention times. These shifts, however, only occur when the columns are changed or if there has been some major maintenance performed on the instrument.

We have included a few sentences to the manuscript to explain this. See lines 284 – 287.

3. Line 310: Carboxylic acids are very difficult to be detected by GC/MS without prior derivatization. If they are detected, their quantities might be highly uncertain. It might be difficult to say whether they are truly minor, or it is just a limitation of GC/MS.

**RESPONSE**
The reviewer raises a good point and this has been incorporated into the manuscript. See lines 380 – 381.

4. Line 274: aren’t saturated cyclic alkanes also alkanes?

**RESPONSE**
Yes, but we refer to straight chain saturated linear alkanes as “linear alkanes” and saturated cyclic alkanes as “cyclic alkanes”.
The manuscript remains unchanged.

References:

Interactive comment on “Mapping and quantifying isomer sets of hydrocarbons (≥C₁₂) in diesel fuel, lubricating oil and diesel exhaust samples using GC×GC-ToFMS” by Mohammed S. Alam et al.

Anonymous Referee #2

Received and published: 20 January 2018

This article describes a novel GCxGC-ToF-MS approach for the characterization of carbonaceous material (oil and aerosol emissions). The research is novel and worth publishing although it is not clear why this necessarily belongs in an atmospheric journal as most of the work and discussion is on oil. The only, but serious weakness is in the writing itself. The text looks more like being part of an extensive report or a thesis than a scientific manuscript with constant referring to 3 other papers by the authors as well as abundant references to the SI. In its current form it is a painful read, even for people interested in the topic, and I find that it is not really able to stand on its own without the information in these other sources. I think the manuscript should be substantially cleaned up to be more readable and tell a consistent and concise story, with the critical information if really not in the paper itself then in 1 single place. For the least, some clarifications are needed and this being an atmospheric measurement journal, it would be useful to emphasize the atmospheric relevance by a clearer comparison to what is known in the field.

RESPONSE

We thank the anonymous reviewer for their comments and have addressed their comments below.

We believe that this study belongs to an atmospheric journal as it presents data not only regarding oil and fuel, but also includes detailed analyses of gas and particulate phase diesel emissions; which are emitted into the atmosphere and can contribute to secondary organic aerosol. The technique of GCxGC-ToFMS has been adopted by the atmospheric science community in the past (the groups of Goldstein (Berkeley) and Lewis (York)), but has not yet exploited the use of variable or low ionisation energy mass spectrometry presented in this study. The work by Goldstein et al., and Zimmerman et al., have exploited photoionisation VUV for low ionisation MS; and this study presents the application of variable ionisation EI mass spectrometry to map and quantify isomer sets of hydrocarbons. This technique can be applied to complex matrices, particularly diesel emissions and atmospheric samples. For these reasons, we believe that this technique can be adopted by many researchers in the field, as anonymous reviewer 1 has already suggested.

The reviewer points out that the text refers to 3 other papers and to the SI substantially. We have addressed this by including some concise information within the manuscript so that the reader no longer has to refer to the other papers and/or the SI (unless they are interested in further information).

Major changes within the manuscript include:

1) The second paragraph in the Experimental section (2.1 sampling) has been edited to include information regarding the engine set up and sampling procedure. This
eliminates the referral to Alam et al. (2016c). What was formerly the second paragraph has been moved to the front of this section and is now paragraph one.

2) Section S3 in Supplementary Information has been modified and moved to the manuscript under a new additional Section 2.5 “Quantification of compounds with no authentic standards”. Section S3 in the SI now only contains the uncertainty calculation for this method.

3) A sentence referring to the potential of scaling ion current to molar quantity (lines 302-304) and its corresponding section in the Supplementary Information (Section S4) has been removed.

4) Supplementary Section S6 has also been removed, as suggested by the reviewer below.

5) We have modified the paper title to: “Mapping and quantifying isomer sets of hydrocarbons (≥C₁₂) in diesel exhaust, lubricating oil and diesel fuel samples using GC×GC-ToFMS”

Major points:
Please review what is really necessary here as method vs what can be in SI and vice versa and in regards to the other manuscripts. Some critical info is lacking as example: what mass resolution is used for the quantification? In no place the resolution of the TOF is mentioned and some spectra in SI are at unity resolution while others are at .1 Da resolution? This would matter.

RESPONSE
We have reviewed what should be in the manuscript and the SI concisely. Section S3 in Supplementary Information has been modified and moved to the manuscript under a new additional Section 2.5 “Quantification of compounds with no authentic standards”. Section S3 in the SI now only contains the uncertainty calculation for this method. Also Sections S4 and S6 have been removed from the SI also.

The mass resolution used for quantification has been added to the manuscript together with the resolution of the TOF. We use a nominal resolution of unity Da for quantification. See lines 194 to 197.

The text should be focused more on the atmospheric community. This starts with the abstract where quantitative info is only given for oil and goes through the text. It is quite unsatisfying to just point out that the results are different from e.g. Gentner’s work but then not attempt at all is being made to explain why? the same is true for other places of the manuscript where the text leaves it at: we see more/less than these other people: : : These differences should be addressed.

RESPONSE
The abstract has been amended to include quantitative information for gas and particulate phase emissions. We have also included additional text in the conclusion to address the use of this technique with atmospheric samples and oxygenated compound. We believe that this manuscript is focused on the atmospheric community with the modifications that we have made. We have also modified the paper title to: “Mapping and quantifying isomer sets of hydrocarbons (≥C₁₂) in diesel exhaust, lubricating oil and diesel fuel samples using GC×GC-ToFMS”

See lines 28 – 29, 44 – 45, 72 – 76, 115 – 125, 146 – 158 and 532 – 547
The differences/similarities that we observe for diesel fuel with Isaacman et al. (2012) are attributable to the different fuel formulations and/or fuel source, as opposed to analytical methods. This is written in the manuscript (lines 342 – 344).

Line 374 has been deleted “in contrast with Worton et al. (2014)”. This statement can be misleading as Worton et al. do not measure constituents in the gas phase and only focus on the particulate phase. In this section we state that we think it is unlikely for there to be a contribution from lubricating oil to the gas phase emissions.

For the analysis of lubricating oil we attribute the differences seen between the work of Worton et al. and ours to be the varying crude oil origins and formulation processes involved in the production of synthetic oil. This is stated in lines 442 – 447.

In that regards too, there are the highly cited papers by the Cass group (Rogge et al papers) that discuss for 1D GC extractable vs elutable, resolvable, identifiable for oils and diesel emissions: : : given that these papers are extremely known and cited, may be data could be compared? To illustrate the advance of this methodology over the 20 year old work that is cited hundreds of times.

**RESPONSE**

We have included the highly cited paper of Rogge et al. (1993) in the introduction of the manuscript but not compared the results in the discussion section as these studies were conducted 25 years ago where only 10-15% of the organics were resolved using GC/MS. Furthermore, the sulphur content in the diesel fuel would have been significantly larger than that of the diesel fuel today and so emissions from diesel exhaust would not be directly comparable. See Jones et al. (2012) for more information. (Jones et al., A large reduction in airborne particle number concentrations at the time of the introduction of “sulphur free” diesel and the London Low Emission Zone, Atmospheric Environment 50, 129-138). See lines 106 – 108.

The methodology of the quantification should be clarified in the present manuscript. example_ L254-2” Therefore, out of the 8026 ng that was injected into the GC, a mass of approximately 7200 (_ 1728) ng was accounted for.” It is unclear here: : : how the 8026 ng was determined? and why does it not have an uncertainty? Similarly how was the uncertainty for the 7200 obtained?

**RESPONSE**

The value of 8026 ng of diesel fuel was the calculated mass that was injected into the GC. A known volume of diesel fuel was weighed and diluted, followed by injection into the GC. The uncertainty associated to 8026 is the combined uncertainty of the apparatus that was used to calculate the mass (i.e. the combined uncertainties of the balance used to weigh the diesel fuel, the pipette used to measure the volume of the fuel and the volumetric flasks used to dilute the fuel; which is 0.3%). This has been added into the manuscript. See line 319-321.

The uncertainty of the mass that was accounted for using the technique is 24% which is the estimated uncertainty of the technique and is discussed in Section 2.5 in the manuscript and explained in detail in the SI Section S3.

Please clarify how some homologous series were quantified for which you do not list standards (hopanes and steranes etc)?

**RESPONSE**
Section S3 in Supplementary Information has been modified and moved to the manuscript under a new additional Section 2.5 “Quantification of compounds with no authentic standards”. Section S3 in the SI now only contains the uncertainty calculation for this method.

We do not quantify hopanes and steranes as outlined in the manuscript. See lines 424 to 430 which states: “Adamantanes, diamantanes, pentacyclic and hexacyclic alkanes, steroids, steranes and hopanes represented 5% of the total ion current, while the remaining 4% remained unidentified. These compounds were not quantifiable using this methodology, as there were no standards available that corresponded to these sections of the chromatography and could not be estimated as they are not present in a homologous series. However, from previous literature, these compound classes are thought to represent a small fraction of the mass (Worton et al., 2015).”

I am deeply confused also on the mass balance how any non-organic/hydrocarbon species are handled? Especially synthetic motor oil could have a substantial fraction Si or other non HCO materials. Related also you refer to synthetic oil and contribution form crude. This is confusing as synthetic oil could be completely synthetic i.e. nothing in there has ever seen crude.

**RESPONSE**

We have not taken into account any non-organic/hydrocarbon species but according to the data presented in this study the fraction of Si or other non HCO materials would be small.

Synthetic lubricating oil can be manufactured using chemically modified petroleum components rather than whole crude oil. This means that it has seen a fraction of crude oil.

A sentence has been added to the manuscript to include this point. See lines 430 – 432.

**Minor points:**
The text needs some attention to detail. In the experimental section, please use a consistent form to refer to manufacturer, location etc.. also check for typos, USA not US etc Do states need to be included for USA?

**RESPONSE**
The text has been amended to refer to manufacturer, location - city, location – country. See line 164
States within the USA have been removed. See line 188
Addition of “A” for USA. See line 200

I am not a native speaker but the English clearly seems off at times e.g. 187-190, two sub sentences starting with while.. reads really odd?

**RESPONSE**
Replaced “while” in the first sub sentence to “at the same time as..” See line 234 – 235.

References cite in chronological and antichronological order in the text not random

**RESPONSE**
Only three references that have been cited from the same author in the same year have all been cited in chronological order in the text. For example: Alam et al. (2016a) has been cited
in line 83, Alam et al. (2016b) has been cited in line 137 and Alam et al. (2016c) has been cited in line 158. We have checked the references thoroughly and have not identified any other issues.

The text remains unchanged.

Details: L129: Ref: EU: : : please write as citation

**RESPONSE**

Amended.

See line 167 – 168, reference added also, see line 617 – 619

L153: “6”? number reference?

**RESPONSE**

Replaced with Alam et al. (2016a)

See line 199

L253 I am not sure it is appropriate to label column bleed as contaminant plus there is no evidence that this is only column bleed. Siloxanes might well be present in synthetic oil or even ambient air.

**RESPONSE**

We have performed extensive analyses on ambient air samples, lubricating oils, diesel fuel and associated field and lab blanks. In all of these analyses we have observed these siloxanes, so we are very confident that this is column bleed.

The manuscript remains unchanged.

L322: this is experimental and does not belong in the results section

**RESPONSE**

We have removed a portion of this section that is reproducing what is already in the experimental section (section 2.4). Lines 388 – 394 have been removed.

**SI section:**

Page 4: formatting issue

**RESPONSE**

This is not a formatting issue. This is the code that is used for the CLIC expressions that have been discussed in the manuscript (section 2.4). These are code that can be extracted and used as is if this technique was to be utilised.

Page 7: Liang et al 2017 in prep?

**RESPONSE**

This has been removed

Figure S4-1: legend X axis? What is 1/MW (C24-C12 Molar Quantity)? Explain clearer and make legend text consistent with actual legend
S4-2 left panel idem to S4-1 plus straight unreadable (font size)
S4-3 make labeling consistent with other figures

**RESPONSE**

This section has been removed altogether to keep the manuscript and supplementary information concise.
All of the double MS plots, please make them look like the ones in the actual paper or I suggest that on one panel at least the eV are indicated as for MS community this way of showing spectra is typically how + and – ion spectra, so could be confusing.

**RESPONSE**

This has been amended and the eV has been labelled on each panel.

See SI Figures S4-1 to S4-3 and Figure S6-1 to S6-7.

The model spectra, are they corrected for background? how were they obtained? some have many peaks (PAH)?

**RESPONSE**

These are not modelled spectra. These mass spectra are experimentally determined at 70eV and 12eV like explained in the text in Section S5. The text has been amended to add the words “Experimentally determined…” See line 224 in SI.

A mass spectrum is determined by selecting a peak of interest in the GC chromatogram. That peak will have its own mass spectrum and can then be compared to the NIST library for positive identification. This is the principle of operation of the GCxGC-TOFMS. If there is co-elution of peaks then the mass spectrum of those peaks would have been merged and will need to be deconvoluted using the GC Image software.

The reason that some of the mass spectra in Section S4 have multiple peaks is because of the extensive fragmentation at 70eV, which is discussed in detail throughout the manuscript. The PAH have many peaks present at 70eV which is in line with the NIST 70eV library. For example, the 70eV MS for phenanthrene (see Figure S4-3 below, top right panel) possessed the mass fragments m/z 76, 89 and 152, corresponding to C6H4, C7H5 and C12H8 fragments, respectively. This mass spectrum is identical to that of phenanthrene in the NIST library at 70eV and is confirmed by the 12eV mass spectrum which only possesses the molecular mass 178.
**Figure S4-3.** Representative mass spectra of four PAH. Top graphs with red peaks are 70eV ionisation mass spectra, while bottom graphs with blue peaks are 12eV spectra.

S6 what is really the use of this if not even the position of one relative to the other are shown in fig 6

**RESPONSE**

This has been removed from the SI

Please even in SI use a uniform formatting of references

**RESPONSE**

Amended
Mapping and quantifying isomer sets of hydrocarbons (≥C\textsubscript{12}) in diesel fuel exhaust, lubricating oil and diesel exhaust fuel samples using GC\texttimes{}GC-ToFMS

Mohammed S. Alam\textsuperscript{1}, Soheil Zeraati-Rezaei\textsuperscript{2}, Zhirong Liang\textsuperscript{3}, Soheil Zeraati-Rezaei\textsuperscript{3}, Christopher Stark\textsuperscript{1}, Hongming Xu\textsuperscript{23}, A. Rob MacKenzie\textsuperscript{1¥} and Roy M. Harrison\textsuperscript{1*†}

\textsuperscript{1} Division of Environmental Health and Risk Management, School of Geography, Earth and Environmental Sciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom

\textsuperscript{2} Department of Mechanical Engineering, University of Birmingham, Edgbaston Birmingham, B15 2TT, United Kingdom

\textsuperscript{2,3} School of Energy and Power Engineering, Beihang University, Beijing, 100191 China

\textsuperscript{*} To whom correspondence should be addressed. Te: +44 121 414 3494; Fax: +44 121 414 3709; Email: r.m.harrison@bham.ac.uk

\textsuperscript{†} Also at: Department of Environmental Sciences / Center of Excellence in Environmental Studies, King Abdulaziz University, PO Box 80203, Jeddah, 21589, Saudi Arabia

\textsuperscript{¥} Also at: Birmingham Institute of Forest Research, University of Birmingham, B15 2TT
ABSTRACT

Airborne particles and vapours, like many other environmental samples, including water, soils, and sediments, and airborne particles and vapours contain complex mixtures of hydrocarbons, often deriving from crude oil either before or after fractionation into fuels, lubricants and feedstocks.

Comprehensive 2D Gas Chromatography- Time-of-Flight Mass Spectrometry (GC×GC-ToFMS), offers a very powerful technique separating and identifying many compounds in complicated hydrocarbon mixtures. However, quantification and identification of individual constituents at high ionization energies would require hundreds of expensive (if—when available) standards for calibration. Although the precise chemical structure of hydrocarbons does matter for their environmental impact and fate, strong similarities can be expected for compounds having very similar chemical structure and carbon number. There is, therefore, a clear benefit in an analytical technique which is specific enough to separate different classes of compounds, and to distinguish homologous series, whilst avoiding the need to handle each isomer individually. Varying EI (electron impact) ionization mass spectrometry significantly enhances the identification of individual isomers and homologous compound groups, which we refer to as ‘isomer sets’. Advances are reported in mapping and quantifying isomer sets of hydrocarbons (≥C_{12}) in diesel fuel, lubricating oil and diesel exhaust emissions. Using this analysis we report mass closures of ca. 90% and 75% for diesel fuel and lubricating oil, and identify 85% and 75% of the total ion current for gas and particulate phase diesel exhaust emissions.
1. INTRODUCTION

Crude oil contains a highly complex mixture of chemical constituents, mainly hydrocarbons (C₄-C₅₅) (Riazi, 2005). There are many reports of crude oil entering the environment through spillage, or deliberate release (Gertler et al., 2010). Most crude oil is treated and fractionated in order to produce fuels and lubricants for use in transport and combustion applications, and as feedstocks for the chemical industry (Riazi, 2005). All of these uses have a potential to contaminate the environment. Understanding fates, pathways, and effects of contamination requires chemical analysis and detailed interpretation of resulting data. Much of the chemical complexity of oil derives from the large numbers of straight and branched chain and cyclic hydrocarbon isomers for a given carbon number (Goldstein et al., 2007). Hence, analytical methods are required that can discriminate structurally similar sets of isomers in complex media.

Application of conventional gas chromatographic methods to oils and oil-derived samples was for many years severely limited by the poor separation capability of one-dimensional chromatography, due to the near-continuous range of physicochemical properties of hydrocarbons. Thus, typically 90% of the hydrocarbon content of the sample is present in the unresolved complex mixture (UCM), creating a large hump in the chromatogram (Fraser et al., 1998; Schauer et al., 1999). The advent of two-dimensional gas chromatography, which provides enhanced separating capability due to the orthogonal separation by two capillary columns of different stationary phases, has transformed the problem by resolving the UCM into many thousands of individual compound peaks. The two columns are connected in series by a modulator which is employed to provide focusing of the primary column eluent (Liu et al., 1991; Phillips et al., 1995). Large amounts of data are produced...
due to the large number of compounds separated. This technique generates very large volumes of mass spectral data, although it is often generic. The information which is required is most useful scientifically in order to compare the main compositional attributes of samples, is often more detailed than that provided by bulk Total Hydrocarbon measurements or fractionation into a few volatility classes, but less detailed than detailed-specific identification of many compounds.

Use of a flame ionisation detector has the advantage of allowing generic quantification of any part of the chromatogram in terms of the carbon mass contained within it, but identification of specific chemical constituents with this detector can only be achieved on the basis of retention times which, in a very complex two-dimensional chromatogram and set-up-dependent chromatogram, are laborious to assign objectively. Mass spectrometric detection, especially when employing both low and high ionisation energies adds a third analytical dimension with the ability to overcome the problem of compound identification (Alam et al., 2016a), but has not generally been applied to generic quantification of compound groups within complex samples. In this study, we show that time-of-flight mass spectrometric detection can be used not only to identify and quantify individual chemical constituents within the chromatogram, but can also be used to quantify generic groups of compounds.

Motor vehicles are a major source of organic carbon in the environment, and the majority of the fine particulate matter (PM) emitted is carbonaceous, directly emitted as primary organic aerosol (POA) or formed as secondary organic aerosol (SOA) (Jimenez et al., 2009). A substantial fraction of the POA in vehicle emissions has been shown to be semi-volatile under atmospheric conditions (Robinson et al., 2007; May et al., 2013), and is mainly comprised of aliphatic species in the carbon number range between $C_{12}$–$C_{35}$, with effective saturation
concentrations ($C^*$) between 0.1 and $10^3 \mu\text{g m}^{-3}$ (Robinson et al., 2007; Weitkemp et al., 2007).

The semi volatile organic compound (SVOC) composition of lubricating oil has been reported to be dominated by branched, cyclic and straight alkanes ($\geq 80\%$), with the largest contribution from cycloalkanes ($\geq 27\%$) (Sakurai et al., 2003; Worton et al., 2014; Sakurai et al., 2003).

Previous research has used a limited range of tracer compounds, or homologous series, for the quantification of emissions, considering representative species that can be distinguished from the bulk of the mass, typically involving analysis of the n-alkanes, polycyclic aromatic hydrocarbons (PAH), hopanes, and steranes (Schauer et al., 1999; 2002) each of which represent only a small fraction of the total mass or number of compounds emitted and might lead to underestimation of the importance of lubricating oil as a source of SOA (Brandenberger et al., 2005; Fujita et al., 2007).

Rogge et al. (1993) investigated the sources of fine organic aerosol from non-catalyst and catalyst equipped vehicles using one-dimensional GC/MS, but could only resolve 10-15\% of the organics, including n-alkanes and PAH. Although some studies have utilized soft ionization to analyse diesel fuel at a molecular level, (Briker et al., 2001); Eschner et al., 2010; Amirav et al., 2008) very few studies have analysed lubricating oil at a molecular level that includes the analysis of SVOCs (Worton et al., 2015; Reddy et al., 2012). In order to address the problems of coelution of constituents of the UCM, Worton et al. (2014) and Isaacman et al. (2012) utilized gas chromatography coupled with vacuum ultraviolet ionization mass spectrometry (GC/VUV-MS) to study the constitutional isomers present in lubricating oil and diesel fuel, respectively, and in a standard crude oil from the Gulf of Mexico (Worton et al., 2015). More recently, Goodman-Rendall et al. (2016) used GC/MS with cold electron impact (EI) ionization, resolving detailed molecular components of diesel fuel. Their results showed that the most important factors in determining SOA
yields were carbon number, the presence (or absence) of a ring moiety and the degree of substitution; and precise information of branching and degrees of unsaturation was of secondary importance. Dunmore et al. (2015) have shown that diesel-related hydrocarbons are responsible for 60% of the winter primary hydrocarbon hydroxyl radical reactivity and possibly up to 50% of the ozone production potential in London. Detailed chemical characterization of diesel emissions would therefore not only resolve factors in determining the contribution to SOA yields, but also to shed light on specific precursors with large photochemical ozone creation potentials and OH reactivity, as well as the identification of compounds that are harmful for human health and the environment.

As well as identifying individual compounds, using GC×GC allows compounds of similar chemical structure to be classified into distinct groups in ordered chromatograms based on their volatility and their polarity, providing information that aids identification and assessment of environmental fate. Dunmore et al. (2015) recently grouped low molecular weight (≤ C_{12}) hydrocarbons in atmospheric samples by carbon number and functionality using GC×GC. They reported the grouping of C_{6}–C_{13} aliphatics and C_{2}–C_{4} substituted monoaromatics, combining the area of all the peaks contained within their selected areas.

In our study, two dimensional Gas Chromatography Time-of-Flight-Mass Spectrometry (GC×GC-ToF-MS) (Adahchour et al., 2008; Alam et al., 2013; Alam et al., 2016b) was utilised and combined with an innovative quantification methodology based on total ion current (TIC) signal response to provide identification and quantification for the compound classes within typical diesel fuel, engine lubricant and engine emissions (gas and particulate phases), providing a near
complete mass closure for diesel fuel and engine lubricant and analyses of diesel engine exhaust composition.

2. EXPERIMENTAL

2.1 Sampling

Gas and particulate phase diesel exhaust emissions were collected from a light-duty diesel engine. This 2.2 L, 4-cylinder, in-line compression ignition engine was equipped with a common rail direct injection system and a variable-nozzle-turbine (VNT) turbocharger. Samples were collected with no diesel oxidation catalyst (DOC) and diesel particulate filter (DPF). The diesel engine emissions were diluted (1:50) with cleaned compressed air using an in-house exhaust dilution system. Samples were collected at steady state engine operating conditions at a low engine load of 3.0 bar mean effective pressure (BMEP) and a speed of 1800 revolutions per minute (RPM). The temperature at the sampling point was 25 ± 5 °C and samples were collected for 30 min at a flow rate of 1.8 L min⁻¹. Briefly, adsorption tubes were used to collect gas phase constituents directly from the diluted (1:50)-diesel engine exhaust, downstream of a polypropylene backed PTFE 47 mm filter (Whatman, Maidstone, UK), which was used to collect and remove constituents in the particulate phase. Further details of the engine exhaust sampling system are given elsewhere (Alam et al. 2016c).

Diesel fuel, engine lubricating oil, and gas-and-particulate diesel exhaust emission samples were analysed using GC×GC-ToF-MS. Briefly, 1 µL of diesel fuel (EN 590-ultra low sulfur diesel, Shell, UK) was diluted (1:1000) in dichloromethane (DCM) and injected onto a stainless steel thermal adsorption tube, packed with 1 cm quartz wool, 300 mg Carbograph 2TD 40/60 (Markes
International, Llantrisant, UK), for analysis on the thermal desorber (TD) coupled to the GC×GC-ToF-MS. The EN 590 ultra-low sulphur diesel fuel is representative of the standardized ultra-low sulphur content fuel (<10 mg kg$^{-1}$ or ppm) that is widely used utilised in the UK and Europe (Ref: EU directive 2009/30/EC European Parliament and the Council of the European Union, 2009). 1 µL of engine lubricant (fully synthetic, 5W30, Castrol, UK) was diluted (1:1000) in DCM and directly injected into the gas chromatographic column, as the high molecular weight constituents found in the lubricating oils would not efficiently desorb into the GC column from the adsorption tubes.

Details of the engine exhaust sampling system are given elsewhere (Alam et al. 2016c). Briefly, adsorption tubes were used to collect gas phase constituents directly from diluted (1:50) diesel engine exhaust, downstream of a polypropylene backed PTFE 47 mm filter (Whatman, Maidstone, UK) used to collect and remove constituents in the particulate phase. The diesel engine emissions were diluted (1:50) with cleaned compressed air using an in-house exhaust dilution system described elsewhere (Alam et al., 2016c). The temperature at the sampling point was 25 ± 5 °C. Samples were collected for 30 min at a flow rate of 1.8 L min$^{-1}$. Details of the engine set up have been given elsewhere (Alam et al., 2016c). Samples were collected at steady state engine operating conditions at a load of 3.0 bar mean effective pressure (BMEP) and a speed of 1800 revolutions per minute (RPM) without a diesel oxidation catalyst (DOC) and diesel particulate filter (DPF).

### 2.2 Instrumentation

Adsorption tubes were desorbed using TD (Unity 2, Markes International, Llantrisant, UK) and
samples were subsequently analysed using a gas chromatograph (GC, 7890B, Agilent Technologies, Wilmington, DE, USA) equipped with a Zoex ZX2 cryogenic modulator (Houston, TX, USA). The primary column (first separation dimension) was equipped with a SGE DBX5, non-polar capillary column (30 m, 0.25 mm ID, 0.25 μm – 5% phenyl polysilphenylene-siloxane). The secondary, more polar column (second separation dimension) was equipped with a SGE DBX50 (4.0 m, 0.1 mm ID, 0.1 μm – 50% phenyl polysilphenylene-siloxane), situated in a secondary internal oven. The GC×GC was interfaced with a BenchTOF-Select, time-of-flight mass spectrometer (ToF-MS, Markes International, Llantrisant, UK), with a scan speed of 50 Hz and a mass resolution of >1200 full width at half maximum (fwhm) at 70 eV and >800 fwhm at 14eV over 100 – 1000 m/z. The mass/charge range was from 30 to 525 m/z, and quantification was conducted on nominal unit mass resolution. Electron impact ionisation energies on this ToFMS can be tuned between 10 eV and 70 eV, the former retaining the molecular ion, while the latter causes extensive fragmentation, but allows comparison with standard library spectra (Alam et al. 2016a). Data were processed by using GC Image v2.5 (Zoex Corporation, Houston, USA).

2.3 Standards & Chromatography Methodology

Nine deuterated internal standards namely, dodecane-d_{26}, pentadecane-d_{32}, eicosane-d_{42}, pentacosane-d_{52}, triacontane-d_{62}, biphenyl-d_{10}, n-butylbenzene-d_{14}, n-nonylbenzene-2,3,4,5,6-d_{5} (Chiron AS, Norway) and p-terphenyl-d_{14} (Sigma Aldrich, UK) were utilised for quantification. Natural standards included 24 n-alkanes (C_{11} – C_{34}), phytane and pristane (Sigma Aldrich, UK), 16 n-alkylcyclohexanes (C_{11} – C_{25} and C_{26}), 5 n-alkylbenzenes (C_{4}, C_{6}, C_{8}, C_{10} and C_{12}), cis- and trans-decalin, tetralin, 4 alkyltetralins (methyl-, di-, tri- and tetra-), 4 n-alkyl naphthalenes (C_{1}, C_{2}, C_{4} and C_{6}) (Chiron AS, Norway) and 16 USEPA polycyclic aromatic hydrocarbons (Thames Restek
These standards were chosen in order to cover as much of the overall chromatogram as possible. The chromatography methodology (i.e., GC oven temperatures, temperature ramp rates etc.) of the analysis of the adsorption tubes, lubricating oil and gas/particulate phase samples is discussed in Section S1 in the Supplementary Information.

2.4 Grouping of Chromatographically Resolved Compounds

Structurally similar compounds belonging to the same chemical group in a mixture possess similar physicochemical properties. This facilitates identification when separating a mixture according to these physical and chemical properties. Diesel fuels, diesel emissions and lubricating oils have been shown to consist of a limited number of compound classes, but an enormous number of individual components within a class.

In this study we use GC×GC coupled to variable ionisation ToF-MS to map and quantify isomer sets previously unresolved in the UCM. Conventional electron ionisation at 70 eV imparts a large amount of excess energy causing extensive fragmentation, with a tendency to generate similar mass spectra. Thus for example the isomeric alkanes all exhibit the same m/z 43, 57, 71, 85, 99 patterns, thus obscuring the match with the NIST library and making identification from the mass spectrum very difficult. To address this issue, a lower ionization energy (10-14eV) was also employed in our study so that the organic compounds are ionized with minimal excess internal energy and thus less fragmentation, hence retaining the distinct identity of the molecule with a much larger fraction of the molecular ion (Alam et al., 2016a). Running samples on the GC×GC with both low and high ionization energy mass spectrometry results in a wealth of data for identification of compounds.
where 14 and 70eV mass spectra can be compared for a given species owing to the identical retention times of the repeat runs. At low ionization energy the molecular ion is enhanced, while-at the same time as still—retaining some fragmentation, while at high ionization energy the mass fragmentation patterns of a species can be compared directly to mass spectral libraries. This allows easier identification of unknown compounds. Low EI spectra therefore give qualitative information, while high EI mass spectra allow for quantitative analyses to be performed.—

Our recent work exploited soft ionisation (14 eV) to identify a large number of isomers, demonstrating the ability to separate and identify individual alkanes (normal, branched and cyclic) with specific carbon numbers, based on their volatility and polarity (Alam et al., 2016a). In this study we expand our previous qualitative analysis and separate the alkane series (as well as other homologous series) into isomer sets containing the same carbon number. Individual alkanes that were identified as having different molecular ions (i.e. different carbon number) to the n-alkane within the area of the chromatogram were included in their appropriate adjacent (usually n±1) area; for example, some dimethyl isomers can be shifted by ~100 delta-Kovats (~1 carbon number), whereas trimethyl- and tetramethyl isomers have been reported to be shifted by ~150 and ~200 delta-Kovats. This has been completed for all the homologous series reported in this study. The grouping of the alkanes according to their respective carbon numbers is shown in Figure 1, where the least polar compounds (fast eluting peaks in the second dimension) are the alkanes, increasing in carbon number as the retention time in the first dimension increases.

This methodology was expanded to more polar homologous series including monocyclic alkanes, bicyclic alkanes, tricyclic alkanes, tetralins/indanes, monocyclic aromatics, bicyclic aromatics and...
alkyl-biphenyls. Like the alkanes, a significant problem in creating the boundaries of groups is the overlapping of one carbon number group into another. Identifying each individual compound in this case (as with the alkanes above) would be resource and time intensive and so carefully constructed ‘Computer Language for Identification of Chemicals’ (CLIC) qualifiers were created and utilised in order to match peaks and their mass fragmentation patterns. A CLIC qualifier is an expression in a computer language that allows users of chromatographic software to build rules for selecting and filtering peaks using retention times and mass fragmentation patterns (Reichenbach et al., 2005). This was exploited to identify specific compounds belonging to a compound class and a polygon selection tool within the GC Image software was drawn around this section of the chromatogram (coloured polygons shown in Figure 1). Any overlap in the graphics was accounted for by forcing peaks to belong to one compound class over another via strict mass fragment and molecular ion selection tools. Examples of CLIC expressions utilised for identifying compound classes are included in the Supplementary Information (Section S2). An example of a selected ion chromatogram with a specific CLIC expression is shown in Figure 2, for C₆-substituted monocyclic aromatics, with their corresponding 70eV and 12eV mass spectra. The characteristic 70eV mass fragments at m/z 92, 105, 119, 133 signify cleavage of the C-C bond next to the benzene ring. The 12eV mass spectra, however, produce poor characteristic fragment ions, but a prominent molecular ion (162) and m/z 92 signifying the overall mass of the molecule and the benzene ring (Ph-CH₂⁺), respectively. In effect, the polygons mark out sets of isomeric compounds having the same empirical formula and shared structural elements; the sets appear to intersect each other in the two-dimensional chromatogram space, but compounds in the intersecting regions are assigned uniquely to a class using a third, mass spectrometric, data dimension (i.e. stringent—mass fragmentation patterns). The resulting isomer sets are more chemically and environmentally
meaningful than the raw polarity/volatility assignment from the chromatography. This approach was completed independently for diesel fuel, lubricating oil, and gas/particulate phase exhaust emissions to ensure the polygon boundaries applicability of polygon boundaries and reproducibility of retention times and mass fragments. Results indicated that isomers within the constructed polygon boundaries possessed identical retention times and interpretable mass spectra for all differing samples. Retention times were reproducible in both chromatographic dimensions in separate runs. The isomer sets (polygons) can be linked together in a large template that can also be linked to the internal standards. This allows easy alignment of the template (all isomer sets) in the event of slight shifts in retention times, due to column changes or instrumental maintenance.

2.5 Quantification of compounds with no authentic standards

The authentic standard mixture contains 74 standard compounds including 9 internal standards (see Section 2.3). These standards were chosen in order to cover as much of the overall chromatogram as possible and are used for obtaining a calibration for quantifying groups of isomers with the same molecular ion and functionality. For example; the response for C$_{11}$ (undecane, m/z 156) was used to quantify all C$_{11}$ alkane isomers which were positively identified in the analysed samples (and have the same molecular mass and retention times in all sample runs). Using retention times for C$_{11}$ alkane isomers as well as mass spectra, only those C$_{11}$ isomer peaks were selected that were observed using the SIC for m/z 156 (or the respective CLIC expressions). Isomer sets were comprehensively identified in the analysed samples using mass fragmentation at 70eV and molecular masses at lower ionisation energies (10-15eV). The total ion current within each polygon was integrated and the isomer set abundance was estimated using the response ratio of the closest
structurally-related deuterated standard to the corresponding compound class natural standard with the same carbon number (usually within the polygon). This methodology has an uncertainty of approximately 24\% and is discussed in detail in the Supplementary Information (Section S3). The potential of scaling ion current to molar quantity for $<C_{25}$ and mass for compounds $>C_{25}$ is also discussed in the Section S4.

3. RESULTS AND DISCUSSION

3.1 Analysis of Diesel Fuel

The chromatography of the diesel fuel analysed by TD-GC×GC-ToF-MS is shown in Figure 1. Compounds identified within the diesel fuel included: $n$-alkanes, branched alkanes (mono-, di-, tri-, tetra- and penta-methyl), $n$-alkyl cycloalkanes, branched monocyclic alkanes, $C_{1}$-$C_{12}$ substituted bicyclic alkanes, $C_{1}$-$C_{4}$ substituted tetralins and indanes, $C_{3}$-$C_{12}$ substituted monocyclic aromatics, $C_{1}$-$C_{3}$ substituted biphenyls/acenaphthenes, $C_{1}$-$C_{4}$ substituted bicyclic aromatics, $C_{1}$-$C_{2}$ substituted fluorenes (FLU), $C_{1}$-$C_{2}$ substituted phenanthrene/anthracenes (PHE/ANT) and unsubstituted PAH. Representative mass spectra at 12eV and 70eV ionization are presented in the Supplementary Information (Section S5S4). These compounds accounted for 93\% of the total response (excluding the siloxanes which derive from contaminants i.e. column bleed) which-and was equivalent to 90 \% of the mass injected. Therefore, out of the 8026 (± 24) ng that was injected into the GC×GC (mass calculated by weighing and diluting a known volume of diesel fuel, while the uncertainty represents the combined uncertainties of the processes involved in this estimation) a mass of approximately 7200 (± 1728) ng was accounted for. We suspect that a significant amount of the mass that was unaccounted may be $<C_{10}$ and/or any unresolved peaks that we were unable to measure and/or
identify using our technique (see Supplementary Information, Section 6). The percentage contribution of each compound class identified to the total mass accounted for is shown in Table 1.

Our results indicate that the majority of the diesel fuel consists of aliphatic compounds, with a low aromatic content (~10%). Very few published studies exist elucidating the contribution of different constituents in diesel fuel (Isaacman et al., 2012; Welthagen et al., 2007; Gentner et al., 2012). Most studies focus on the characterisation of specific compounds within diesel fuels such as nitrogen containing species (Wang et al. 2004), cyclic compounds (Edam et al., 2005), or to identify strengths and weaknesses in analytical techniques (Frysinger et al., 1999). Recently, VUV ionization at 10 – 10.5 eV has been exploited to elucidate some of the structures within diesel fuel, by separating components using GC (Isaacman et al., 2012). The authors report their observed mass of diesel fuel as 73% aliphatic and 27% aromatic, broadly consistent with the results of this study. Up to 11% of the observed mass fraction of diesel fuel was attributed to bicyclic alkanes, a factor of 2 larger than observed in this study. Their observed mass fractions of cycloalkanes and benzene, however, are in excellent agreement. The contribution of branched alkanes (i-alkanes) and linear n-alkanes to the total mass of the alkanes was 39.1 and 23.1%, respectively. A significant proportion of the total mass observed was attributed to alkanes (62%), a factor of 1.5 larger than reported by Isaacman et al. (2012). However, the differences observed between diesel fuel analysed in this study and that reported by Isaacman et al. (2012) is attributable to different fuel formulations and/or fuel source, as opposed to analytical methods. Although not shown here, a significant number of alkane isomers were identified for each carbon number using soft ionisation mass spectrometry, accounting for a total of ~200 alkanes across the C_{11} – C_{30} range. The ratio of
\( \text{i-alkanes to } n\text{-alkanes sharply decreases after } C_{25}, \text{ indicating a reduced amount of mass represented by branched isomers present in diesel fuel for } >C_{25}\text{ alkanes, which could be related to the formulation process, or reflect the composition of the feedstock.} \)

3.2 Analysis of diesel engine emissions (gas phase)

A GC×GC contour plot of the gas phase diesel exhaust emissions is shown in the Supplementary Information (Figure S7S5). The observed chromatogram for the gas phase emissions looked extremely similar to that of the diesel fuel chromatogram (Figure 1), suggesting that the majority of compounds found in the gas phase emissions are of diesel fuel origin. All of the compounds found in the diesel fuel were observed in the gas phase emissions, albeit with a reduced number of \( \text{i-alkanes } >C_{20}, \) which may signify efficient combustion of these high molecular weight compounds, or partitioning into the particulate phase. The measured constituents of the gas phase diesel exhaust emissions are shown in Table 1. Approximately 15\% of the total ion current (response, excluding siloxanes) was unaccounted for. Table 1 illustrates the percentage mass of each compound class identified, in the 85\% of the response that was accounted for. As the total mass of the gas phase sample is unknown, a mass for the remaining 15\% of the total response cannot be estimated, as the individual components that are unidentified will have different responses per unit mass. For example, 23.5\% of the mass identified was attributed to \( C_4 – C_{12}\) alkyl substituted monocyclic aromatics and accounted for 9.7\% of the total ion current response; and 10.0\% of the mass identified was bicyclic alkanes, representing 9.2\% of the response.

Although the diesel fuel constituents present in the gas phase exhaust emissions broadly were compositionally consistent with the fuel, there were significant differences observed in their relative
amounts. Of the total mass identified in the gas phase emissions, \( n \)-alkanes and \( i \)-alkanes represented 9.8% and 30.1%, respectively. These are factors of 2.4 and 1.3 lower than that for diesel fuel, respectively; which may be due to preferred combustion of these compounds (Burcat et al., 2012). Enhancements of monocyclic aromatics, monocyclic alkanes, bicyclic alkanes and bicyclic aromatics were observed in the emissions, possibly due to them being intermediate species formed during the combustion of larger molecules (Gentner et al., 2013), and unlikely to be a contribution from lubricating oil as very little mass was attributed to compounds with \(<C_{18}\) (see Section 3.3), in contrast to previous studies (Worton et al., 2014). A very limited number of oxygenates were also identified \( (e.g. \) ketones \( m/z \) 58, 72), carboxylic acids \( (m/z \) 60)), most probably combustion products of diesel fuel, but representing a very small fraction of the total measured gas phase emissions \(<1\%). However, carboxylic acids are difficult to detect using GC/MS without prior derivatization and may therefore be underestimated. Gentner et al. (2013) suggest that compounds such as alkenes, aromatics and oxygenates comprise \(~30\%\) of the total measured gas phase emissions, in agreement with this study; however, they suggest that these products are unlikely to contribute to primary organic aerosol (POA). We observe these aromatic compounds in the particulate phase also, indicating a contribution.

### 3.3 Analysis of Lubricating Oil

A comprehensive analysis of base oil is presented elsewhere (Alam et al, 2016a) and a comparison of different aged and fresh oils is reported in Liang et al. (2017 in prep). A brief account of analysis of 5W30 synthetic lubricating oil is presented in this study. The analysis was conducted in exactly the same way as the diesel fuel and gas phase diesel emissions as outlined in Section 2.4. In brief, analysis was conducted using 12eV electron impact ionization energy mass spectrometry in order to
retain the molecular ion for the compounds analysed. This enabled the clear identification of specific compounds with different carbon numbers from within a compound class. The isomer sets (polygons), identified in the lubricating oil chromatogram, are shown in Figure 3A. Molecular ions present in the mass spectra enabled the grouping of isomers by carbon number, while the presence of the characteristic mass fragments, presented in Table 1, were used to confirm the identity of the type of hydrocarbon, see Supplementary Information, Section S8, for the representative mass spectra for compounds presented in Table 1, for the lubricating oil is shown in the Supplementary Information, Section S6. Polygons were drawn around groups of compounds that possessed the same molecular ion for a given compound class, see Figure 3A and 3B. The lubricating oil was analysed using two independent temperature ramps of the GCxGC (methodologies outlined above); one to achieve the best possible comprehensive separation of compounds in the oil (Figure 3A) and the other using methodologies developed for analysis of the particulate phase components of engine exhaust (Figure 3B), to ascertain where the compounds identified in the oil are present in the particulate phase emissions filter. Figure 3B also illustrates the positioning of the SVOC measured in the gas phase, that are observed in the particulate phase filter as well as the positioning for the PAH. The grouping template that is illustrated in Figure 1 covers the SVOC range indicated in Figure 3B.

Using the signature mass fragment ions (Table 1) together with the calculated molecular mass, specific compounds with the same carbon number were isolated, see Figure S9S7-A and S9S7-B. For example, selecting the ion fragment m/z 92 and 119 for monocyclic aromatics gives rise to the selected ion chromatogram illustrated in Figure S9S7-A. This can be achieved using 70eV mass
spectrometry identifying a homologous series across a large carbon number range. However, selecting the molecular mass for a specific carbon number allows the identification of all isomers in a region of the chromatogram with that specific molecular mass, as shown in Figure S9B for C_{22} monocyclic aromatics (m/z 302). A mass of 8511 \(^{\pm}255\) ng of lubricating oil was injected into the GC×GC, of which 6356 (±1525) ng was quantified. This methodology was used to identify and quantify the following homologous series: C_{16}–C_{33} straight and branched chain alkanes, C_{16}–C_{33} monocyclic alkanes, C_{17}–C_{33} bicyclic alkanes, C_{17}–C_{33} tricyclic alkanes and C_{16}–C_{33} monocyclic aromatics. These compound groups represented approximately 91\% of the total ion current (excluding siloxanes) and 75\% of the mass fraction. Adamantanes, diamantanes, pentacyclic and hexacyclic alkanes, steroids, steranes, and hopanes represented 5\% of the total ion current, while the remaining 4\% remained unidentified. These compounds were not quantifiable using this methodology, as there were no standards available that corresponded to these sections of the chromatography and could not be estimated as they are not present in a homologous series. However, from previous literature, these compound classes are thought to represent a small fraction of the mass (Worton et al., 2015). Furthermore, we have not taken into account any non-organic/hydrocarbon species but according to these data, the fraction of any non-HCO material is likely to be small.

Worton et al. (2015) exploited VUV photoionization mass spectrometry to characterize comprehensively hydrocarbons in a standard reference crude oil sample. They reported a total mass closure of 68 ± 22\%, comprised of linear and branched alkanes (19\%), 1-6 ring cycloalkanes (37\%), monoaromatics (6.8\%) and PAH (4.7\%). The mass fractions observed for linear and branched
alkanes in this study were 11% and 12%, respectively, which is in excellent agreement. There is also excellent agreement with the mass attributed to bicyclic (2-ring) and tricyclic (3-ring) alkanes however, for monocyclic alkanes the results presented here are a factor of 2 larger than Worton et al.\textsuperscript{14}(2015) and 2.5 larger than Reddy et al. (2012). Both previous studies analysed similar crude oil samples associated with the Deepwater Horizon disaster (McNutt et al., 2012), and would be expected to differ appreciably from a lubricating oil. Furthermore, no PAH were observed in the lubricating oil in this study, in agreement with Zielinska et al. (2004) but in contrast to Worton et al. (2015). We attribute this difference to the varying crude oil origins and formulation processes involved in the production of synthetic oil.

Previous work from this group identified a large number of isomeric species in base oil using 14eV EI ionization energy mass spectrometry (Alam et al., 2016a). Although we were able to identify a large number of compounds, there still existed a small amount of fragmentation at 14eV, particularly for alkyl-methyl-, alkyl-dimethyl-, and alkyl-trimethyl-cyclohexanes. In this study the fragmentation was significantly reduced for these compounds at 12eV (i.e. relative intensities of m/z 97, 111, 125 reduced by >50%) and completely eradicated (relative intensities of mass fragments reduced by >95%) at 10eV, leaving the m/z 82 ion (for monocyclic alkanes) and the molecular ion. This demonstrates the significant differences observed in fragmentation over small changes in lower ionisation EI energies and may also account for slight discrepancies between studies (Worton et al., 2015; Isaacman et al., 2012; Alam et al., 2016b). Utilising these differences in fragmentation from using low ionization energies (10 – 15eV) may provide more information in regards to the structure of many isomeric compounds.
3.4 Analysis of Diesel Engine Emissions (Particulate Phase)

90% of the total ion current of the particulate phase filter was identified and attributed to a wide range of classes. Of the total mass identified, 47 (±11)% was straight and branched chain alkanes, 20 (±4.8)% monocyclic alkanes, 7.5 (±1.8)% bicyclic alkanes, <3 (±0.7)% tricyclic alkanes, 6 (±1.4)% monocyclic aromatics, 7 (±1.7)% oxygenates, <1 (±0.2)% furanones, 4 (±1.0)% PAH and 2 (±0.5)% fatty acid methyl esters (FAMES). Figure 4 illustrates the percentage mass contribution of homologous series (including isomers) identified as a function of carbon number. Peak concentrations of alkanes (cyclic and straight/branched chain) were observed between C_{24} – C_{27} consistent with the lubricating oil, while a small peak in concentration was also observed in the C_{15} – C_{20} range, consistent with the fuel and gas phase emissions. Oxygenated compounds were found to be present in the C_{11} – C_{22} range, suggesting that these compounds are combustion products. The concentration of monocyclic aromatics was steady throughout the carbon number distribution (C_{15} – C_{32}), with a small peak at C_{25} – C_{27}. The presence of PAH in the particulate phase suggests their formation via diesel fuel combustion or unburnt fuel, owing to their absence in the lubricating oil.

There are numerous studies reporting the absence of PAH in unused lubricating oil and presence in used oil, which suggests the absorption of blow-by exhaust containing fuel combustion associated PAH (Fujita et al., 2006). FAMES were identified by their characteristic fragmentation at 70eV EI ionization and with their characteristic ion (m/z 174) at low EI ionization (12 eV).

There have been few studies investigating the contribution of lubricating oil and fuel to the emitted diesel POA, suggesting 20 to 80% influence from lubricating oil (Worton et al., 2014; Brandenberger et al., 2005; Kleeman et al., 2008; Sonntag et al., 2012). Most recently, it has been
suggested that ≥80% of the SVOC composition is dominated by branched cycloalkanes with one or more rings and one or more branched alkyl side chain (Sakurai-Worton et al., 2003-2014). This is significantly larger than that reported in this study (≥30%), where the majority of the emissions are dominated by straight and branched chain alkanes (47%) over a volatility range that also suggests a significant contribution from the diesel fuel (C_{11} – C_{20}, see Figure 4). The diesel fuel and lubricating oil contained respectively 62% and 47.5% straight and branched chain alkanes (summed), suggesting a larger possible contribution of diesel fuel to the vapour phase engine emissions (which is dominated by straight and branched chain alkanes). The contribution of unburned lubricating oil, however, most likely dominates the SVOC emissions in the particulate phase, as shown in Figure 4.

5. CONCLUSION

The SVOC contents in diesel fuel, 5W30 synthetic lubricating oil, and diesel exhaust emissions (both in the gas and particulate phases) were characterized using TD-GC×GC-ToFMS. By exploiting the mass spectrometric fingerprint of eluting compounds in highly structured and ordered chromatograms, a methodology has been constructed in order to quantify the contributions of ‘isomer sets’ (i.e. structural isomers in specific compound classes) to the overall composition of a sample. We found that the ion current for identified homologous series exhibited very similar responses, illustrating that quantitative calibrations derived from the n-alkane series could be used to estimate the concentrations of isomeric aliphatic compounds with similar molecular weight. Using this methodology together with a range of standards, and aggregating compound classes of similar functionality together (i.e. n-alkanes, branched alkanes etc.), we present comprehensive characterization of diesel fuel, lubricating oil and diesel exhaust emissions.
Furthermore, combining conventional 70eV EI ionization mass spectrometry with lower ionization energy (10-14eV), allowed the identification of constitutional isomers of the same molecular weight and compound class, enabling a clear distinction between carbon number and functionality. By utilising this innovative method, a number of findings were achieved: 1) a mass closure accountings for ca. 90% and 75% for the analysis of diesel fuel and lubricating oil, respectively. 2) A cyclic and monocyclic alkanes were found to be predominant in both the diesel fuel and synthetic lubricating oil (76% and 68%, respectively). 3) Diesel exhaust emissions in the gas phase were extremely similar to the composition of diesel fuel. 4) Diesel exhaust emissions in the particulate phase were similar to the composition of lubricating oil. 5) The presence of combustion products of diesel fuel (e.g. aromatics and oxygenates) in the particulate phase indicates a contribution to POA.

Diesel exhaust hydrocarbons are a significant precursor of secondary organic aerosol (Dunmore et al., 2015; Gentner et al., 2012). Diesel fuel and lubricant, contributors to diesel exhaust, contain large numbers of isomers. Separation into isomer sets is potentially a key step forward in improves our understanding of the fates of these oil-derived materials in the environment (Lim & Ziemann, 2009; Kroll & Seinfeld, 2008). By utilizing GC×GC-ToFMS with soft ionisation, we enable the identification of the composition of the UCM, characterising the chemical composition by carbon number and compound class, and the possibility of branching structural information. Along with a grouping methodology using CLIC expressions and unique compound fragmentation patterns, we demonstrate the reliable quantitative integration of structural isomers. These methods exploit the improved resolution and isomer separation capabilities of the advanced instrumentation and have
potential applications to the observations of petroleum degradation, and SOA formation and
evolution. This method is ideal for investigating fossil fuel sources (e.g. lubricating oil, motor
vehicles, fuel etc.) and can be extended to atmospheric measurements where there exist many
oxygenates. The analysis of atmospheric particulate phase samples (not using thermal desorption),
may benefit from exploring different extraction solvents (e.g. methanol or a hexane/methanol mix)
to allow the larger expected proportion of oxygenated species to be efficiently extracted. Although
the presence of oxygenated species adds chromatographic complexity, as co-elution
can be a limitation, using carefully constructed CLIC expressions and mass fragmentation patterns,
various oxygenates can be identified (e.g. 2-ketones (m/z 58), 3-ketones (m/z 72), carboxylic acids
(m/z 60) etc.), making this technique applicable. It also has application in any scientific field that
routinely characterizes complex hydrocarbon mixtures, e.g., atmospheric chemistry, microbial and
chemical ecology, bioremediation, and aquatic pollution. Although not explored here, this technique
of low ionization mass spectrometry may benefit from reducing the temperature of the ionisation
source, as demonstrated by Isaacman et al. (2012); who retained a larger fraction of the molecular
parent ion on reducing the ion source temperature of their VUV photoionization technique. This
would particularly aid identification for atmospheric samples which would contain more
oxygenated compounds.

ACKNOWLEDGEMENTS

The authors would like to thank L. McGregor and N. Bukowski on valuable discussions on aspects
of GCxGC-ToFMS and assistance in CLIC qualifiers. Financial support from the European
Research Council (ERC-2012-AdG, Proposal No. 320821) for the FASTER project is gratefully acknowledged. The assistance of Y. Al-Qahtani in collecting samples from the engine facility is also acknowledged.

Author Contribution —needs to be completed

MSA prepared the manuscript with contributions from ARMK and RMH; RMH, MSA and SZR designed the engine experiments; SZR, MSA and ZL carried out the engine experiments; MSA and CS developed the GC×GC methodology and completed subsequent analyses. HX overlooked the engine facility and RMH oversaw the entire project.

Conflict of Interests

The authors declare no competing financial interest.
REFERENCES


Alam, M. S. & Harrison, R. M.: Recent advances in the application of 2-dimensional gas chromatography with soft and hard ionisation time-of-flight mass spectrometry in environmental analysis, Chemical Science, 7, 3968-3977, 2016b.


Brandenberger, S., Mohr, M., Grob, K., & Neukom, H. P. Contribution of unburned lubricating oil and diesel fuel to particulate emission from passenger cars, Atmospheric Environment, 39, 6985-6994, 2005.


Worton, D. R., Isaacman, G., Gentner, D. R., Dallmann, T. R., Chan, A. W., Ruehl, C., Kirchstetter,


TABLE LEGEND:

| Table 1. | Hydrocarbons identified in diesel fuel, lubricating oil and diesel emissions (gas and particulate phases) with their respective m/z fragment ions and percentage mass contributions. |

FIGURE LEGENDS:

- **Figure 1.** A contour plot (chromatogram with 70eV ionisation mass spectrometry) of diesel fuel separation. Peak height (intensity) increases with warmth (blue to red) of the colour scale. Each region fenced by a coloured polygon marks out the 2-dimensional chromatogram space in which are found isomers of a particular compound type having a particular carbon number (e.g. C₄–substituted monocyclic aromatics).

- **Figure 2.** A contour plot (chromatogram with 12eV ionisation mass spectrometry) displaying C₆-substituted monocyclic aromatics identified by the CLIC expression. All C₆-substituted monocyclic aromatics are located within the pink polygon displayed. 70eV (red peaks) and 12eV (blue peaks) mass spectra corresponding to the peaks identified by the CLIC expressions in the SIC is shown for 6 different C₆-substituted monocyclic aromatics isomers.

- **Figure 3.** A chromatogram (with 70eV ionisation mass spectrometry) of lubricating oil (5W30) (A) with labelled compound classes, using a methodology specific for characterising the composition of lubricating oil; (B) using methodology developed for characterising particulate phase emissions from diesel engine exhaust. Polygons drawn around sections of the chromatograms indicate compounds with different molecular masses within compound classes.

- **Figure 4.** Percentage mass contribution of the compounds identified in homologous series as a function of carbon number in diesel exhaust particles.
Table 1. Hydrocarbons identified in diesel fuel, lubricating oil and diesel emissions (gas and particulate phases) with their respective m/z fragment ions and percentage mass contributions.

<table>
<thead>
<tr>
<th>Compound Class</th>
<th>m/z (M⁺)*</th>
<th>% Mass closure</th>
<th>% Contribution to mass identified in emissions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Diesel Fuel</td>
<td>Lubricating Oil</td>
</tr>
<tr>
<td>Total</td>
<td>89.7 74.7</td>
<td>85.0 of TIC</td>
<td>75.0 of TIC</td>
</tr>
<tr>
<td>n + i-Alkanes (C₁₁ – C₃₃)</td>
<td>57 (CₙH₂ₙ₊₂)</td>
<td>62.2</td>
<td>23.0</td>
</tr>
<tr>
<td>Monocyclic Alkanes (C₁₁ – C₃₃)</td>
<td>82 (CₙH₂ₙ)</td>
<td>13.8</td>
<td>35.6</td>
</tr>
<tr>
<td>Bicyclic Alkanes (C₁₁ – C₃₃)</td>
<td>137 (CₙH₂ₙ₊₂)</td>
<td>5.0</td>
<td>9.2</td>
</tr>
<tr>
<td>Tricyclic Alkanes (C₁₁ – C₃₃)</td>
<td>191 (CₙH₂ₙ₋₄)</td>
<td>&lt;0.1</td>
<td>2.7</td>
</tr>
<tr>
<td>Monocyclic Aromatics (C₁₁ – C₃₃)</td>
<td>92, 119 (CₙH₂ₙ₋₆)</td>
<td>4.4</td>
<td>4.2</td>
</tr>
<tr>
<td>Bicyclic Aromatics (C₁₁ – C₃₃)</td>
<td>128, 141</td>
<td>0.8</td>
<td>&lt;0.1</td>
</tr>
<tr>
<td>adamantanes</td>
<td>135, 149, 163, 177</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diamantanes</td>
<td>187, 188, 201, 215, 229</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pentacyclic Alkanes</td>
<td>258, 272, 286</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hexacyclic Alkanes</td>
<td>298, 312</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Steroids</td>
<td>239, 267</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monoaromatic Steranes</td>
<td>253</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Steranes</td>
<td>217, 218</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methyl Steranes</td>
<td>217, 218, 231, 232</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25-norhopanes</td>
<td>177</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28, 30-norhopanes</td>
<td>163, 191</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hopanes</td>
<td>336</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAHs</td>
<td>0.8</td>
<td>0.6</td>
<td>4.0</td>
</tr>
<tr>
<td>Biphenyls / Acenaphthenes</td>
<td>0.1</td>
<td>0.1</td>
<td>&lt;0.1</td>
</tr>
<tr>
<td>Tetratin / Indanes</td>
<td>132, 145</td>
<td>2.5</td>
<td>6.9</td>
</tr>
<tr>
<td>Oxygenates</td>
<td>7.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Furanones</td>
<td>84</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>FAMEs</td>
<td>174</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>Miscellaneous Compounds</td>
<td>&lt;3.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*m/z ratios presented here are the main mass fragments present in the low ionisation energy mass spectra. CLIC expressions and 70eV mass spectra use more m/z fragments which were also used for qualification and quantification.*
Figure 1. A contour plot (chromatogram with 70eV ionisation mass spectrometry) of diesel fuel separation. Peak height (intensity) increases with warmth (blue to red) of the colour scale. Each region fenced by a coloured polygon marks out the 2-dimensional chromatogram space in which are found isomers of a particular compound type having a particular carbon number (e.g. C₄-substituted monocyclic aromatics).
Figure 2. A contour plot (chromatogram with 12eV ionisation mass spectrometry) displaying C₆-substituted monocyclic aromatics identified by the CLIC expression. All C₆-substituted monocyclic aromatics are located within the pink polygon displayed. 70eV (red peaks) and 12eV (blue peaks) mass spectra corresponding to the peaks identified by the CLIC expressions in the SIC is shown for 6 different C₆-substituted monocyclic aromatics isomers.
Figure 3. A chromatogram (with 70eV ionisation mass spectrometry) of lubricating oil (5W30) (A) with labelled compound classes, using a methodology specific for characterising the composition of lubricating oil; (B) using methodology developed for characterising particulate phase emissions from diesel engine exhaust. Polygons drawn around sections of the chromatograms indicate compounds with different molecular masses within compound classes.
Figure 4. Percentage mass contribution of the compounds identified in homologous series as a function of carbon number in diesel exhaust particles.
SUPPLEMENTARY INFORMATION

Mapping and quantifying isomer sets of hydrocarbons (≥C_{12}) in diesel exhaust, fuel, lubricating oil and diesel exhaust fuel samples using GC×GC-ToFMS

Mohammed S. Alam, Soheil Zeraati-Rezaei, Zhirong Liang, Soheil-Zeraati-Rezaei, Christopher Stark, Hongming Xu, A. Rob MacKenzie and Roy M. Harrison

CONTENTS

Section S1 – Chromatographic Methodology

Section S2 – Computer language for identification of chemicals (CLIC qualifiers)

Section S3 – Uncertainty of Quantification of compounds with no authentic standards

Section S4 — Scaling ion current to molar quantity for <C_{25} and mass for >C_{25}

Section S5 — Representative Mass Spectra for diesel and gas phase hydrocarbon identification

Section S6 — Unresolved peaks

Figure S7 S5

Section S8—S6 — Representative mass spectra for compound classes identified in the lubricating oil

Figure S9 S7
Exposed adsorption tubes were spiked with 1 ng of deuterated internal standard mix for quantification, and desorbed onto the cold trap at 350 °C for 15 min (trap held at 20 °C). The trap was then purged onto the column in a split ratio of 102:1 at 350 °C and held for 4 min. An initial temperature of 90 °C of the primary oven was held for 2 min and then increased at 2 °C min\(^{-1}\) to 240 °C, followed by 3 °C min\(^{-1}\) to 310 °C and held for 5 min. The initial temperature of the secondary oven of 40 °C was held for 2 min, and then increased to 250 °C by 3 °C min\(^{-1}\), followed by an increase of 1.5 °C min\(^{-1}\) to 315 °C and held for 5 min. The modulation period was 10 s. The complete run time was approximately 105 min.

Diluted lubricating oil samples and particulate phase filters were spiked with 50 µL of 1 ng µL\(^{-1}\) deuterated internal standard mix for quantification. Filters were immersed in dichloromethane (DCM) and ultrasonicated at 20 °C for 20 min. DCM was chosen as an extraction solvent as this was found to have a greater extraction efficiency over hexane and methanol in extracting the large range of polar and non-polar compounds of interest. The extract was then concentrated to 50 µL under a gentle flow of N\(_2\) for analysis on the GC×GC-ToF-MS. 1 µL of the extracted sample was injected in a split ratio 100:1 at 300 °C. An initial temperature of 120 °C of the primary oven was held for 2 min and then increased at 2 °C min\(^{-1}\) to 210 °C, followed by 1.5 °C min\(^{-1}\) to 325 °C and held for 5 min. The initial temperature of the secondary oven of 120 °C was held for 2 min, and then increased to 200 °C by 3 °C min\(^{-1}\), followed by an increase of 2 °C min\(^{-1}\) to 300 °C and a final increase of 1 °C min\(^{-1}\) to 330 °C, ensuring all species pass through the column efficiently. The modulation period was 13 s.
1 µL of diluted (1:1000) lubricating oil was injected in a split ratio 100:1 at 300 °C. An initial temperature of the primary and secondary ovens were kept the same (175 °C) held for 5 min. The primary oven temperature was increased by 1°C min\(^{-1}\) to 325°C, while the secondary oven temperature was increased by 1°C min\(^{-1}\) to 330°C. A modulation time of 8 s was used, while a total run time of each sample was 120 min. The transfer line and ion source temperatures were 325 °C and 320 °C, respectively and were kept consistent for all sample analyses. Helium was used as the carrier gas at a constant flow rate of 1 mL min\(^{-1}\).

Section S2: Computer Language for Identification of Chemicals (CLIC qualifiers)

**Language functions:** See Reference 2 (below) for more details on selection language functions.

**Ordinal** – Returns the ordinal position of the indicated channel (m/z in a mass spectrum) in the intensity-ordered multi-channel array of the current object (blob)

**Retention** – Returns the retention time of the current object (blob) with respect to the chromatographic column indicated by the dimension parameter. Retention time for dimension 1 is expressed in minutes and dimension 2 is expressed in seconds.

**Relative** – Returns the intensity value of the indicated channel (m/z in a mass spectrum) in the multi-channel intensity array of the current object (blob) as a relative percentage of the largest intensity value of the array.

See Reference 2 (below) Reichenbach et al. (2005) for more details on selection language functions.

Some examples of CLIC qualifiers are shown below:
C2-Alkyl Benzenes
=(ordinal(91)=1)&(ordinal(106)=2);

C3-Alkyl Benzenes
=(ordinal(91)=1)&(ordinal(120)<=3)&(retention(1)<30))|(ordinal(105)=1)&(ordinal(120)<=3)&(retention(1)<30));

C4-Alkyl
Benzenes=((ordinal(91)=1)&(ordinal(134)<=3))|(ordinal(105)=1)&(ordinal(134)<=3)&(retention(1)<80))|(ordinal(119)=1)&(ordinal(134)<=3));

C5-Alkyl
Benzenes=((ordinal(92)=1)&(ordinal(148)<=3))|(ordinal(105)=1)&(ordinal(148)<=4))|(ordinal(119)=1)&(ordinal(148)<=4))|(ordinal(133)=1)&(ordinal(148)<=3));

C6-Alkyl
Benzenes=((ordinal(92)=1)&(ordinal(162)<=4))|(ordinal(105)=1)&(ordinal(162)<=5))|(ordinal(119)=1)&(ordinal(162)<=5))|(ordinal(133)=1)&(ordinal(162)<=5))|(ordinal(147)<=2)&(ordinal(162)<=5)&(retention(2)<6))|(ordinal(106)=1)&(ordinal(162)<=4))|(ordinal(91)=1)&(ordinal(162)<=4));

C7-Alkyl
Benzenes=((ordinal(105)=1)&(ordinal(176)<=4))|(ordinal(133)=1)&(ordinal(176)<=4))|(ordinal(147)=1)&(ordinal(176)<=4)&(Retention(2)<5.5))|(ordinal(141)=1)&(ordinal(176)<=3))|(ordinal(119)=1)&(ordinal(176)<=5));

C1-Alkyl NAP=(ordinal(141)<=3)&(ordinal(115)<=3)&(ordinal(142)<=3);

C2-Alkyl
NAP=((ordinal(141)=1)&(ordinal(156)=2)&(ordinal(115)=3)&(relative(141)>90)&(relative(156)>50))|(ordinal(156)=1)&(ordinal(141)=2)&(ordinal(155)=3)&(relative(156)>90)&(relative(141)>50))

C3-Alkyl NAP
=|(ordinal(155)=1)&(ordinal(170)=2)&(Relative(155)>80)&(Relative(170)>25))|(ordinal(170)=1)&(ordinal(155)=2)&(Relative(170)>80)&(Relative(155)>60))|(ordinal(141)=1)&(ordinal(170)=2)&(ordinal(115)=3)&(Relative(141)>80)&(Relative(170)>25)&(Relative(115)>15));

C4-Alkyl
NAP=((ordinal(155)=1)&(ordinal(184)=2)&(Relative(155)>80)&(Relative(184)>15))|(ordinal(141)=1)&(ordinal(184)=2)&(Relative(141)>80)&(Relative(184)>15))|(ordinal(169)=1)&(ordinal(184)=2)&(Relative(169)>80)&(Relative(184)>30))|(ordinal(184)=1)&(ordinal(169)=2)&(Relative(184)>80)&(Relative(169)>50));
Section S3: Uncertainty of quantification of compounds with no authentic standards

The authentic standard mixture contains 74 standard compounds including 9 internal standards (see Section 2.3 for more details). These standards were chosen in order to cover as much of the overall chromatogram as possible and are used for obtaining a calibration for quantifying groups of isomers with the same molecular ion and functionality. For example, the response for C_{11} (undecane, m/z 156) was used to quantify all C_{11}-alkane isomers which were positively identified in the analysed samples (and have the same molecular mass and retention times in all sample runs). Using retention times for C_{11}-alkane isomers as well as mass spectra, only those C_{11}-isomer peaks were selected that were observed using the SIC for m/z 156 (or the respective CLIC expressions). Isomer sets were comprehensively identified in the analysed samples using mass fragmentation at 70eV and molecular masses at lower ionisation energies (10-15eV); the methodology of identifying individual isomers has been published elsewhere (Alam et al. 2016). The total ion current within each polygon (C_{11}-isomer sets in this case) was then integrated and the isomer set abundance was estimated using the response ratio of the closest structurally related deuterated standard (dodecane-d_{26}) to the corresponding compound class natural standard with the same carbon number (C_{11} in this case).

To assess the uncertainty of this method, we estimated known concentrations of compounds for which authentic standards were available. Table S3-1 shows the difference between concentrations estimated with the generic standard and the authentic standard. The overall uncertainty is difficult to estimate. This depends upon both the uncertainty associated with quantifying an individual compound (U), and the number of compounds in a polygon (n). Then

\[ U(\text{polygon}) = (U_1^2 + U_2^2 + U_3^2 \ldots U_n^2)^{1/2} \]

where \( U_1, U_2 \ldots U_n \) are the uncertainties associated with individual compounds. There were
three polygons for which calibration standards were available for all compounds. These gave collective uncertainties calculated as above of 13.9%, 18.9% and 39.3% (mean = 24%). Although statistically rigorous, we feel that this overestimates the uncertainty as the mass closure figures for the samples as a whole appear realistic and none deviates appreciably from 75-100%, including samples not described in this paper.

Table S3-1. Comparison of true calibrated concentrations and estimated concentrations using this methodology.

<table>
<thead>
<tr>
<th>Compound</th>
<th>True Calibrated Concentrations in sample (ng/µL)</th>
<th>Concentration using TIC and n-alkane calibration (ng/µL)</th>
<th>% difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pristane</td>
<td>9.9</td>
<td>9.3</td>
<td>-6%</td>
</tr>
<tr>
<td>Phytane</td>
<td>10.0</td>
<td>10.4</td>
<td>4%</td>
</tr>
<tr>
<td>Cyclohexane, pentyl-</td>
<td>3.1</td>
<td>3.3</td>
<td>6%</td>
</tr>
<tr>
<td>Cyclohexane, hexyl-</td>
<td>15.2</td>
<td>18.1</td>
<td>19%</td>
</tr>
<tr>
<td>Cyclohexane, heptyl-</td>
<td>14.0</td>
<td>16.5</td>
<td>18%</td>
</tr>
<tr>
<td>Cyclohexane, octyl-</td>
<td>11.1</td>
<td>12.5</td>
<td>13%</td>
</tr>
<tr>
<td>Cyclohexane, nonyl-</td>
<td>8.9</td>
<td>10.0</td>
<td>12%</td>
</tr>
<tr>
<td>Cyclohexane, decyl-</td>
<td>10.0</td>
<td>9.5</td>
<td>-5%</td>
</tr>
<tr>
<td>Cyclohexane, undecyl-</td>
<td>3.8</td>
<td>4.0</td>
<td>5%</td>
</tr>
<tr>
<td>Cyclohexane, dodecyl-</td>
<td>4.1</td>
<td>3.8</td>
<td>-7%</td>
</tr>
<tr>
<td>Cyclohexane, tridecyl-</td>
<td>3.7</td>
<td>3.7</td>
<td>0%</td>
</tr>
<tr>
<td>Cyclohexane, tetradecyl-</td>
<td>3.1</td>
<td>2.5</td>
<td>-19%</td>
</tr>
<tr>
<td>Cyclohexane, pentadecyl-</td>
<td>3.0</td>
<td>2.8</td>
<td>-7%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Compound</th>
<th>True Calibrated Concentrations in sample (ng/µL)</th>
<th>Concentration using TIC and n-alkylcyclohexane calibration (ng/µL)</th>
<th>% difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>naphthalene, 1-methyl-</td>
<td>20.0</td>
<td>24.0</td>
<td>20%</td>
</tr>
<tr>
<td>naphthalene, 1-ethyl-</td>
<td>15.0</td>
<td>13.0</td>
<td>-13%</td>
</tr>
<tr>
<td>naphthalene, 1-propyl-</td>
<td>12.0</td>
<td>10.0</td>
<td>-17%</td>
</tr>
<tr>
<td>naphthalene, 1-hexyl-</td>
<td>32.0</td>
<td>35.0</td>
<td>9%</td>
</tr>
</tbody>
</table>
Section S4: Scaling ion current to molar quantity for \(<C_{25}\) and mass for \(>C_{25}\)

Previous work shows that the total ion current (TIC) signal can be scaled to molar quantity (depending on saturation and effective vapour pressure). The TIC intensity of the individual authentic n-alkane standards (C\(_{12}\)-C\(_{34}\)), with seven injected concentrations covering 0.5ng/µL, 1ng/µL, 1.5ng/µL, 2ng/µL, 2.5ng/µL, 4ng/µL and 5ng/µL were investigated. Below \(C_{25}\), the TIC intensity of alkanes increased linearly as the molar quantity (Figure S1, red curves). The TIC molar quantity of alkyl-cyclohexane (C\(_{12}\)-C\(_{34}\)) are also plotted here to compare with the curves of n-alkanes (Figure A, black curves). The curves of the n-alkyl-cyclohexanes overlap with those of the n-alkanes in the same injection, illustrating that n-alkyl-cyclohexane and n-alkanes have a very similar response per mole. This is also found to be the same for other species with increasing alkyl chain length (e.g. alkyl benzenes and alkyl naphthalenes). However, as the carbon number increases \(>C_{25}\), the TIC response is linearly related to the compound mass, which is plotted in Figure B as an illustration. The perfect overlap of the curves also demonstrates that the TIC intensity was almost unaffected by the individual chemical structures under the same injection mass. (For the high molecular weight alkyl-cyclohexanes, only C\(_{25}\) and C\(_{33}\) were investigated due to the limitation of the available authentic standards).

The TIC molar quantity responses of deuterated standards are also shown in Figure S4-3 to further support the above view. Additionally, the TIC versus molar quantity for the alkane and alkyl-cyclohexane series are shown in Figure S4-2. This has been explored in more detail in Liang et al. (2017, in prep).
Figure S4-1. TIC intensity response to the molar quantity of individual n-alkanes and n-alkyl-cyclohexanes (C12-C24).
Figure S4.2. TIC intensity response to the molar quantity (left) and mass (right) of individual n-alkanes and n-alkyl cyclohexanes (C$_{25}$-C$_{34}$).

Figure S4.3. TIC intensity response to the normalized molar quantity of individual deuterated standards (C$_{12}$, C$_{15}$, C$_{20}$, C$_{25}$, C$_{30}$).
Figure S4-4. Plots of total ion current vs. molar quantity for the individual alkanes with carbon number from C12-C34.
Section S5S4: Representative Mass Spectra for diesel and gas phase hydrocarbon identification

Experimentally determined representative mass spectra at 70eV and 12eV ionisation are shown below. These mass spectra were identical at their respective retention times in different samples. Mass spectra at the bottom panel (blue peaks) are from 12 eV ionisation, while the top panels (red peaks) are from 70eV ionisation.

Figure S5S4-1. Representative mass spectra of n-alkanes, n-alkylcyclohexanes and n-alkylbenzenes. Top graphs with red peaks are 70eV ionisation mass spectra, while bottom graphs with blue peaks are 12eV spectra.
Figure S5-S4-2. Representative mass spectra of n-alkynaphthalenes, cis and trans-decalin, tetralin, methyl-, dimethyl-, trimethyl-, and tetramethyl tetralins. Top graphs with red peaks are 70eV ionisation mass spectra, while bottom graphs with blue peaks are 12eV spectra.
When two compounds co-elute, it is often possible to quantify each by using the low ionisation energy mass spectrum. The molecular ion of the higher molecular weight component is used to provide an ion current which is scaled upward to include the other peaks in the mass spectrum to give a total ion current due to the component. The second component can be determined by difference. The molecular ions for a C\textsubscript{11} and a C\textsubscript{12} cycloalkane are at m/z 154 and 168, respectively. These peaks would be present in the mass spectrum of co-eluting compounds, but can be separated by isolating the ion current attributed to a peak by assessing the mass spectrum that contains only the molecular ion of one of the compounds. This avoids the issue of co-elution of some compounds.
Section S6: Unresolved peaks

For the diesel fuel 90% of the total mass injected was accounted for. We suspect that a significant amount of the mass that was unaccounted may be \( \text{C}_{10} \) and/or any unresolved peaks that we were unable to measure and/or identify using our technique. There were 123 unresolved peaks in the 2D chromatogram and their positioning is shown in Figure S6-1. The intensities of the peaks, however, are relatively small and thus suggest a small contribution to the overall mass. Representative mass spectra of some of these unresolved peaks are shown in Figure S6-2.

Figure S6-1. A 2D chromatogram showing the positioning of unresolved peaks (red oval).
Figure S6-2. Representative mass spectra of unresolved peaks by this methodology. These unresolved peaks were present in the chromatographic space outlined in Figure S6-1.
**Figure S755.** A chromatogram of diesel exhaust emissions in the gas phase collected on an adsorption tube and analysed via thermal desorption with 70eV ionization mass spectrometry.

**Section S8-S6** – Representative mass spectra for compound classes identified in the lubricating oil

n + i-Alkanes were identified by the SIC of m/z 57 and selecting the corresponding molecular ion of the alkane. Figure S8-1 shows two representative alkanes with the molecular ions 282 (C20) and 296 (C21).
Figure S8S6-1. n-Alkane mass spectra at 12eV for m/z 57 and the molecular ion 282 and 296 for C20 n-alkane (top mass spectrum, red) and C21 n-alkane (bottom mass spectrum, blue), respectively.

Monocyclic alkanes were identified by the SIC of m/z 82 and selecting the corresponding molecular ion of the specific monocyclic alkane. For example, Figure S8-2 shows two representative monocyclic alkanes with the molecular ions 280 (C20) and 294 (C21).

Figure S8S6-2. Monocyclic alkane mass spectra at 12eV for m/z 82 and the molecular ions 280 and 294 for C20 monocyclic alkane (top mass spectrum, red) and C21 monocyclic alkane (bottom mass spectrum, blue), respectively.

Bicyclic alkanes were identified by the SIC of m/z 137 and selecting the corresponding molecular ion of the specific bicyclic alkane. For example, Figure S8-3 shows two representative bicyclic alkanes with the molecular ions 292 (C21) and 306 (C22).
Figure S8S6-3. Bicyclic alkane mass spectra at 12eV for m/z 137 and the molecular ions 292 and 306 for C21 bicyclic alkane (top mass spectum, red) and C22 bicyclic alkane (bottom mass spectrum, blue), respectively.

Tricyclic terpanes were identified by the SIC of m/z 191 and selecting the corresponding molecular ion of the specific tricyclic terpane carbon number. For example, Figure S8-4 shows two representative tricyclic terpanes with the molecular ions 276 (C20) and 290 (C22).

Figure S8S6-4. Tricyclic terpanes mass spectra at 12eV for m/z 191 and the molecular ions 276 and 290 for C20 tricyclic terpane (top mass spectum, red) and C21 tricyclic terpane (bottom mass spectrum, blue), respectively.

This analysis was repeated for all the compounds shown in Table 1 and representative mass spectra are shown in Figures S8-5 to S8-7 below.
**Figure S8S6-5.** Monocyclic aromatics mass spectra at 12eV for m/z 92 and the molecular ions 288 and 302 for C21 (top mass spectrum, red) and C22 (bottom mass spectrum, blue), respectively.

**Figure S8S6-6.** Pentacyclic alkanes mass spectra at 12eV for the molecular ions 258 (top mass spectrum, red) and methyl pentacyclic alkane with molecular ion 272 and m/z 257 mass fragment (bottom mass spectrum, blue).

**Figure S8S6-7.** Hexacyclic alkanes mass spectra at 12eV for the molecular ions 298 (top mass spectrum, red) and methyl hexacyclic alkane with molecular ion 312 and m/z 297 mass fragment (bottom mass spectrum, blue).
Figure S9S7. A selected ion chromatogram of lubricating oil (5W30) with mass fragments (A) m/z 92 and 119 signifying monocyclic and methyl-monocyclic aromatics (at 70eV ionization mass spectrometry) and (B) m/z 302 signifying the M⁺ for C₂₂ monocyclic aromatic isomers (at 12eV ionization mass spectrometry).

References

