

# Report

Report no. 10/19

**Investigating the HCBM -  
GCxGC relationship: an  
elution model to  
interpret GCxGC  
retention times of  
petroleum substances**



# Investigating the HCBM - GCxGC relationship: an elution model to interpret GCxGC retention times of petroleum substances

This report is prepared by:

**J. S. Arey and S. Forbes**

For:

Concawe REACH Delivery Management Group

Concawe Ecology Group

Concawe Substance Identification Group

Under the supervision of

**E. Vaiopoulou (Concawe Science Executive for REACH delivery)**

[eleni.vaiopoulou@concawe.eu](mailto:eleni.vaiopoulou@concawe.eu)

Reproduction permitted with due acknowledgement

## ABSTRACT

Risk assessments on petroleum substances (PS) are challenging due to the fact that these complex materials contain thousands of individual chemical components having widely differing physico-chemical characteristics, environmental degradation rates, and toxicities. However, comprehensive two-dimensional gas chromatography (GCxGC) can be used to separate hydrocarbon complex mixtures, thereby supporting studies of environmental risk associated with PS.

To facilitate the development of GCxGC for the evaluation of PBT properties (persistence, bioaccumulation, toxicity) of PS, Concawe developed the hydrocarbon block method (HCBM). The HCBM delineates the GCxGC chromatogram into hydrocarbon blocks (HCBs) of closely related hydrocarbon compounds: a PS containing thousands of individual constituents can be described using about 200 HCBs. Nonetheless, the mass assignments attributed by the HCBM may contain uncertainties. The objective of the present study is to investigate the assumptions and approximations that underlie the application of the HCBM to GCxGC data for the purposes of supporting PBT assessment.

To investigate the HCBM-GCxGC relationship, a theoretically-based elution model of GCxGC retention times was developed and tested. According to the elution model, the GCxGC retention times of known or hypothesized analytes can be predicted from their chemical structures, formatted as SMILES input. The retention time model was calibrated with a set of previously measured retention times for 56 hydrocarbon compounds having diverse chemical structures. The calibrated model was then applied to the entire Concawe GGraph library of 15397 individual hydrocarbon structures, providing an unprecedented theoretical prediction of the elution patterns of petroleum hydrocarbon compounds spanning the GCxGC chromatogram of diverse PS. We plotted the simulated GCxGC retention times for the 14190 chemical structures that were predicted to fall in the  $n$ -C<sub>10</sub>– $n$ -C<sub>30</sub> elution window. To visually differentiate the chromatographic region occupied by the structural members of each structural class and carbon number, a solid-line coloured polygon was overlaid onto the plots of modelled retention times. Each coloured polygon represents the convex hull that envelops the two-dimensional retention times of all members of a single class and carbon number.

By visualizing the location of each polygon for each class and carbon number in the  $n$ -C<sub>10</sub> to  $n$ -C<sub>30</sub> elution window, we conclude that the majority of modelled retention time polygons exhibit overlap with neighbouring groups of differing compound classes having the same carbon number. This approach made it possible to quantify the extent of overlap among the different chemical classes. As an example, we assumed a worst case of a PS that contains all 14190 library constituents in the  $n$ -C<sub>10</sub>– $n$ -C<sub>30</sub> elution window at relevant concentrations. Ignoring differences in carbon number within each class, we found that 79% of individual compound structures were enveloped by polygons representing two or more distinct classes, whereas 21% of individual compound structures fell into an area of the chromatogram occupied by only a single class. However, in practice, the majority of PS contain far fewer constituents in relevant concentrations, and many PS contain only a subset of the classes that are encompassed by the library.

In conclusion, the GCxGC retention time model enables an improved understanding of the elution patterns of diverse hydrocarbon compound structures on the GCxGC chromatogram. In particular, the retention time model reveals substantial overlap among the elution regions of most of the hydrocarbon classes that the HCBM was designed to quantify. This finding contrasts with current implementations of the HCBM, which assume that the designated hydrocarbon classes do not overlap in the GCxGC chromatogram. Further work would be needed to elucidate the uncertainties of the HCBM that arise from overlapping elution patterns of compound classes.

## KEYWORDS

GCxGC, Kovats, retention time, hydrocarbon block, PBT, diesel, blob, peak, affine, align, UFZ-LSER, QSAR

## INTERNET

This report is available as an Adobe pdf file on the Concawe website ([www.concawe.org](http://www.concawe.org)).

### NOTE

*Considerable efforts have been made to assure the accuracy and reliability of the information contained in this publication. However, neither Concawe nor any company participating in Concawe can accept liability for any loss, damage or injury whatsoever resulting from the use of this information.*

*This report does not necessarily represent the views of any company participating in Concawe.*

<b>CONTENTS</b>		<b>Page</b>
<b>1.</b>	<b>EXECUTIVE SUMMARY</b>	<b>V</b>
<b>2.</b>	<b>INTRODUCTION</b>	<b>1</b>
<b>3.</b>	<b>BACKGROUND</b>	<b>3</b>
3.1.	HYDROCARBON BLOCK METHOD	3
3.2.	COMPREHENSIVE TWO-DIMENSIONAL GAS CHROMATOGRAPHY (GCXGC)	3
<b>4.</b>	<b>WORK TO INVESTIGATE THE HCBM - GCXGC RELATIONSHIP</b>	<b>6</b>
4.1.	BACKBONE PROJECT	6
4.2.	A MODEL OF GCXGC RETENTION TIMES FOR PETROLEUM HYDROCARBON COMPOUNDS	6
4.2.1.	Introduction and Background	6
4.2.2.	Methodology of the GCxGC retention time model	9
4.2.3.	Calibration of the GCxGC retention time model	12
4.2.4.	Application of the GCxGC retention time model to the Concawe GGraph library	15
4.2.5.	Modelled GCxGC retention times of hydrocarbon compound classes in the $n\text{-C}_{10}\text{-}n\text{-C}_{30}$ elution window	18
4.2.6.	Using the GCxGC retention time model to interpret potential uncertainties when analyzing GCxGC data for individual chemical classes: implications for use of the HCBM in PBT assessment	27
4.2.7.	Potential future investigations	30
<b>5.</b>	<b>GLOSSARY</b>	<b>32</b>
<b>6.</b>	<b>REFERENCES</b>	<b>33</b>

## 1. EXECUTIVE SUMMARY

Risk assessments on petroleum substances (PS) are challenging due to the fact that these complex materials contain thousands of individual chemical components having widely differing physico-chemical characteristics, environmental degradation rates, and toxicities. However, comprehensive two-dimensional gas chromatography (GCxGC) can be used to separate hydrocarbon complex mixtures, thereby supporting studies of environmental risk associated with PS.

GCxGC provides excellent separation of  $C_8$  to  $C_{35}$  petroleum constituents. Due to the underlying physical principles of the separation process, distinct classes of hydrocarbon analytes form structured patterns in the two-dimensional GCxGC chromatogram, such that members of a given compound class elute in a visually recognizable cluster. Additionally, the elution position of analytes in the GCxGC chromatogram can be used to estimate the properties that influence their environmental transport behaviors, bioavailabilities, and baseline toxicities. Consequently, GCxGC has been used in several previous studies of environmental fate and environmental risk of petroleum substances.

To take advantage of these features of GCxGC, Concawe developed the hydrocarbon block method (HCBM), which facilitates the risk assessment of complex PS. It is not possible to identify and quantify all the individual components present in a complex PS, and therefore it would be impractical to conduct a risk assessment on each of the many thousands of components present. Instead, the HCBM delineates the GCxGC chromatogram into “hydrocarbon blocks” (HCBs) of closely related hydrocarbons: each HCB is a two-dimensional sub-region of the GCxGC chromatogram that is intended to encompass a select compound class and carbon number range. The HCBs are intended to group together constituents in such a way that each individual HCB envelops chemicals having similar physico-chemical, persistence, and toxicity properties. The HCB grouping is based largely upon results of quantitative structure activity relationship (QSAR) models. This enables a tractable approach to hazard assessment: a PS containing thousands of individual constituents can be described using about 200 HCBs. Each HCB might then be assessed for, persistence, bioaccumulation and toxicities properties based on knowledge of its composition.

Nonetheless, the mass assignments attributed by the HCBM may contain uncertainties: a HCB that is assigned to a single chemical class may also contain chemical members of other classes, because differing chemical classes frequently exhibit overlapping elution regions in the GCxGC chromatogram. Additionally, the quantitation of chemical mass from GCxGC raw data involves uncertainties that can originate from minor instrument instabilities, limitations in data analytics, and limitations of the detector. Commonly, a GCxGC separation is coupled to a micro Flame Ionization Detector (FID) for the quantification of HCBs, due to the wide range of dynamic linear response offered by FID, excellent signal-to-noise properties, and the near-constancy of the FID response factor for differing hydrocarbon structures.

Therefore, the objective of the present study is to investigate the assumptions and approximations that underlie the application of the HCBM to GCxGC-FID data for the purposes of supporting PBT assessment. Specifically, the present study is intended to address the methodology for the assignment of chemical constituents to hydrocarbon blocks, including:

- application of the HCBM to GCxGC data, including resolution, carbon numbers;
- assignments of chemical constituents to hydrocarbon blocks, including precision of the assignment and overlap of structures;

- mapping chemical structures onto the GCxGC chromatogram by use of a retention time model;
- uncertainty and variability related to the application of the HCBM to GCxGC data; and
- integration of this approach, identifying gaps where further work is needed.

To improve our understanding of the applicability of the HCBM to GCxGC data, the present work reports on the development and testing of a theoretically-based elution model of GCxGC retention times. The two-dimensional retention time of an analyte describes its precise position on the GCxGC chromatogram. According to the elution model, the GCxGC retention times of known or hypothesized analytes can be predicted from their chemical structures (e.g., as SMILES input). This approach relies on many of the same principles that also enable the prediction of chemical properties of hydrocarbon analytes based on their retention time positions in the GCxGC chromatogram. The retention time model contrasts with conventional approaches, in which expert judgment is required to anticipate the elution regions of individual analytes based on empirical data and experience.

The retention times of the PS analytes depend on the particular GCxGC instrument conditions that are employed for the analysis. Therefore, the GCxGC retention time model must be calibrated to a set of measured retention time data for a set of known analyte structures, in order to account for the particular instrument conditions that are applied for that analysis. The calibrated model can then be used to predict retention times of other analytes (which have not been measured) for that particular instrument program. To calibrate the retention time model, we employed a set of previously measured retention times for 56 hydrocarbon compounds having diverse chemical structures. The calibrated model is then applied to the entire Concawe GGraph library of 15397 individual hydrocarbon structures, providing an unprecedented theoretical prediction of the elution patterns of petroleum hydrocarbon compounds spanning the GCxGC chromatogram of diverse PS.

By applying the calibrated retention time model to the entire Concawe GGraph library, we plotted the simulated GCxGC retention times for the 14190 chemical structures that were predicted to fall in the  $n\text{-C}_{10}\text{-}n\text{-C}_{30}$  elution window, out of 15397 available hydrocarbon structures in the library. To visually differentiate the chromatographic region occupied by the structural members of each structural class and carbon number, a solid-line coloured polygon is overlaid onto the plots of modelled retention times. Each coloured polygon represents the convex hull that envelops the two-dimensional retention times of all members of a single class and carbon number. The plots of model retention time polygons provide insight into the ability of GCxGC to separate the individual compound groups (by class and carbon number) that are represented in the GGraph library (encompassing 11 classes and 21 carbon numbers).

By visualizing the location of each polygon for each class and carbon number in the  $n\text{-C}_{10}$  to  $n\text{-C}_{30}$  elution window, we conclude that the majority of modelled retention time polygons exhibit overlap with neighbouring groups of differing compound classes having the same carbon number. Additionally, the majority of polygons also exhibit overlap with neighbouring groups of the same class having different carbon numbers. This model exercise offers a unique perspective into the potential for overlapping two-dimensional elution patterns of the diverse hydrocarbon chemical structures that are representative of middle-distillate PS such as a diesel fuel.

With the retention time model, it is possible to calculate statistics representing the extent of overlap among the different chemical classes. As an example, we assumed a worst case of a PS that contains all 14190 library constituents in the  $n\text{-C}_{10}\text{-}n\text{-C}_{30}$  elution window at

relevant concentrations. If we ignore differences in carbon number within each class, we find that 79% of individual compound structures are enveloped by polygons representing two or more distinct classes, whereas 21% of individual compound structures fall into a set of polygons (or non-polygon groups) representing a single type of class. This analysis might be justified by the view that differentiating the exact carbon number of an analyte is less important than identifying its class, for the purposes of applying the HCBM. According to this perspective, we would infer that GCxGC retention time information alone can be used to attribute an unknown hydrocarbon analyte to a unique class in ~20% of cases, assuming the GGraph library is representative of the analyzed PS. However, these results assume an unrealistic case of a PS that contains all 14190 constituents at relevant concentrations. In practice, the majority of PS contain far fewer constituents in relevant concentrations, and many PS contain only a subset of the classes that are encompassed by the library. Additionally, the GGraph library may potentially contain some hypothetical structures that are not representative of real PS. Therefore, an expert practitioner of GCxGC may be able to successfully segregate and attribute the analyte mass assigned to individual compound classes for many PS. Whereas the Concawe GGraph library is intended to explore the universe of compositional complexity that can be encountered in PS (for  $C_{10} - C_{30}$  hydrocarbon compounds), most real PS likely contain a much smaller subset of these structures at relevant concentrations.

In conclusion, the GCxGC retention time model enables an improved understanding of the elution patterns of diverse hydrocarbon compound structures on the GCxGC chromatogram. In particular, the retention time model reveals substantial overlap among the elution regions of most of the hydrocarbon classes that the HCBM is designed to quantify. This finding contrasts with current implementations of the HCBM, which assume that the designated hydrocarbon classes do not overlap in the GCxGC chromatogram. However, further work would be needed to elucidate the uncertainties of the HCBM that arise from overlapping elution patterns of compound classes. For example, it would be possible to use the retention time model to explore the correspondence between HCB (as defined by the HCBM) and chemical class as defined by the GGraph library, including an assessment of the uncertainty of HCBM assignments that arise from the simulated overlap of elution regions for different chemical classes on the GCxGC chromatogram.

In addition, the methods of the present study might be used together with GCxGC measurements of real PS and expert judgment (from member company practitioners) to try to provide further insight into the representativeness of the Concawe GGraph library. Such work could lead to improvements in the GGraph library. In a similar vein, the retention time model could also be applied to heteroatom-containing (e.g., S-, O-, or N-containing) structures. The model is not limited to hydrocarbons.

Finally, the retention time model could be used together with the GGraph library to formulate hypotheses about how PBT properties of a PS may vary throughout the GCxGC chromatogram. The HCBM assumes that individual hydrocarbon blocks have distinct PBT properties, and it also assumes that the hydrocarbon constituents contained within an individual block have homogeneous PBT properties. The GCxGC retention time model could enable in-silico testing of these assumptions. For example, the retention time model could be used to map persistence properties and bioaccumulation properties onto the GCxGC chromatogram of a hypothetical PS composition, by combining the retention time model together with QSAR-based predictions of persistence properties or bioaccumulation properties to a PS structure library (such as the Concawe GGraph library). This approach could provide predictions of the extent to which certain PBT properties vary throughout different regions of the GCxGC chromatogram. Such predictions could be compared to the regions that are delineated by the HCBM, showing visually and quantitatively how the same analytes are expected to fall into hydrocarbon blocks.



## 2. INTRODUCTION

Comprehensive two-dimensional gas chromatography (GCxGC) is a widely accepted chemical analysis approach for petroleum substances (PS) and other nonpolar organic complex mixtures. Compared to other available techniques for chemical analysis, GCxGC provides unrivalled separation of C<sub>8</sub> to C<sub>35</sub> petroleum constituents.

GCxGC is ideally suited to support studies of environmental transport, fate, and risk associated with PS. The analysis of PS by GCxGC coupled to micro Flame Ionization Detector (GCxGC-FID) produces a two-dimensional chromatogram with several favorable properties: the FID is sensitive for hydrocarbon analytes and produces excellent signal-to-noise ratios down to low analyte concentrations; the FID exhibits dynamic linear response over a wide range of analyte concentrations; and the FID produces quantitatively similar response factors for a wide variety of different types of hydrocarbon structures (Tong and Karasek 1984, Reichenbach et al 2003, Gros et al 2014a). In addition, the two-dimensional retention time information of the GCxGC-FID chromatogram can be used to estimate the properties of the analyzed hydrocarbons that influence their environmental transport behaviors, bioavailabilities, and baseline toxicities (Arey et al 2005, Nabi et al 2014, Nabi et al 2017). As a consequence, GCxGC-FID has been used to support studies of environmental risk of PS (Redman et al 2012, Redman et al 2014, Leon Paumen 2015, Redman et al 2017) and studies of environmental transport, degradation, and fate of PS (Arey et al 2007a, b, Wardlaw et al 2008, Mao et al 2009, Gros et al 2014a, b, Gros et al 2016, Nelson et al 2016, Swarthout et al 2016).

Therefore, GCxGC can provide useful compositional data for the purposes of assessing PS for their potential to exhibit persistence, bioaccumulation, and baseline toxicity (Leon Paumen et al 2016). However, the raw GCxGC-FID chromatogram of a typical PS is a complex dataset that cannot be directly interpreted without further analysis. Therefore, in order to use GCxGC to support PBT (persistence, bioaccumulation, and toxicity) assessment of PS, it is necessary to clarify how the complex compositional datasets afforded by GCxGC can be appropriately analyzed and interpreted for this purpose.

In practice, a wide range of environmentally relevant petroleum compounds can be analyzed by GCxGC. GCxGC-amenable petroleum constituents are thermally stable compounds that fall approximately within a boiling point range from *n*-octane (*n*-C<sub>8</sub>) to pentatriacontane (*n*-C<sub>35</sub>). GCxGC cannot be applied to thermally labile compounds such as carboxylic acids that rapidly break down at the temperature of the injector inlet (typically 300-350 °C), nor to high-molecular weight (>500 g mol<sup>-1</sup>) compounds such as asphaltenes. GCxGC can be used to analyze for all petroleum hydrocarbon compounds that fall within the *n*-C<sub>8</sub> to *n*-C<sub>35</sub> elution window, as well as many hetero-substituted petroleum compounds within this window. In GCxGC terminology, the “elution window” refers to the range of *n*-alkanes that can be resolved in the first dimension, expressed as carbon number (e.g., *n*-C<sub>8</sub>). Therefore, petroleum compounds having boiling points between *n*-C<sub>8</sub> and *n*-C<sub>35</sub> can be analyzed by GCxGC, assuming a typical instrument configuration in which the first-dimension column contains a nonpolar stationary phase that separates chemical analytes by differences in volatility (i.e., boiling point), combined with a short second dimension column containing a semipolar stationary phase that separates analytes by differences in polarity. In practice, this elution window envelops the 2-ring to 5-ring parent PAHs, substituted mono-aromatic hydrocarbon compounds containing ≥8 carbon atoms, saturated hydrocarbon compounds containing 8 to 35 carbons, and many naphthenic hydrocarbon compounds ranging from decalin (C<sub>10</sub>) to larger compounds such as the sterane and hopane biomarkers.

Due to the underlying physical principles of the separation process, distinct classes of hydrocarbon analytes form structured patterns in the two-dimensional GCxGC chromatogram, such that members of a given compound class elute in a visually recognizable cluster. This advantageous feature has enabled specialists to delineate the GCxGC chromatogram into “hydrocarbon blocks” (HCBs) of closely related hydrocarbons: each HCB is a two-dimensional sub-region of the GCxGC chromatogram that is intended to encompass a selected compound class and carbon number range. When GCxGC is used in tandem with FID, the hydrocarbon block method (HCBM) can be used to aggregate the FID signal of each block within the GCxGC chromatogram, such that the total mass concentration attributed to a particular chemical class and carbon number range can be approximately determined.

The HCBM mass assignments attributed to each chemical class may contain uncertainties for three major reasons: (1) hundreds or thousands of structurally diverse chemicals occur in PS, whereas current analytical techniques are best suited for identifying the most structurally simple members; (2) a hydrocarbon block that is assigned to a single chemical class may also contain chemical members of other classes, because differing chemical classes frequently exhibit overlapping elution regions in the GCxGC chromatogram; and (3) the quantitation of chemical mass from GCxGC-FID raw data involves uncertainties that can originate from small run-to-run shifts in retention time (e.g., from minor instrument instabilities), limitations in data analytics (e.g., procedures for baseline definition and FID signal integration), and limitations of the detector (e.g., slight variations in FID response factor among different hydrocarbon compound structures).

The objective of the present study is to investigate the assumptions and approximations that underlie the application of the HCBM to GCxGC-FID data for the purposes of supporting PBT assessment. The initial results from an ongoing Concawe GCxGC Inter-Laboratory Study involving the analysis of four different PS by nine laboratories show good inter-laboratory agreement. That study and the present study are intended to address the methodology for the assignment of chemical constituents to hydrocarbon blocks, in particular:

- application of the HCBM to GCxGC data, including resolution, carbon numbers;
- assignments of chemical constituents to hydrocarbon blocks, including precision of the assignment and overlap of structures;
- mapping chemical structures onto the GCxGC chromatogram by use of a retention time model;
- uncertainty and variability related to the application of the HCBM to GCxGC data; and
- integration of this approach, identifying gaps where further work is needed.

The present work aims toward the goals stated above. However, the methods currently employed for delineating hydrocarbon blocks from GCxGC data were not made available for use together with the GCxGC retention time model that was developed for the present study, because the HCBM methodology is proprietary information within Concawe member companies.

### 3. BACKGROUND

#### 3.1. HYDROCARBON BLOCK METHOD

PS are complex materials that contain many thousands of individual chemical components. These components usually have widely differing physico-chemical characteristics, degradation rates, and toxicities, all of which influence the distribution, fate and impact of these chemicals if released into the environment. Given that it is not possible to identify and quantify all the individual chemicals present in a complex PS, and given that it would be totally impractical to conduct a risk assessment on each of the many thousands of components present, Concawe developed the hydrocarbon block method (HCBM) to facilitate the risk assessment of complex PS.

The HCBM groups together closely-related hydrocarbon chemicals into blocks which are used to describe the chemical composition of complex PS (King et al 1996). The blocking is based on various criteria, including:

- discrimination between chemical components having different physico-chemical and environmental degradation properties (e.g. vapour pressure, water solubility, octanol-water partition coefficient, biodegradation half-life, photo-oxidation potential);
- the concentrations of individual components and groups of components and the combination of such components if present at low concentrations; and
- the use of single chemical components, if unique toxicological properties or specific hazard and exposure data suggest this is appropriate (e.g. EU priority substances such as benzene, toluene and naphthalene).

Because only limited experimental data are available on the physico-chemical, (eco)toxicological and (bio)degradation properties of most the components present in complex PS, estimated property values are also employed in deciding which components should be included in the same hydrocarbon block. These estimated values are usually obtained from established and widely accepted quantitative structure activity relationship (QSAR) models.

The resulting hydrocarbon blocks are intended to group together constituents in such a way that each individual block envelops chemicals having similar physico-chemical, persistence, and toxicity properties. This enables a tractable approach to hazard assessment: a PS containing thousands of individual constituents can be described using >200 blocks. Each block can then be assessed for bioaccumulation, persistence, and toxicity properties based on knowledge of the chemical composition of that block.

#### 3.2. COMPREHENSIVE TWO-DIMENSIONAL GAS CHROMATOGRAPHY (GCXGC)

A sample analyzed by GCxGC is separated using two serially joined GC columns, with each column having different properties. All the sample components eluting from the first column are transferred at discrete time intervals to the second column, where the components undergo a further separation step. All parts of the sample are therefore subjected to two independent GC separations, providing greatly enhanced separation relative to conventional one-dimensional GC. In a typical instrument configuration applied to PS, the first-dimension column separates the sample constituents based on differences in their volatilities, and the

second dimension column separates the constituents based on differences in their molecular polarities and polarizabilities (Arey et al 2005). For truly comprehensive GCxGC, the second separation must be fast enough to preserve the information contained in the first dimension separation. So, typically the second-dimension separation takes 5-20 seconds, whereas the first-dimension separation might last for several hours. Quantification and identification of the eluting components is carried out using either universal detection systems, such as the flame ionization detector (FID) or time-of-flight mass spectrometer, or selective detection systems, such as the sulphur chemiluminescence detector.

In comparison with conventional single-column gas chromatography, GCxGC offers the following advantages:

- it provides highly detailed, readily interpretable images of very complex samples;
- it allows a complex sample to be separated into many individual peaks which can subsequently be classified for group-type analysis (such as the HCBM);
- it provides superior resolution and increased sensitivity, allowing accurate determination of low concentrations of specified components in a complex mixture; and
- it can be used to provide estimates of boiling point and partitioning properties (air-water partition coefficient, octanol-water partition coefficient) for many different classes of hydrocarbon compounds simultaneously.

GCxGC is particularly well suited to the examination of complex mixtures of middle-distillate hydrocarbons, ideally those with components in the  $C_8$  to  $\sim C_{30}$  range, such as diesel fuel, jet fuel, kerosene, and marine bunker fuel which contain many thousands of individual chemical compounds. Despite the high separation power of GCxGC, it is not possible, nor necessarily desirable, to resolve all the individual components present, so compositional reports typically include concentrations of both individual compounds (e.g. naphthalene) and groups of compounds (e.g.  $C_{11}$  iso-paraffins).

Blocking schemes for GCxGC-FID chromatogram data have been reported previously, based on groupings by estimated physico-chemical properties or by compound class (Arey et al 2007a, b, Mao et al 2009, Gros et al 2016). Although there is no standard reporting format for GCxGC results, the composition of a complex substance may be characterised using >200 hydrocarbon groups and individual components (Wang 2018). These “data elements” are usually grouped in a matrix based on individual carbon numbers and chemical functionalities such as:

- n-paraffins (nP)
- iso-paraffins (iP)
- mono-naphthenes (mN)
- di-naphthenes (dN)
- mono-aromatics (mA)
- tri-naphthenes (TriN)
- naphthenic mono-aromatics (NmA)
- di-aromatics (DiAr)
- naphthenic di-aromatics (NdAr)

- tri-aromatics and tetra-aromatics (PolyA)
- naphthenic tri-aromatics (NTriAr)

In addition to the vast amount of qualitative and quantitative information generated by GCxGC on the components present in complex PS, the technique is also well suited to environmental and human health risk assessment studies. The two-dimensional retention time data measured by GCxGC provide direct information about the volatilities, air-water partitioning properties, and octanol-water partitioning properties of the components present in a PS sample. These properties ultimately determine the rate and extent to which the petroleum hydrocarbon components in the samples may partition between the atmospheric, water, and soil compartments of the environment (Arey et al 2005, Nabi et al 2014, Nabi et al 2017). In this respect the HCBM is well-supported by the separation of the hydrocarbon components achieved by GCxGC analysis.

## 4. WORK TO INVESTIGATE THE HCBM - GCXGC RELATIONSHIP

### 4.1. BACKBONE PROJECT

Although the GCxGC retention characteristics of most constituent classes in PS are well understood, there are some compounds whose elution behaviours are not known. These include 78 compounds from the Concawe Library of Constituents in Petroleum Substances (LMC 2014), hereafter referred to as the “Concawe GGraph Library”, which are of concern owing to their perceived environmental or human health risks. Most of these compounds contain fused cyclic structures, including 2- to 6-membered ring systems with naphthenic, aromatic and hetero-cyclic characteristics.

Concawe attempted to procure as many of the 78 substances as possible and finally succeeded in obtaining 25 compounds; extensive internet searches failed to identify a supplier for some of the 78 compounds, and others were only listed by obscure foreign chemical supply houses, would have required custom synthesis and/or were prohibitively expensive.

The 25 compounds obtained were investigated by GCxGC using the same operating conditions as those usually employed for the detailed characterisation of PS (de Kruijff 2018). This study showed that:

- nine compounds eluted in the expected retention positions based on their chemical structures and established GCxGC profiles for known chemical functionalities;
- eleven compounds which contained a hetero-cyclic moiety, and therefore had no expected retention positions, eluted in the same area as their non-heteroatom analogues; and
- the remaining compounds which had no expected retention positions eluted at or close to the position of other compounds containing similar chemical functionalities.

In summary, this study indicated that despite not knowing the exact GCxGC retention characteristics of many constituents present in PS, it is possible to anticipate with a high degree of confidence where these compounds would elute, based on expert judgment.

### 4.2. A MODEL OF GCXGC RETENTION TIMES FOR PETROLEUM HYDROCARBON COMPOUNDS

#### 4.2.1. Introduction and Background

In order to gain a general understanding of the elution patterns that arise in a GCxGC chromatogram, specialists have relied on empirical data and prior experience with conventional GC. At present, expert judgment is required to interpret the large amount of compositional information contained in existing GCxGC data for a single PS. Individual signal peaks of interest on the GCxGC chromatogram, each of which is assumed to represent a single compound or small number of co-elutants, must be identified separately by time-consuming and costly analytical procedures. Normally, the chemical identity of an individual analyte peak in a PS can be ascertained by either: (1) a successful match of the mass spectrum

(a tentative identification that requires a GCxGC-TOFMS instrument); or (2) an additional injection of a synthesized standard. Neither approach is fool-proof, in the case of co-eluting structures. Technically, the definitive identity confirmation of a peak requires both techniques (a mass spectrum match and an injected standard) to be applied. Analyte identification/confirmation can be thus costly and time-consuming for even a single analyte. Therefore, a comprehensive analytical identification of a plausibly complete library of PS constituents would be cost-prohibitive.

To provide an alternative perspective that can complement these conventional techniques, a theoretically-based model of GCxGC retention times was developed and tested. The retention time model provides the basis for improved understanding of the elution patterns of diverse hydrocarbon compound structures on the GCxGC chromatogram. The model takes as input a list of hydrocarbon chemical structures (SMILES strings) of hypothetical analytes, and it produces as output a set of predicted two-dimensional GCxGC retention times for these analytes. To calibrate and test the model, we employed a set of previously measured retention times for 56 hydrocarbon compounds having diverse chemical structures. The tested retention time model is then applied to the entire Concawe GGraph library of 15397 individual hydrocarbon structures, which provides an unprecedented theoretical prediction of the elution patterns of petroleum hydrocarbon compounds spanning the GCxGC chromatogram.

We briefly provide a definition of “retention time”, so as to ensure that the model results are communicated clearly. The two-dimensional retention time of an analyte refers to the amount of time that is required for the analyte to elute from (exit) both the first-dimension column (the first-dimension retention time) and the second-dimension column (the second-dimension retention time). In practice, the elution of an individual analyte on the GCxGC-FID chromatogram is visually identified as a localized hump of elevated FID signal (a “peak”) that is approximately Gaussian-shaped with respect to both the first and second dimensions. This peak usually spans a non-negligible region of the chromatogram space - for example, a typical peak width may be 0.3–1.0 min in the first dimension and 0.3–1.0 s in the second dimension, depending on instrument conditions. For the purposes of the present report, the “retention time” of the analyte is operationally defined as the individual FID signal datum (“pixel”) having the highest magnitude within a distinct, identifiable peak. Therefore, a two-dimensional retention time of a single analyte is defined as a set of two floating point numbers: one value indicates the retention time in the GCxGC first dimension (in units of minutes); the second value indicates the retention time in the GCxGC second dimension (in units of seconds).

The retention time of a particular analyte depends on GCxGC instrument conditions (oven temperature program, inlet pressure program, column lengths, and column stationary phase types) that are employed for the analysis. Therefore, the retention time of that analyte may vary widely from one instrument analysis to another, unless identical instrument conditions are employed for both analyses. As a consequence, a model of GCxGC retention times must account for the particular instrument conditions that are applied for a given analysis. To accomplish this, the retention time model of the present work uses a set of “calibration analytes” (**Table 1**): a set of known hydrocarbon compounds for which the GCxGC retention times have been measured and recorded using a particular instrument program. These measured retention time data are used to calibrate the model; the model can then be used to predict retention times of other analytes (which have not been measured) for the particular instrument program that is represented by the calibration. We emphasize that the predictions of the GCxGC retention time model

are intended to represent the conditions of a particular GCxGC instrument program. Therefore, it would be inappropriate to compare the predictions of the retention time model based on one set of instrument conditions to measured GCxGC retention time data that have been acquired using a different set of instrument conditions.

A GCxGC chromatogram represents a two-dimensional array of retention time information that has been recorded for the separated analytes. For a given sample, the analyte patterns of two different GCxGC chromatograms are typically similar to one another, even if the instruments or instrument programs are not the same, assuming that the two analyses employ the same types of first and second dimension columns. In other words, the relative elution order of the sample constituents (“relative retention times”) remain similar when two different instruments or programs are used, despite that the two analyses may produce very different absolute values of retention times for individual constituents (again, assuming that the two column types used are identical for both runs). This feature of GCxGC provides an important basis for the HCBM, which can be applied to the same sample by different laboratories using different GCxGC instruments and different instrument programs.

Finally, the precision of measured GCxGC retention time data is affected by the experimental resolution afforded by the instrument program and also by the stability of the instrument.

- The precision of GCxGC retention time data represents the experimental variability in the measured two-dimensional retention time that can be expected for a single analyte using a given instrument program. For example, the precision of the retention time data may restrain the analyst’s ability to track changes in analyte composition from one chromatogram to another chromatogram (e.g., between two samples that experienced different experimental treatments).
- For a given instrument program, the precision of measured GCxGC retention time data depends on both the stability of the instrument conditions and the resolution of the two-dimensional signal data. The resolution of the retention time depends on the data acquisition rate of the detector and the modulation period of the 2<sup>nd</sup> column. For example, in one previous study (Gros et al 2012), the FID sampling rate was 200 Hz and the modulation period was 12.5 s.

Therefore, the experimental resolution of the two-dimensional retention time was 0.21 min for the first dimension and 0.005 s for the second dimension. In the same study (Gros et al 2012), repeated measurements of the same 7–14 hydrocarbon analytes in 12 differently weathered oil samples produced average absolute deviations of 0.04 min in the first dimension retention time values and average absolute deviations of 0.08 s in the second dimension retention time values. In this example, we can see that run-to-run instrument instability produced variability in the first dimension retention times that (on average) fell below the experimental resolution, whereas the instrument instability produced variability in the second dimension retention times that exceeded the experimental resolution. Therefore, in the Gros et al 2012 study, the precision of the first dimension retention time was controlled largely by experimental resolution (0.21 min), whereas the precision of the second dimension retention time was controlled by instrument stability (~0.08 s).

#### 4.2.2. Methodology of the GCxGC retention time model

The implementation of the GCxGC retention time model entailed four separate steps. To apply the model to a library of hydrocarbon structures, the list of SMILES strings of those structures must be used as input at the first step.

1. The UFZ-LSER QSAR model (Ulrich et al 2017) was used to estimate five Abraham solvation parameters ( $E$ ,  $S$ ,  $B$ ,  $V$ ,  $L$ ) for the 15397 hydrocarbon structures of the Concawe GGraph library. This was accomplished by applying a list of SMILES strings as input to the UFZ-LSER QSAR model.
2. The solvation parameters from step #1 were combined with published coefficients (Abraham, 1999) of linear solvation energy relationships (LSERs) to predict gas-stationary phase partition coefficients at reference conditions ( $L_1$  and  $L_2$ ) for the two stationary phases employed in the first- and second-dimension columns of the GCxGC. The two modelled GCxGC phases were: 100% dimethylpolysiloxane (Restek Rtx-1) for the first dimension, represented by published LSER coefficients for SE-30 (Abraham 1999); and 50% phenyl polysilphenylene-siloxane (SGE BPX50) in the second dimension, represented by published LSER coefficients for OV-17 (Abraham 1999).
3. A published GCxGC model (Nabi et al 2014) was re-adapted for the present study. The resulting retention time model provides a quantitative prediction of two-dimensional GCxGC retention times, using as input the theoretical  $L_1$  and  $L_2$  values from step #2. To calibrate the retention time model, three parameters ( $\alpha_1$ ,  $\alpha_2$ , and  $\alpha_3$ ) must be fitted to measurement data by regression. Fitting  $\alpha_1$ ,  $\alpha_2$ , and  $\alpha_3$  to constant values represents a specific set of GCxGC instrument conditions (e.g., temperature program, inlet pressure program, column lengths, and column types). To constrain these three model parameters, we obtained previously-measured two-dimensional GCxGC retention time data for 56 injected hydrocarbon standards (“calibration analytes”) that were reported in the literature (Arey et al., 2005). To enable a slight improvement in the model results, the initially fitted retention time model was then used to back-fit two of the five Abraham LSER coefficients (parameters  $V$  and  $L$ ), and steps 2 and 3 were repeated to provide an updated calibration (a second and final fit) of the retention time model (see text below).
4. The set of  $L_1$  and  $L_2$  values (from step 2) for the 15397 hydrocarbon structures of the Concawe GGraph library were used as an input to the calibrated GCxGC retention time model from step #3. This produced a prediction of two-dimensional GCxGC retention times for the entire structure library.

**Table 1.** List of analytes used to calibrate the GCxGC retention time model: compound names, measured retention time values ( $t_1$ ,  $t_2$ ), and estimated Abraham solvation parameter values given by the UFZ-LSER QSAR tool ( $E$ ,  $S$ ,  $B$ ,  $V$ ,  $L$ )

GGraph Library ID	Compound Name	$t_1$ (min)	$t_2$ (s)	$E$	$S$	$B$	$V$	$L$
1	<i>n</i> -decane	6.50	2.08	0	0	0	1.5176	4.746
2	<i>n</i> -undecane	13.67	2.12	0	0	0	1.6585	5.249
3	<i>n</i> -dodecane	21.83	2.13	0	0	0	1.7994	5.751
4	<i>n</i> -tridecane	30.17	2.19	0	0	0	1.9403	6.253
5	<i>n</i> -tetradecane	38.17	2.22	0	0	0	2.0812	6.755
6	<i>n</i> -pentadecane	46.00	2.25	0	0	0	2.2221	7.257
7	<i>n</i> -hexadecane	53.33	2.28	0	0	0	2.363	7.76
8	<i>n</i> -heptadecane	60.33	2.28	0	0	0	2.5039	8.262
9	<i>n</i> -octadecane	67.00	2.29	0	0	0	2.6448	8.764
10	<i>n</i> -nonadecane	73.33	2.31	0	0	0	2.7857	9.266
11	<i>n</i> -eicosane	79.33	2.33	0	0	0	2.9266	9.769
12	<i>n</i> -hen[e]icosane	85.17	2.33	0	0	0	3.0675	10.271
13	<i>n</i> -docosane	90.67	2.33	0	0	0	3.2084	10.773
14	<i>n</i> -tricosane	96.00	2.34	0	0	0	3.3493	11.275
15	<i>n</i> -tetracosane	101.00	2.34	0	0	0	3.4902	11.778
10315	naphthalene	16.50	5.35	1.28	0.93	0.18	1.0854	5.392
n.a.	tetralin	15.33	4.24	0.88	0.64	0.19	1.1714	5.17
7414	<i>n</i> -butylbenzene	8.67	3.04	0.63	0.48	0.17	1.28	4.8
n.a.	1,2,4,5-tetramethylbenzene	12.83	3.31	0.72	0.65	0.35	1.28	5.006
3193	<i>cis</i> -decalin	13.67	2.72	0.53	0.36	0	1.3004	4.985
10317	2-methylnaphthalene	25.50	5.34	1.32	0.94	0.2	1.2263	5.883
10316	1-methylnaphthalene	26.50	5.67	1.32	0.94	0.21	1.2263	5.986
7420	<i>n</i> -pentylbenzene	16.33	3.20	0.63	0.48	0.17	1.4209	5.302
1709	<i>n</i> -pentylcyclohexane	15.33	2.46	0.26	0.18	0	1.5499	5.374
n.a.	acenaphthylene	36.50	7.06	1.97	1.18	0.3	1.2156	6.262
11676	acenaphthene	39.33	6.66	1.58	1.07	0.27	1.2586	6.262
n.a.	biphenyl	32.33	5.77	1.32	1.07	0.27	1.3242	6.002
n.a.	1-ethylnaphthalene	33.50	5.57	1.32	0.86	0.23	1.3672	6.126
10326	2,6-dimethylnaphthalene	34.33	5.21	1.34	0.95	0.24	1.3672	6.24
7442	<i>n</i> -hexylbenzene	24.67	3.26	0.63	0.48	0.17	1.5618	5.804
n.a.	hexamethylbenzene	38.67	4.30	0.78	0.67	0.4	1.5618	6.082
1745	<i>n</i> -hexylcyclohexane	23.83	2.49	0.26	0.18	0	1.6908	5.876
11678	fluorene	46.50	6.80	1.69	1.16	0.37	1.3565	7.027
n.a.	<i>n</i> -heptylcyclohexane	32.33	2.52	0.26	0.18	0	1.8317	6.292
13129	phenanthrene	59.17	8.33	2.07	1.25	0.22	1.4544	7.866

GGraph Library ID	Compound Name	$t_1$ (min)	$t_2$ (s)	$E$	$S$	$B$	$V$	$L$
13128	anthracene	59.83	8.19	2.52	1.25	0.27	1.4544	7.789
7496	<i>n</i> -octylbenzene	41.33	3.28	0.63	0.48	0.17	1.8436	6.809
n.a.	<i>n</i> -octylcyclohexane	40.50	2.57	0.26	0.18	0	1.9726	6.996
13133	1-methylphenanthrene	68.33	8.18	2.02	1.26	0.26	1.5953	8.46
n.a.	9-methylanthracene	70.17	8.56	2.24	1.26	0.27	1.5953	8.485
7542	<i>n</i> -nonylbenzene	49.00	3.29	0.63	0.48	0.17	1.9845	7.311
n.a.	<i>n</i> -nonylcyclohexane	48.50	2.58	0.26	0.18	0	2.1135	7.499
14297	fluoranthene	75.83	9.16	2.27	1.53	0.32	1.5846	8.392
13137	pyrene	78.33	10.00	2.85	1.43	0.26	1.5846	8.944
n.a.	9,10-dimethylanthracene	80.17	8.85	2.25	1.27	0.31	1.7362	9.181
n.a.	<i>n</i> -decylbenzene	56.50	3.27	0.63	0.48	0.17	2.1254	7.813
n.a.	<i>n</i> -decylcyclohexane	56.00	2.60	0.26	0.18	0	2.2544	8.001
n.a.	<i>n</i> -undecylbenzene	63.50	3.27	0.63	0.48	0.17	2.2663	8.316
n.a.	<i>n</i> -undecylcyclohexane	63.17	2.56	0.26	0.18	0	2.3953	8.503
13227	benzo[a]anthracene	96.33	10.31	2.95	1.74	0.31	1.8234	10.336
13204	chrysene	96.67	10.77	2.68	1.44	0.27	1.8234	10.34
n.a.	<i>n</i> -dodecylbenzene	70.33	3.26	0.63	0.48	0.17	2.4072	8.818
n.a.	<i>n</i> -dodecylcyclohexane	69.83	2.62	0.26	0.18	0	2.5362	9.005
n.a.	<i>n</i> -tridecylcyclohexane	76.33	2.58	0.26	0.18	0	2.6771	9.507
n.a.	<i>n</i> -tetradecylbenzene	83.00	3.29	0.63	0.48	0.17	2.689	9.822
n.a.	<i>n</i> -tetradecylcyclohexane	82.33	2.65	0.26	0.18	0	2.818	10.01

Notes: “n.a.” indicates that this compound structure is not listed in the Concawe GGraph library.

The retention time model relies upon several key assumptions.

First, the model assumes that the GCxGC instrument program applies linear temperature ramps in both the first and second dimensions throughout the entire analysis. This is a typical feature of a GCxGC instrument program applied to PS.

Second, the model assumes a constant gas hold-up time in the second dimension. The gas hold-up time depends on the coupling of the inlet pressure program and the temperature program, and good chromatography practice typically aims for a constant gas hold-up throughout the analysis run. This assumption corresponds to observing a horizontal column bleed line that elutes early in the second dimension.

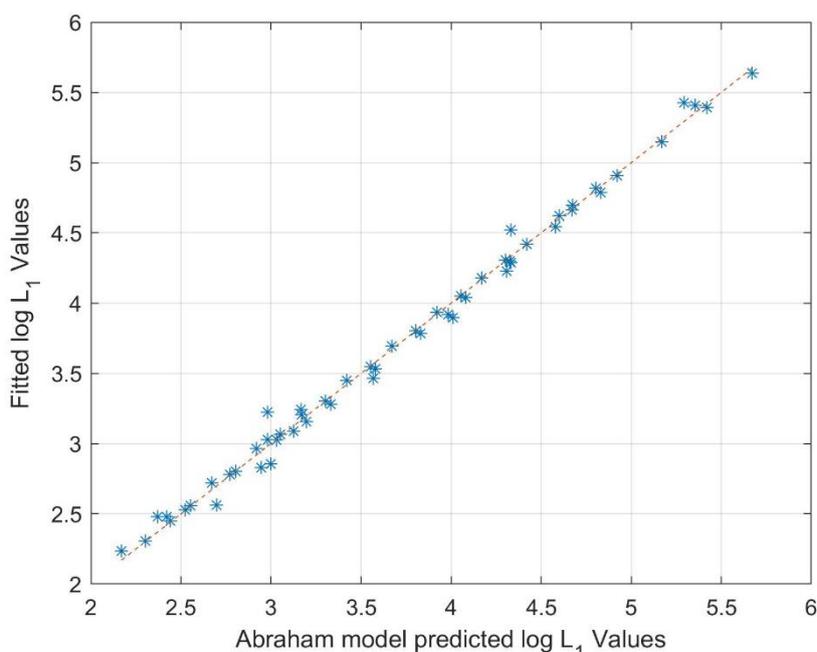
Third, the two GCxGC columns used for the first- and second-dimension separations are assumed to contain stationary phase compositions that are listed above (100% dimethylpolysiloxane for the first dimension and 50% phenyl polysilphenylene-siloxane for the second dimension).

Finally, the model incorporates the assumptions of the UFZ-LSER QSAR model that is used to generate estimated Abraham model parameters based on the SMILES string input (step #1 in the Methods); hence any errors or limitations of the UFZ-LSER QSAR model are propagated through the retention time model. The retention time model is not restricted to hydrocarbons - the model is applicable to diverse organic compound structures, assuming that they are GCxGC-amenable and neutrally charged (non-ionic) compounds.

#### 4.2.3. Calibration of the GCxGC retention time model

To calibrate the GCxGC retention time model, measured two-dimensional retention time data for 56 individual hydrocarbon standards were obtained from a previous study (Arey et al 2005). These GCxGC measurements had been conducted by Robert K. Nelson at Woods Hole Oceanographic Institution. The standards set contained several hydrocarbon classes that spanned the  $n$ -C<sub>10</sub> to  $n$ -C<sub>24</sub> elution window, encompassing: 15  $n$ -alkanes; 10 alkyl-substituted cyclohexanes; 11 alkyl-substituted benzenes; 9 parent PAHs ranging from 2-ring (naphthalene) to 4-ring (benzo[a]anthracene, chrysene); 7 alkyl-substituted PAHs; 3 naphthenic compounds (tetralin, decalin, acenaphthene); and biphenyl (Table 1). For these 56 hydrocarbon standards, the UFZ-LSER QSAR model (Ulrich et al 2017) was employed to estimate five Abraham solvation parameters ( $E$ ,  $S$ ,  $B$ ,  $V$ ,  $L$ ), using the SMILES string of each hydrocarbon structure as model input. These estimated Abraham model parameters were combined with published coefficients (Abraham 1999) of linear solvation energy relationships (LSERs) to predict gas-stationary phase partition coefficients at reference conditions ( $L_1$  and  $L_2$ ) for the two stationary phases employed in the first- and second-dimension columns of the GCxGC (see Methods).

**Figure 1.** Predicted  $\log L_1$  values according to the Abraham model (horizontal axis) versus fitted  $\log L_1$  values from measured GCxGC retention time data (vertical axis) for 56 structurally diverse hydrocarbon compounds (blue stars) in the  $n\text{-C}_{10}\text{-}n\text{-C}_{24}$  elution window. The red dotted line indicates where predicted values would equal fitted values (the 1:1 line).

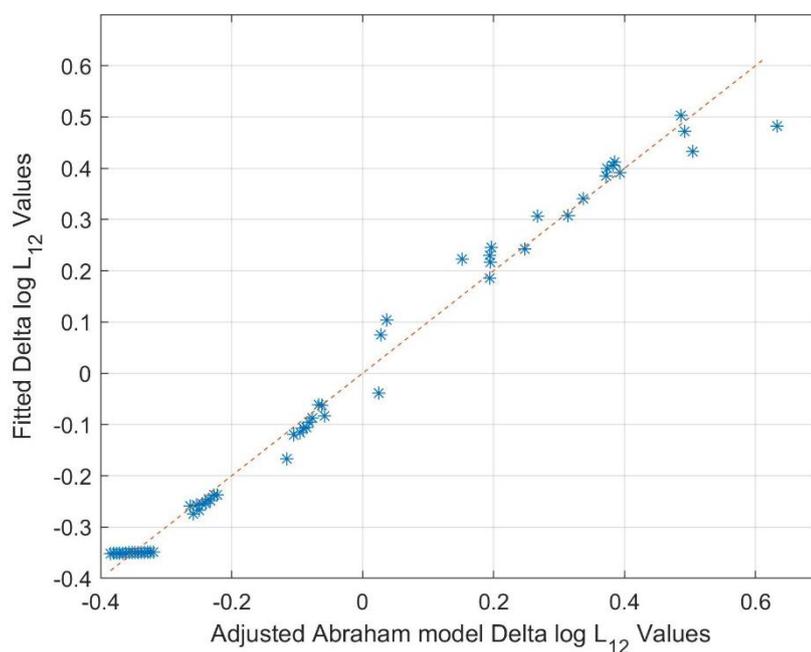


Using the measured retention time data and the theoretical  $L_1$  and  $L_2$  values for these 56 injected standards (“calibration analytes”), the retention time model parameters,  $\alpha_1$ ,  $\alpha_2$ , and  $\alpha_3$ , were adequately fitted, based on previous technical guidance (Nabi et al 2014, Arey and Nabi 2015). With this initial calibration, the fitted GCxGC retention time model produced a set of fitted  $L_1$  and  $L_2$  values for the 56 calibration analytes. These fitted  $L_2$  values, which were inferred from GCxGC measurements of the 56 calibration analytes, were then used to re-fit two of the five LSER coefficients (for parameters  $V$  and  $L$ ) that are used to predict theoretical  $L_2$  values, while maintaining as fixed the previously determined LSER coefficients for other solvation parameters ( $E$ ,  $S$ , and  $B$ ). The calibration of the retention time model was then performed a second time with updated theoretical  $L_2$  values that resulted from the updated LSER coefficients. This enabled an updated calibration procedure that produced a slightly better fit of the retention time model (Figures 1 and 2). The resulting parameter values were  $\alpha_1 = 0.24 \pm 0.05$  carbon number<sup>-1</sup>,  $\alpha_2 = -0.19 \pm 0.08$  (unitless), and  $\alpha_3 = 0.97 \pm 0.10$  s for the finally calibrated retention time model, with estimated  $\pm 95\%$  probability intervals determined by a bootstrap procedure (Nabi et al 2014). These parameter fit results indicate that the retention time model is well-constrained (Nabi et al 2014, Arey and Nabi 2015).

With this calibration of the GCxGC retention time model, the regression fit statistics for the 56 injected standards were: a correlation coefficient ( $r^2$ ) value of 0.995 for the set of model fit values of  $\log L_1$  versus the observed values of  $\log L_1$ ; and  $r^2 = 0.987$  for the model fit versus observed values of  $\Delta \log L_{12}$ . The term  $\Delta \log L_{12}$  is defined as the difference,  $\log L_2 - \beta \log L_1$ , where  $\beta$  is a coefficient such that the product  $\beta \log L_2$  is orthogonal to  $\log L_1$  for a previously defined test set (Nabi et al 2014). For the calibration set of the present study, modelled values of  $\log L_1$  were

found to be statistically uncorrelated with model fit  $\Delta \log L_{12}$  values ( $r^2 = 0.08$ ), consistent with the interpretation that  $\log L_1$  and  $\Delta \log L_{12}$  effectively provide the two independent vectors of information needed to represent first and second dimension retention times.

**Figure 2.** Predicted  $\Delta \log L_{12}$  values according to the adjusted Abraham model (horizontal axis) versus fitted  $\Delta \log L_{12}$  values from measured GCxGC retention time data (vertical axis) for 56 structurally diverse hydrocarbon compounds (blue stars) in the  $n\text{-C}_{10}\text{-}n\text{-C}_{24}$  elution window. The red dotted line indicates where predicted values would equal fitted values (the 1:1 line).



The regression fit statistics for the modelled retention times were as follows:  $r^2 = 0.994$  and root-mean-squared-error (rmse) value of 2.0 min for the model fitted values versus measured first dimension retention time ( $t_1$ ) values, and  $r^2 = 0.96$  and  $\text{rmse} = 0.59$  s for the model fitted values versus measured second dimension retention time ( $t_2$ ) values (**Figure 3**).

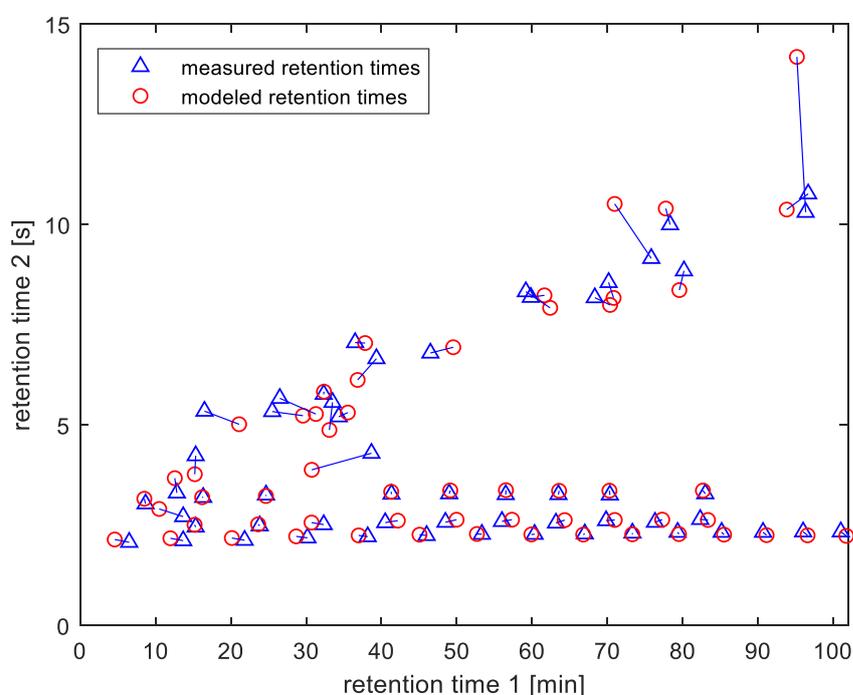
Among the 56 compounds in the calibration set, the largest error for the modelled first-dimension retention time was exhibited by hexamethylbenzene, which differed from the measured first-dimension retention time by 7.9 minutes (the measured values were  $t_1 = 38.67$  min and  $t_2 = 4.30$  s; **Figure 3**).

The largest error for the modelled second-dimension retention time was exhibited by benzo[a]anthracene, which differed from the measured second-dimension retention time by 3.9 s (measured values were  $t_1 = 96.33$  min and  $t_2 = 10.31$  s; **Figure 3**).

These results indicate that the model was able to represent the retention times of the 56 calibration analytes with a typical accuracy of  $\sim 2$  min in the first dimension and  $\sim 0.6$  s in the second dimension. However, the model exhibited larger deviations for heavily-substituted aromatic compounds and parent polycyclic aromatic hydrocarbon compounds (e.g., hexamethylbenzene, 1-methylnaphthalene,

fluoranthene, benzo[a]anthracene), on average, than for the straight chain saturated hydrocarbons (normal alkanes), the singly-substituted single-ring non-aromatic hydrocarbons (*n*-alkylcyclohexanes), or the singly-substituted single-ring aromatic hydrocarbons (*n*-alkyl-benzenes). The model also produced good results for unsubstituted prototype compounds representing 2-ring and 3-ring naphthenes (tetralin, decalin, acenaphthene). Overall, the results for the calibration set suggest that the model is able to represent the retention times of structurally simple compounds (e.g., single-ring, mono-substituted) more accurately than those of structurally complex compounds (e.g., multiple-ring aromatic, multiply-substituted aromatic).

**Figure 3.** Calibration of the retention time model with measured GCxGC retention time data for 56 structurally diverse hydrocarbon compounds in the *n*-C<sub>10</sub>–*n*-C<sub>24</sub> elution window

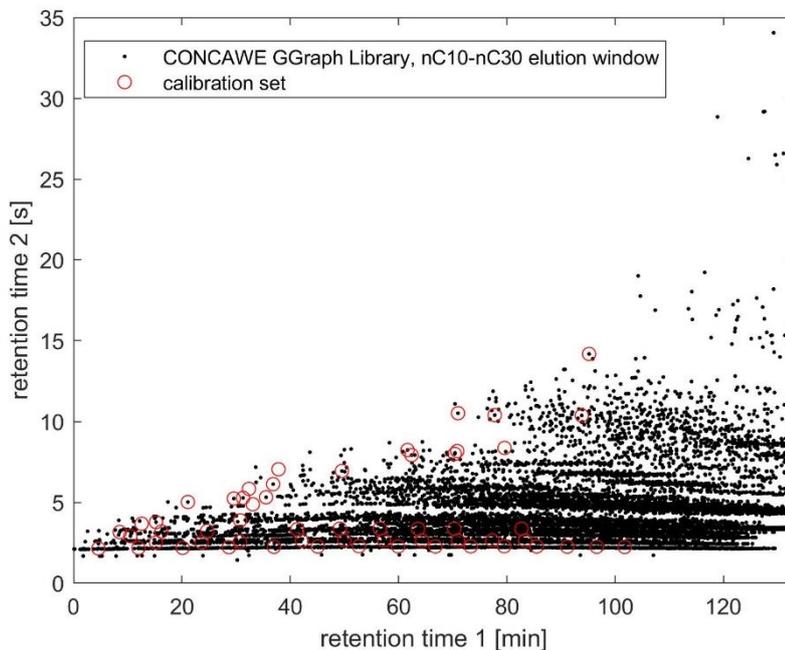


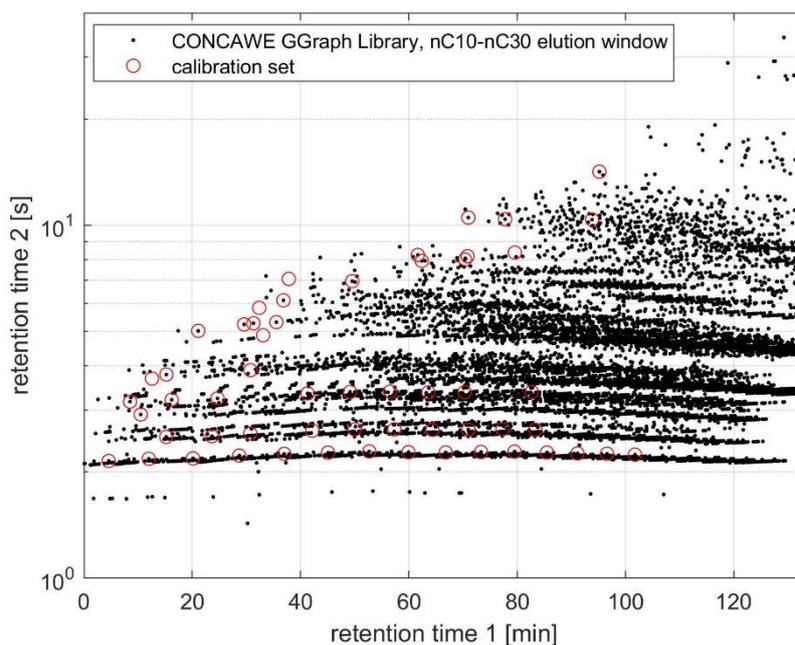
#### 4.2.4. Application of the GCxGC retention time model to the Concawe GGraph library

By applying the calibrated retention time model to the entire Concawe GGraph library, it is possible to plot the simulated GCxGC retention times for the 14190 chemical structures that were predicted to fall in the *n*-C<sub>10</sub>–*n*-C<sub>30</sub> elution window (Figure 4), out of the 15397 hydrocarbon structures of the library. The *n*-C<sub>10</sub>–*n*-C<sub>30</sub> elution window envelops the hydrocarbon constituents that are environmentally relevant in diesel-range petroleum products and heavier distillation cuts. Higher elution windows (beyond *n*-C<sub>30</sub>) comprise of constituents having low volatilities ( $<10^{-7}$  Pa) and low aqueous solubilities ( $\log K_{ow} > 7$ ); these constituents would exhibit low bioavailabilities and would therefore pose little threat to the environment (Redman et al 2012, Redman et al 2017).

Upon inspection of the plot of simulated retention times of the library (**Figure 4**, top panel), we see the triangle-shaped distribution of retention times that would be typically observed for the GCxGC chromatogram of a middle distillate PS such as a diesel fuel. It is useful to plot the simulated GCxGC chromatogram with a logarithmically-distributed second dimension retention time (**Figure 4**, bottom panel), which provides an improved visualization of the second-dimension separation of compound classes. Although experimental GCxGC chromatograms are not typically plotted with a log-transformed second-dimension axis, there is a theoretical justification for visualizing retention time results in this way. Unlike the first dimension separation (which normally employs a temperature ramp), the second-dimension separation is isothermal. Under isothermal conditions, incremental changes in compound structure will produce exponential changes in retention time, following long-established theoretical principles of gas chromatography. Consequently, the separation of chemical classes appears as a more regular sequence of structured horizontal lines, when the second dimension axis of the GCxGC chromatogram is depicted as a log-transformed retention time versus as a real retention time (**Figure 4**).

**Figure 4.** Modelled GCxGC retention times for 14190 distinct hydrocarbon structures in the  $n\text{-C}_{10}\text{-}n\text{-C}_{30}$  elution window (1<sup>st</sup> dimension retention time window of 0 to 133 minutes) of the Concawe GGraph library. Top panel: vertical axis depicted with units distributed in real time [s]. Bottom panel: vertical axis depicted with logarithmically-distributed units of time [s].





The retention time model is applied to the  $n\text{-C}_{10}\text{-}n\text{-C}_{30}$  elution window of the GGraph library structure, whereas the set of calibration analytes spanned only the  $n\text{-C}_{10}\text{-}n\text{-C}_{24}$  elution window (Figures 3, 4). Although the calibration procedure included the measured retention times of hydrocarbon compounds in the  $n\text{-C}_{10}\text{-}n\text{-C}_{24}$  elution window only, the model can be extrapolated to predict GCxGC retention times beyond  $n\text{-C}_{24}$ , by applying the constant values of the three instrument program parameters ( $\alpha_1$ ,  $\alpha_2$ , and  $\alpha_3$ ) that were fitted during the calibration procedure. Extrapolating the model to the  $n\text{-C}_{25}\text{-}n\text{-C}_{30}$  elution window corresponds to the assumption that the instrument program continues uninterrupted after the elution of  $n\text{-C}_{24}$ ; this assumes, for example, a continuation of the linear temperature ramp and pressure program that was applied for  $n\text{-C}_{10}\text{-}n\text{-C}_{24}$ . A plot of the predicted retention times for the library structures in the  $n\text{-C}_{10}\text{-}n\text{-C}_{24}$  elution window (Figure 4) reveals that they are predominantly bounded by the retention times of the 56 compounds of the calibration set. This demonstrates that the calibration set effectively brackets the fundamental chemical properties (molecular volume, molecular polarity, and molecular polarizability) that differentiate retention times (Nabi et al 2014, Arey et al 2005) for the GGraph library of hydrocarbon compounds in the  $n\text{-C}_{10}\text{-}n\text{-C}_{24}$  elution window. This finding supports our interpretation that the presently calibrated model can robustly predict GCxGC retention times for diverse hydrocarbon structures in the  $n\text{-C}_{10}\text{-}n\text{-C}_{30}$  elution window.

Not all of the 15397 hydrocarbon structures of the GGraph library fall within the  $n\text{-C}_{10}\text{-}n\text{-C}_{30}$  elution window, which is defined as the window of first-dimension retention times spanning  $n\text{-C}_{10}$  (measured as 6.5 minutes) to  $n\text{-C}_{30}$  (predicted as 132.6 minutes). For example, the  $\text{C}_{10}$  iso-paraffins are predicted to elute prior to  $n\text{-C}_{10}$ , including some  $\text{C}_{10}$  iso-paraffin compounds that are predicted to remain unresolved by this GCxGC instrument program (i.e., chromatographically unretained compounds that would elute immediately after injection), according to the model fit. In other words, some of the  $\text{C}_{10}$  iso-paraffin compounds were not assigned meaningful retention time values by the model, because the model predicts that they would not be physically resolved by the particular instrument

program that was used for the calibration analytes (Other GCxGC instrument configurations could have been selected that would effectively separate the  $C_{10}$  isoparaffins). Finally, some of the  $C_{26}$ - $C_{30}$  polycyclic aromatic compounds and other large multi-ring compounds (e.g., trinaphthenics) having <30 carbons are predicted to elute later than  $n$ - $C_{30}$ . This explains why only 14190 chemical structures are predicted fall in the  $n$ - $C_{10}$ - $n$ - $C_{30}$  elution window (**Figure 4**), out of the 15397 hydrocarbon structures of the GGraph library.

#### 4.2.5. Modelled GCxGC retention times of hydrocarbon compound classes in the $n$ - $C_{10}$ - $n$ - $C_{30}$ elution window

The GCxGC retention time model enables the visualization of individual hydrocarbon compound classes, based on the collection of candidate structures in the Concawe GGraph library (**Figures 5** and **6**). This model exercise offers unprecedented insights into the potential for overlap among diverse hydrocarbon chemical structures in the two-dimensional elution patterns that are representative of middle-distillate PS. At present, it is experimentally infeasible to selectively measure the GCxGC chromatograms of the individual classes of petroleum hydrocarbon compounds, because it is not possible to physically separate (with available fractionation methods) nor to concoct (with available pure standards) most of the individual classes of petroleum hydrocarbon compounds that are represented in the GGraph library. However, the GCxGC retention time model allows us to perform this hypothetical experiment on a computer.

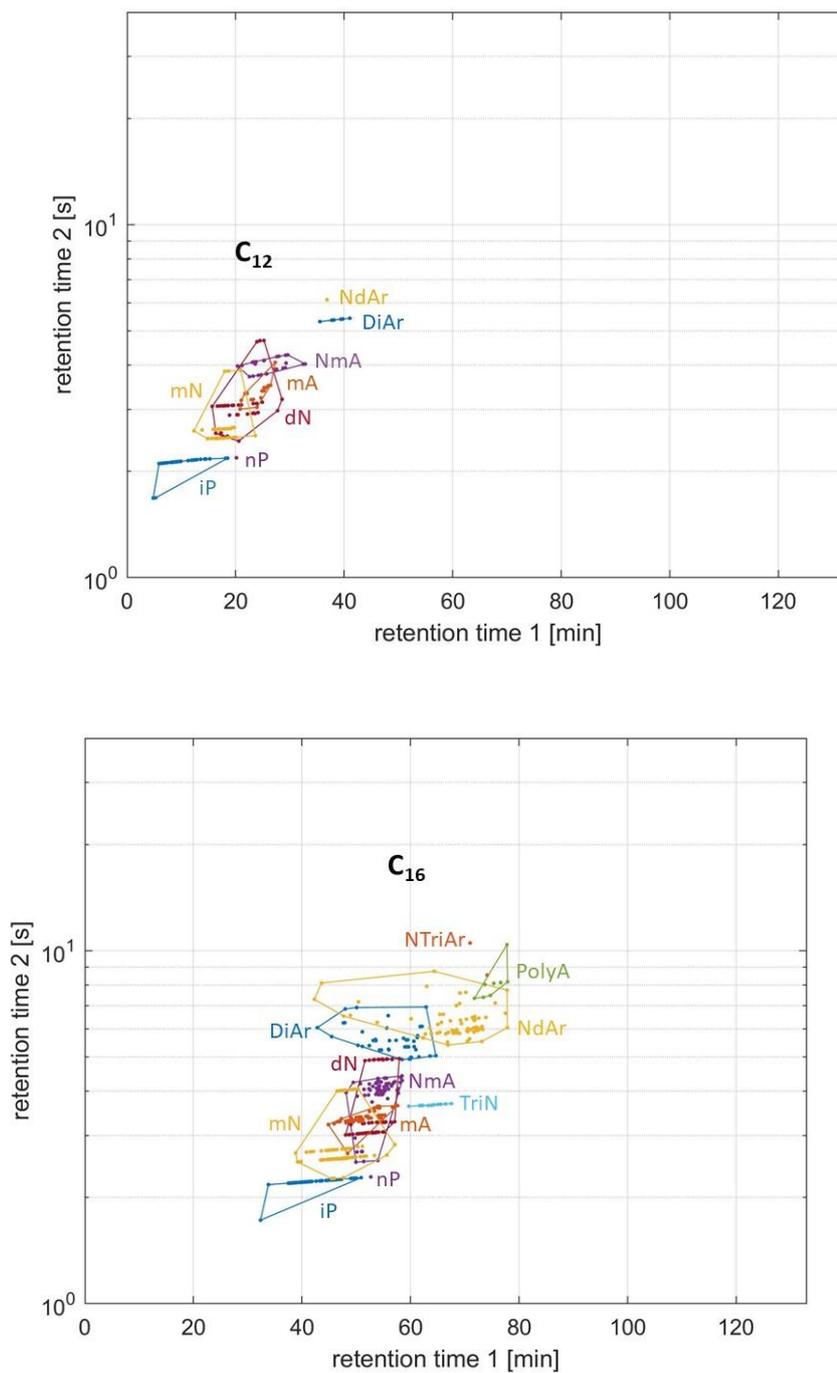
To visually distinguish the modelled GCxGC retention time patterns for each of 11 different compound classes in the GGraph library, the retention times of all structures having a given carbon number (e.g.,  $C_{16}$ ) within a single class (e.g., mono-naphthenes) are each plotted using a single color (**Figure 5**). However, even for the collection of structures having only a single carbon number, each class may contain a large number of distinct retention times, each of which is shown as a single point on the plot of modelled two-dimensional retention times. To visually differentiate the chromatographic region occupied by the structural members of a single structural class having a single carbon number, a solid-line coloured polygon is additionally overlaid onto these plots of modelled retention times (**Figures 5** and **6**). Each coloured polygon represents the convex hull that envelops the two-dimensional retention times of all members of a single class and carbon number. For example, all of the  $C_{16}$  mono-naphthenes are enveloped by a light orange polygon in the first bottom panel of **Figure 5**. The retention time polygons are drawn only for those groups that contain three or more distinct chemical structures. [Note: the lines of each polygon are drawn here for the case in which the second-dimension retention time is depicted with a log-transformed axis, but these “straight” lines would be drawn slightly differently for polygons overlaid onto retention times plotted with real time axes].

The plots of model retention time polygons provide insight into the ability of GCxGC to segregate the individual compound groups (by class and carbon number) that are represented in the GGraph library. Inspection of the separately plotted retention times of 11 individual compound classes having either 12, 16, 20, or 24 carbons (**Figure 5**) reveals that there is overlap among several of the polygons that each contain a different class. For example, the four polygons that envelop the mono-naphthenes, mono-aromatics, naphthenic mono-aromatics, and di-naphthenes all exhibit heavy overlap with one another for both the  $C_{12}$  and  $C_{16}$  groups.

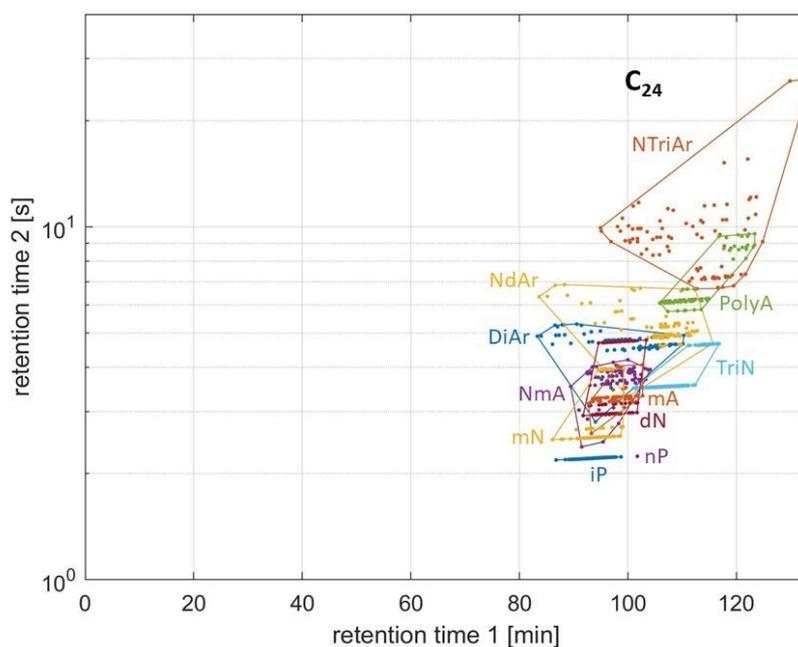
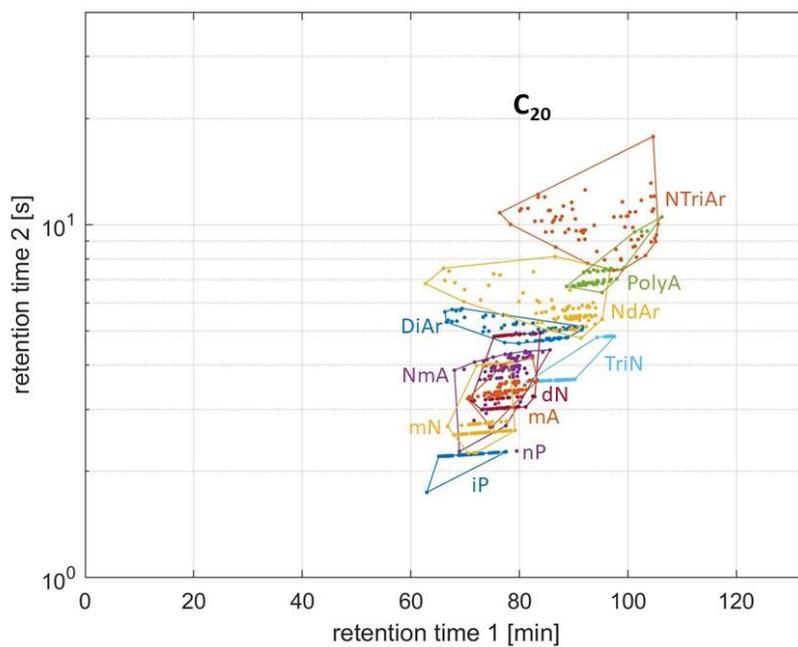
Additionally, within the  $C_{16}$  group, the two polygons that contain the di-aromatics and polycyclic aromatics exhibit overlap with the polygon enveloping the naphthenic di-aromatics. All of these types of class overlap are also observed for the  $C_{20}$  and  $C_{24}$  groups. New types of polygon overlap arise among these compound classes for the  $C_{20}$  and  $C_{24}$  structure groups, including overlaps that now also involve the tri-naphthenes and the naphthenic tri-aromatics. Across this set ( $C_{12}$ ,  $C_{16}$ ,  $C_{20}$ ,  $C_{24}$ ), we see that increasing carbon number leads to both increasing frequency of overlap and increasing extent of overlap among the 11 different class polygons. This shows that individual compound classes represented by the GGraph library become increasingly difficult to segregate by GCxGC as the carbon number is increased from  $C_{10}$  to  $C_{30}$ .

Selective visualization of the model retention times for each GGraph class reveals the extent to which GCxGC can segregate an individual class by carbon number (**Figure 6**). In the majority of cases, the modelled retention time polygons exhibit substantial overlap for groups having different carbon numbers within a single type of class (e.g., the  $C_{10}$  to  $C_{30}$  classes of di-naphthenes). For a few exceptional cases (e.g., normal paraffins, the  $C_{10}$  mono-naphthenes, or the  $C_{10}$  and  $C_{11}$  mono-aromatics), the modelled GCxGC program provides complete segregation among groups having differing carbon number within a single type of class. Inspection of these trends indicates that the extent of polygon overlap increases with increasing carbon number throughout the  $n$ - $C_{10}$  to  $n$ - $C_{30}$  elution window (**Figure 6**). Based on these results, we conclude that the majority of modelled retention time polygons exhibit overlap with neighbouring groups of differing compound classes having the same carbon number (**Figure 5**). Additionally, the majority of modelled polygons exhibit overlap with neighbouring groups of the same class having different carbon numbers (**Figure 6**). In the next section (below), we gain further insight into the extent to which the regions occupied by different compound classes experience overlap in the GCxGC chromatogram, based on further quantitative analysis of modelled retention times of the GGraph structure library.

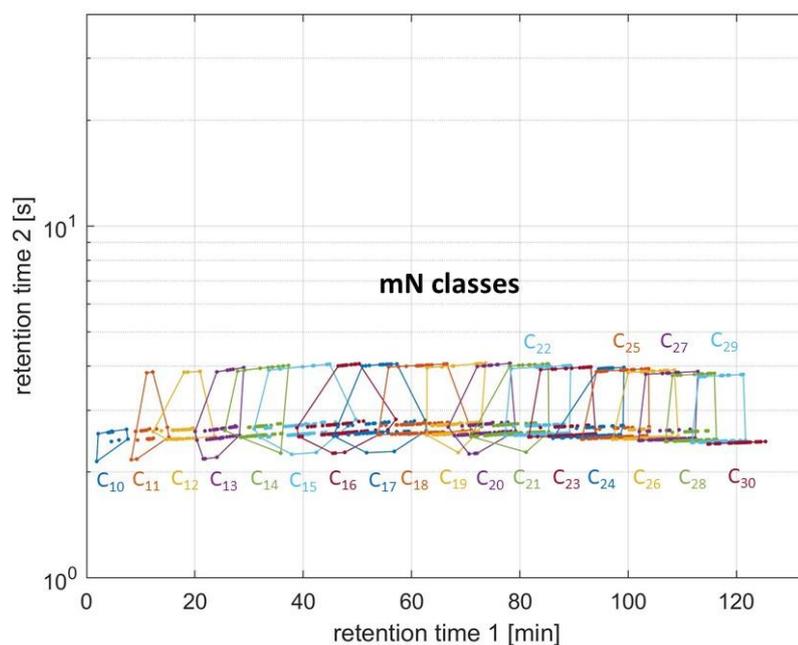
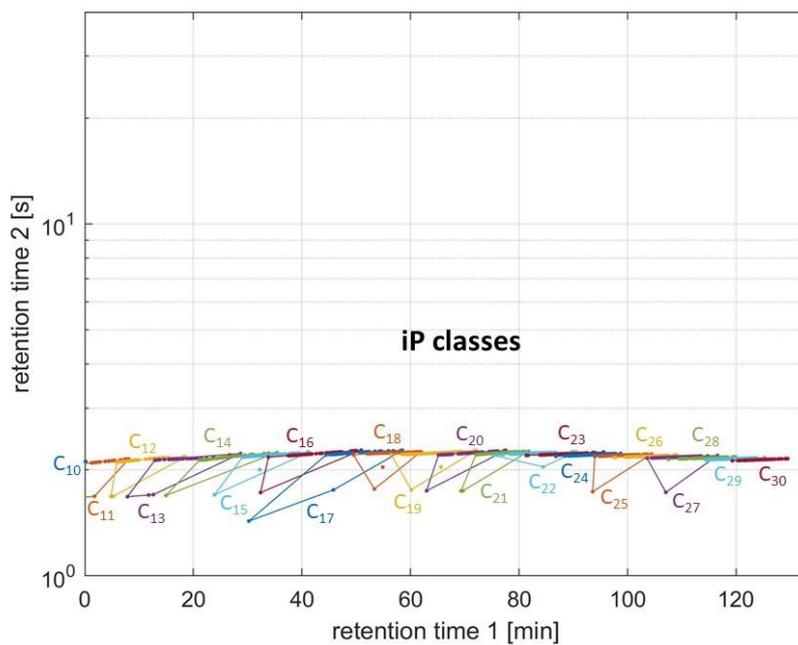
**Figure 5.** Modelled GCxGC retention times of the 11 compound classes that contain  $C_{12}$ ,  $C_{16}$ ,  $C_{20}$ , and  $C_{24}$  hydrocarbon structures in the Concawe GGraph library, depicted as an overlay on the  $n$ - $C_{10}$ - $n$ - $C_{30}$  elution window. Each coloured polygon delineates the convex hull enveloping the members of a single class (for classes containing  $\geq 3$  members). Continued on next page.



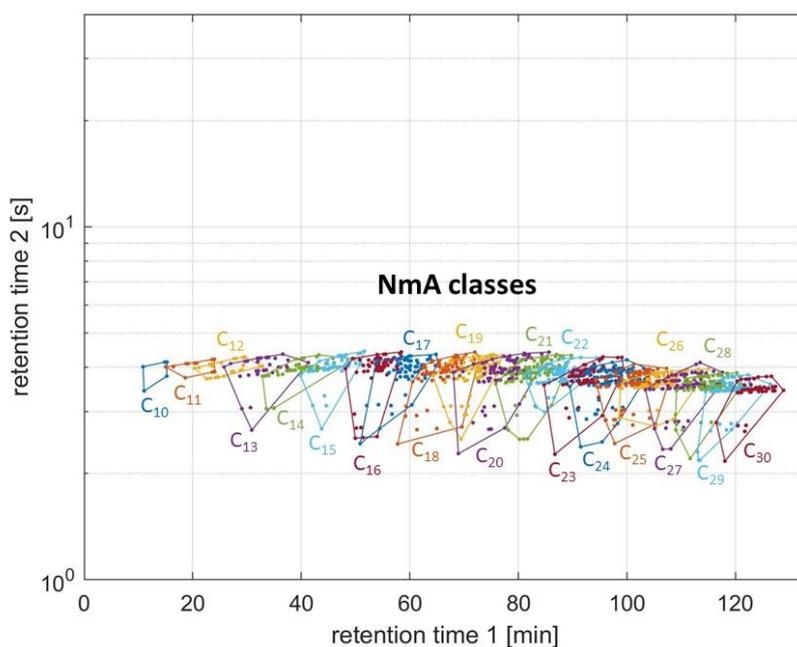
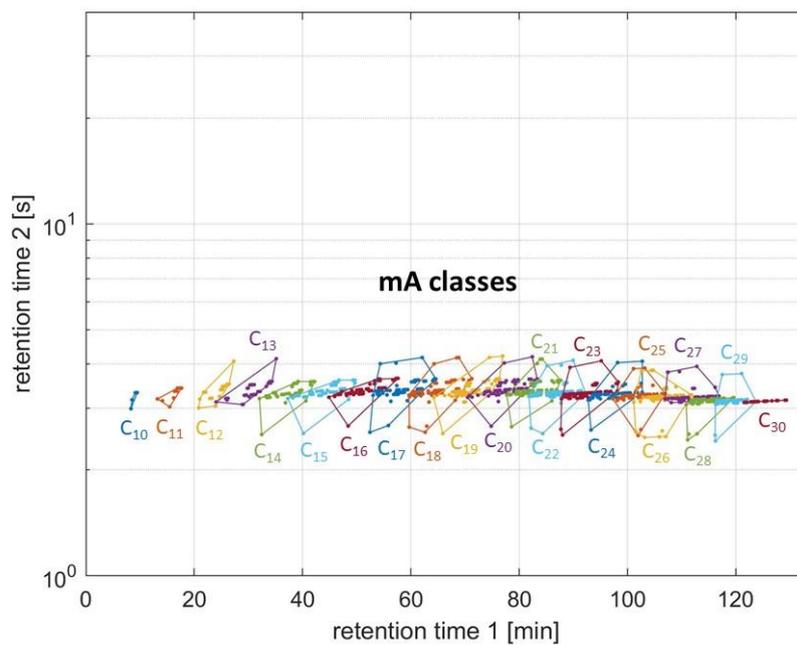
**Figure 5 (cont'd).** Modelled GCxGC retention times of the 11 compound classes that contain C<sub>12</sub>, C<sub>16</sub>, C<sub>20</sub>, and C<sub>24</sub> hydrocarbon structures in the Concawe GGraph library, depicted as an overlay on the *n*-C<sub>10</sub>–*n*-C<sub>30</sub> elution window. Each coloured polygon delineates the convex hull enveloping the members of a single class (for classes containing ≥3 members).



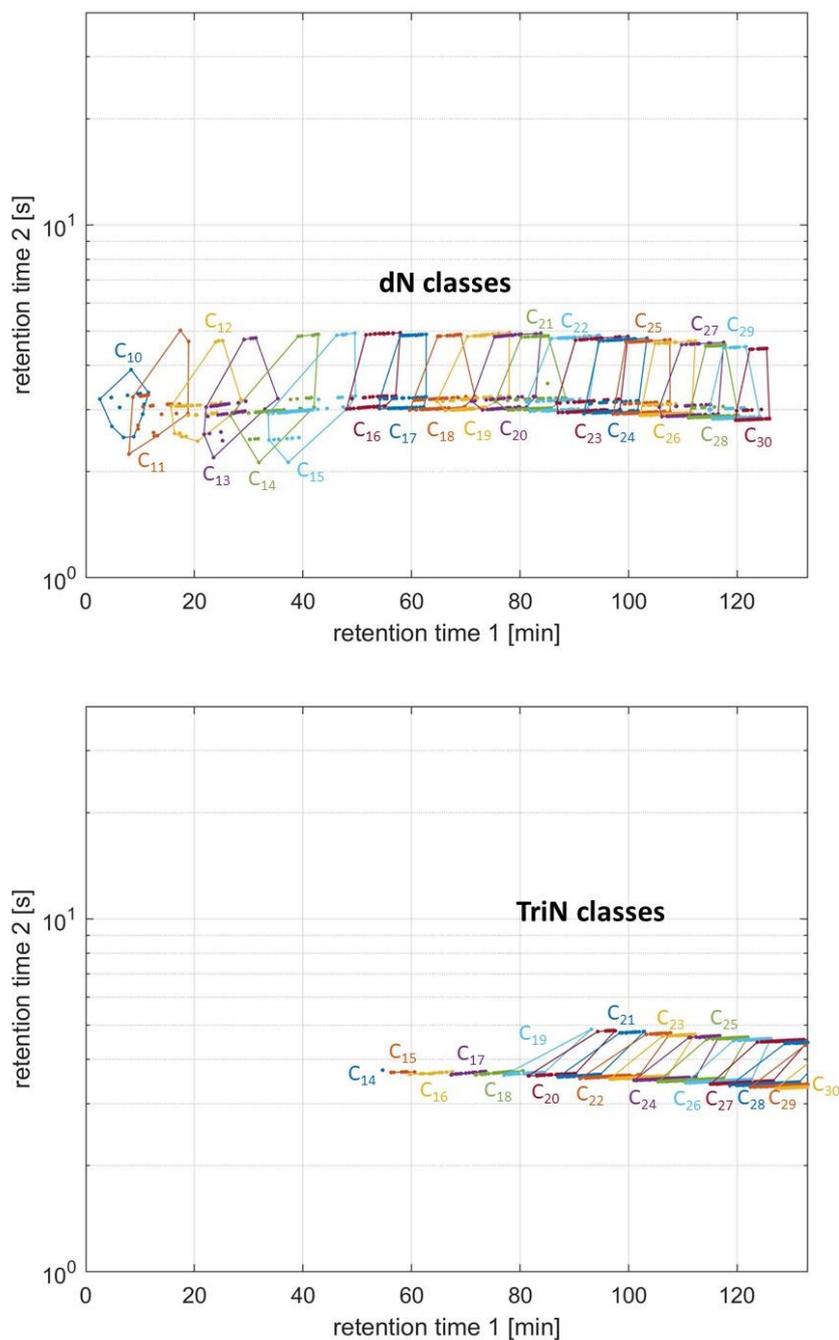
**Figure 6.** Modelled GCxGC retention times of the C<sub>10</sub>–C<sub>30</sub> structures for 10 different compound class types of the Concawe GGraph library, depicted as an overlay on the *n*-C<sub>10</sub>–*n*-C<sub>30</sub> elution window. Each coloured polygon delineates the convex hull enveloping the members of a single class (for classes containing ≥3 members). Continued on next page.



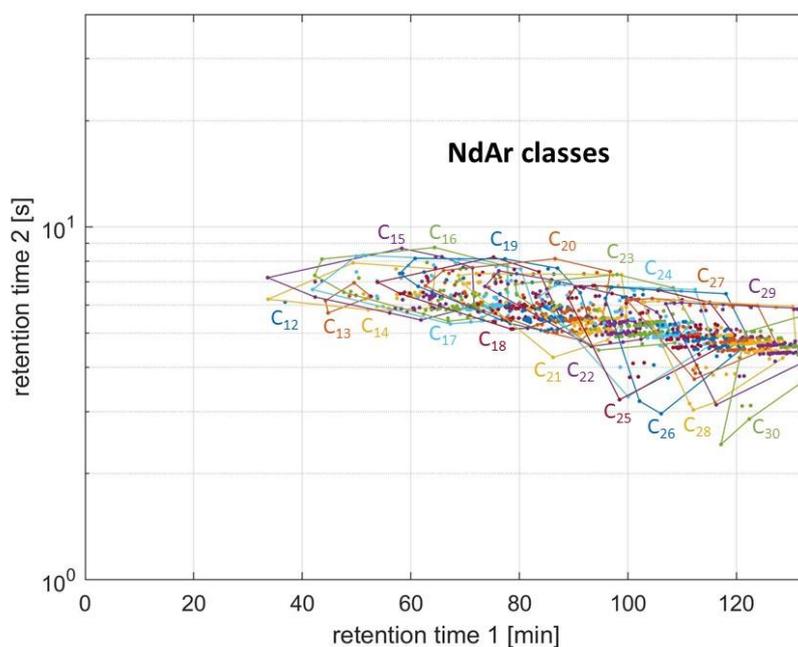
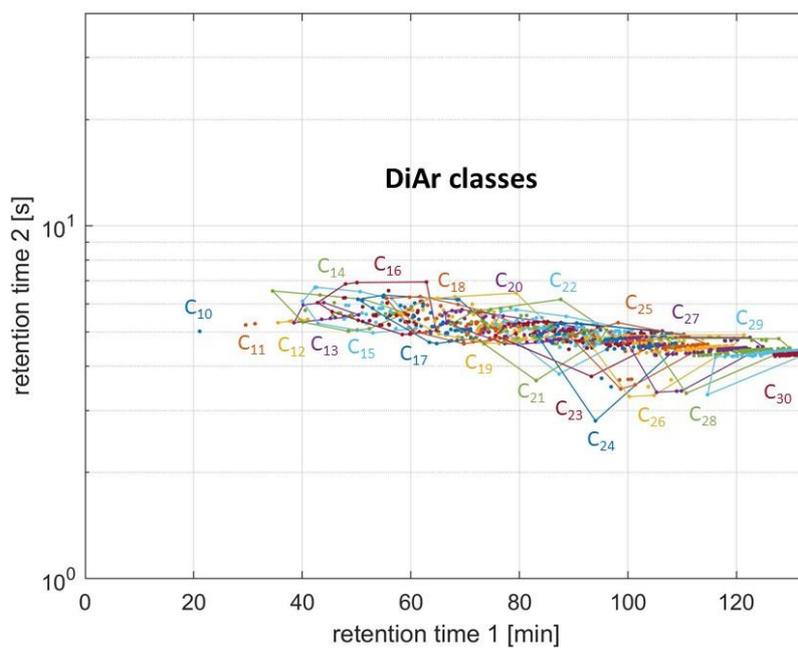
**Figure 6 (cont'd).** Modelled GCxGC retention times of the C<sub>10</sub>–C<sub>30</sub> structures for 10 different compound class types of the Concawe GGraph library, depicted as an overlay on the *n*-C<sub>10</sub>–*n*-C<sub>30</sub> elution window. Each coloured polygon delineates the convex hull enveloping the members of a single class (for classes containing ≥3 members).



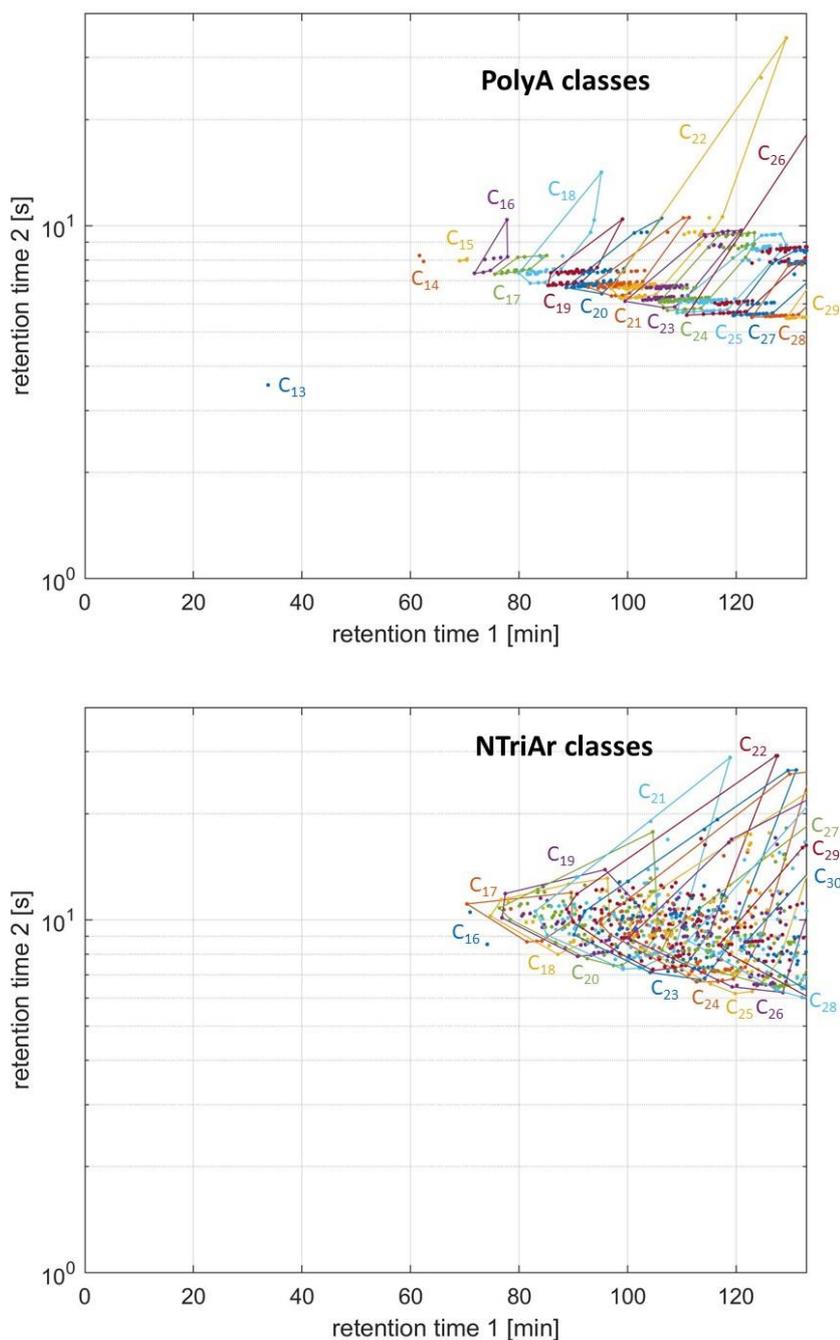
**Figure 6 (cont'd).** Modelled GCxGC retention times of the C<sub>10</sub>–C<sub>30</sub> structures for 10 different compound class types of the Concawe GGraph library, depicted as an overlay on the *n*-C<sub>10</sub>–*n*-C<sub>30</sub> elution window. Each coloured polygon delineates the convex hull enveloping the members of a single class (for classes containing ≥3 members).



**Figure 6 (cont'd).** Modelled GCxGC retention times of the C<sub>10</sub>–C<sub>30</sub> structures for 10 different compound class types of the Concawe GGraph library, depicted as an overlay on the *n*-C<sub>10</sub>–*n*-C<sub>30</sub> elution window. Each coloured polygon delineates the convex hull enveloping the members of a single class (for classes containing ≥3 members).



**Figure 6 (cont'd).** Modelled GCxGC retention times of the  $C_{10}$ – $C_{30}$  structures for 10 different compound class types of the Concawe GGraph library, depicted as an overlay on the  $n$ - $C_{10}$ – $n$ - $C_{30}$  elution window. Each coloured polygon delineates the convex hull enveloping the members of a single class (for classes containing  $\geq 3$  members).



#### 4.2.6. Using the GCxGC retention time model to interpret potential uncertainties when analyzing GCxGC data for individual chemical classes: implications for use of the HCBM in PBT assessment

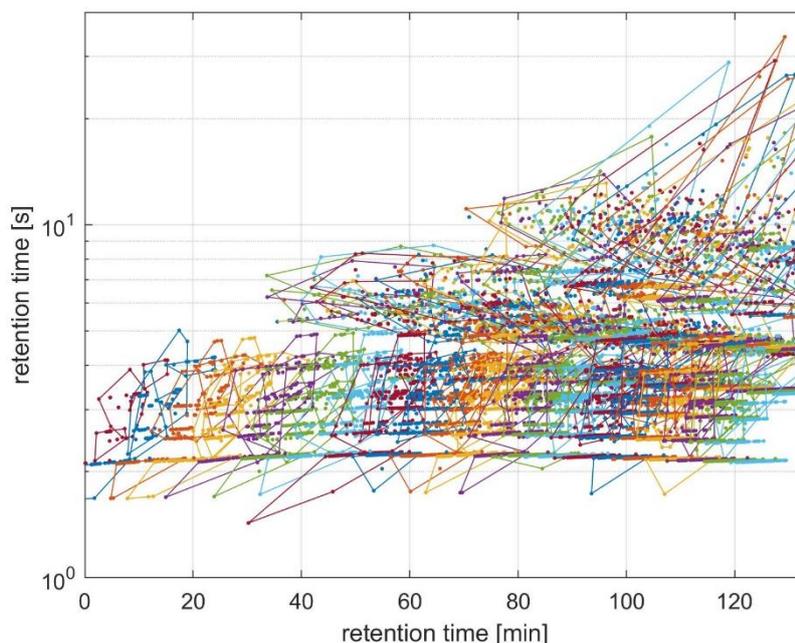
Retention time model predictions for the Concawe GGraph library may provide insights into the uncertainties that could arise when applying the HCBM to GCxGC-FID chromatograms of real PS. However, this discussion requires reflection upon several considerations that we will explore in the present section and also in the subsequent section on potential future work.

In the present study, we developed and applied a model to predict GCxGC retention times of 14190 hypothetical hydrocarbon compounds for a particular instrument program (**Figure 4**). To visualize the two-dimensional region of the GCxGC chromatogram that would envelop all of the members of a single structural class and carbon number, an individual polygon is drawn on the modelled two-dimension retention time space (**Figures 5 and 6**). By visualizing the location of each polygon for each class and carbon number in the  $n$ -C<sub>10</sub> to  $n$ -C<sub>30</sub> elution window, we conclude that the majority of modelled retention time polygons exhibit overlap with neighbouring groups of differing compound classes having the same carbon number (**Figure 5**), and they also exhibit overlap with neighbouring groups of the same class having different carbon numbers (**Figure 6**). A plot of all 216 compound groups (encompassing 11 classes and 21 carbon numbers) confirms that substantial overlap arises among the individual polygons and non-polygon groups (**Figure 7**).

The extent of overlap among these modelled compound groups can be quantified: 90.4% of individual compound structures in the  $n$ -C<sub>10</sub> to  $n$ -C<sub>30</sub> elution window are enveloped by two or more polygons (**Figure 8**), whereas only 9.6% fall into a single polygon or a non-polygon-associated group. Based on this result, one would infer that GCxGC retention time information alone can be used to attribute an “unknown” hydrocarbon analyte to a unique combination of class and carbon number in only ~10% of cases, assuming that the GGraph library is representative of the analyzed PS.

Alternatively, we can consider the extent of overlap between the sets of polygons represented by the 11 different chemical classes, ignoring differences in carbon number within each class. In this case, 79.2% of individual compound structures are enveloped by polygons representing two or more distinct classes, whereas only 20.8% of individual compound structures fall into a set of polygons (or non-polygon group) representing a single type of class (**Figure 9**). The latter analysis might be justified by the view that differentiating the exact carbon number of an analyte is less important than identifying its class, for the purposes of applying the HCBM.

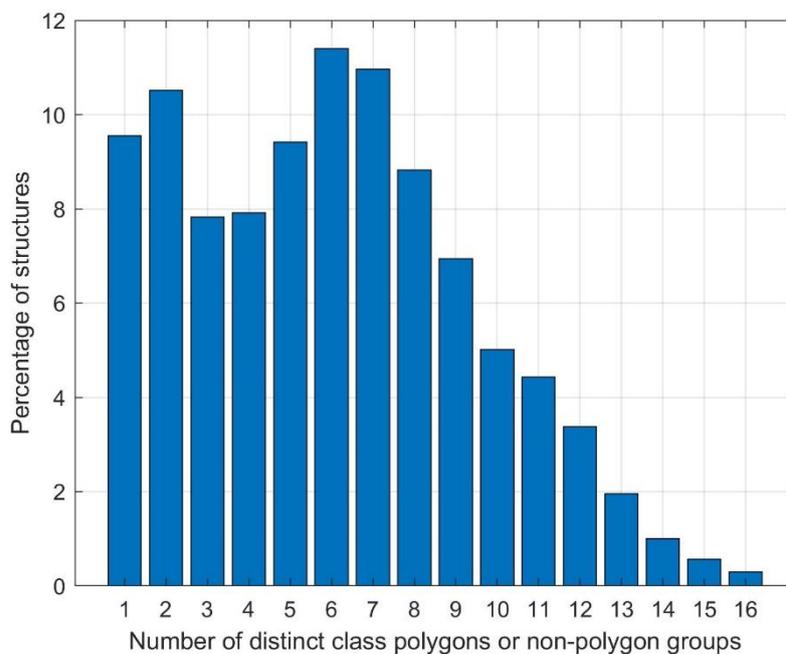
**Figure 7.** Modelled elution regions of 216 distinct hydrocarbon classes, each bounded by a separate polygon (for classes containing  $\geq 3$  members) or plotted as a non-polygon group (for classes containing  $< 3$  members), in the  $n\text{-C}_{10}\text{-}n\text{-C}_{30}$  elution window of the Concawe GGraph library.



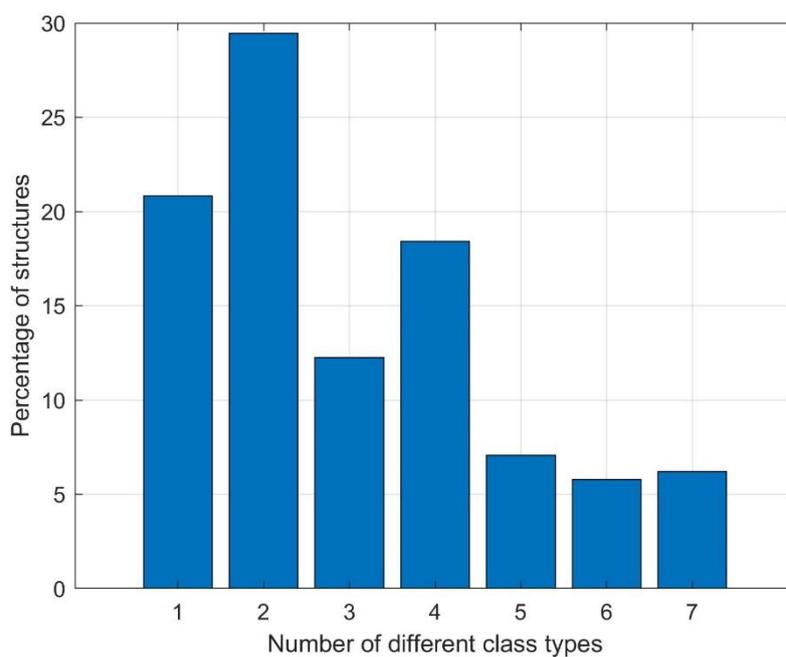
According to this perspective, we would infer that GCxGC retention time information alone can be used to attribute an unknown hydrocarbon analyte to a unique class in  $\sim 21\%$  of cases, again assuming the GGraph library is representative of the analyzed PS.

However, the results summarized above assume that the GGraph library realistically represents the compositional complexity of the analyzed PS. This assumes the worst case of a PS that contains all 14190 library constituents in the  $n\text{-C}_{10}$  to  $n\text{-C}_{30}$  elution window at relevant concentrations. In practice, the majority of PS contain far fewer constituents in relevant concentrations, and many PS contain only a subset of the classes that are encompassed by the library. Additionally, the library may potentially contain many hypothetical structures that are not representative of real PS (Forbes 2015). Therefore, a practitioner of GCxGC may be able to successfully segregate and attribute the analyte mass assigned to individual compound classes for many PS. Whereas the Concawe GGraph library is intended to explore the universe of compositional complexity that can be encountered in PS (for  $\text{C}_{10}$  to  $\text{C}_{30}$  hydrocarbon compounds), most real PS likely contain a much smaller subset of these structures at relevant concentrations. Further work would be needed to better understand how the present study provides insight into the results of the HCBM, and this is discussed in the next section.

**Figure 8.** Percentage of individual compound structures enveloped by one, two, or more different class polygons or non-polygon classes (of 216 possible), for the 14190 hydrocarbon structures in the modelled  $n$ - $C_{10}$ - $n$ - $C_{30}$  elution window of the Concawe GGraph library.



**Figure 9.** Percentage of individual compound structures enveloped by one, two, or more different structural class types (of 11 possible structural class types, ignoring differences in carbon number), for the 14190 hydrocarbon structures in the modelled  $n$ - $C_{10}$ - $n$ - $C_{30}$  elution window of the Concawe GGraph library.



#### 4.2.7. Potential future investigations

In future work, it would be possible to use the GCxGC retention time model to explore the correspondence between hydrocarbon blocks (as defined by the HCBM) and chemical classes as defined by the GGraph library, including an assessment of the uncertainty of HCBM assignments by this technique. It would be necessary for Concawe (or a member company) to supply the coordinates for a set of candidate polygons that delineate hydrocarbon blocks for real GCxGC chromatograms of PS, with a given instrument program. This implementation of the HCBM could be a non-proprietary version designed for environmental assessment that members would allow to be shared (for example, based on the ongoing inter-laboratory study), and it would not necessarily correspond to the delineation techniques that they retain as proprietary. With a non-proprietary version of the HCBM polygons, it would be possible to analyze how all of the constituents of the Concawe GGraph library would be allocated by the HCBM, according to the retention time model. These results could also be compared with the modelled class polygons that are constructed in the present report. This approach would be the best way to use the retention time model to explore the correspondence between the blocks of the HCBM and the constituent classes of the GGraph library. This work would also enable the application of the HCBM to become automated (i.e., less time-consuming and costly) and also standardized (i.e., ensuring reproducibility) among different Concawe members.

The GCxGC retention time model could also be used to discover GCxGC instrument programs that can optimally separate particular petroleum fractions or constituent classes (e.g., mono-naphthenes) according to pre-defined criteria, using the Concawe GGraph library structures as the basis to optimize a chosen objective function. For example, optimization criteria might be selected to achieve the optimal segregation of two or more constituent classes from one another, or the criteria might be selected to achieve the optimal separation of the constituent members of a particular compound class or petroleum fraction. This problem could be solved feasibly with a mathematical analysis in which the Abraham parameters of available column types ( $E$ ,  $S$ ,  $B$ ,  $V$ ,  $L$ ) and instrument program (encapsulated in the tunable model parameters,  $\alpha_1$ ,  $\alpha_2$ , and  $\alpha_3$ ) would be varied systematically to find optimal solutions to the selection criteria, using the structures of the Concawe GGraph library (or another compositional database) to define a composition. The optimized values of the varied parameters ( $E$ ,  $S$ ,  $B$ ,  $V$ ,  $L$ ,  $\alpha_1$ ,  $\alpha_2$ , and  $\alpha_3$ ) would correspond to a set of columns for the first and second dimension and a prescribed instrument program. Once suitable solutions to the optimization criteria are obtained, a practical implementation of the GCxGC instrument method then could be determined in collaboration with laboratory practitioners.

Additionally, the methods of the present study could be used together with GCxGC measurements of real PS and expert judgment (from member company practitioners) to try to provide further insight into the representativeness of the Concawe GGraph library. This might lead to efforts to further curate, extend, or otherwise improve the GGraph library.

The GCxGC retention time model can also be applied to any GCxGC-amenable compound, including heteroatom-containing (e.g., S-, O-, or N-containing) structures. The model is not limited to hydrocarbons. For example, this approach could be tested with the set of analytes that are provided in the backbone report. If needed, the method could be developed into a tool that is useful for predicting where a candidate chemical is predicted to elute on a GCxGC chromatogram.

Finally, the GCxGC retention time model could be used together with the GGraph library to formalize hypotheses about PBT properties. The HCBM assumes that individual hydrocarbon blocks have distinct PBT properties, and it also assumes that the hydrocarbon constituents contained within an individual block have homogeneous PBT properties. The GCxGC retention time model could enable in-silico testing of these assumptions, which would build on previous work (Leon Paumen et al 2016). The retention time model could be used to map persistence properties and bioaccumulation properties onto the GCxGC chromatogram of a hypothetical PS composition, by combining the retention time model together with QSAR-based predictions of persistence properties or bioaccumulation properties to a PS structure library (such as the Concawe GGraph library). For example, if combined with a model of persistence (e.g., predicted susceptibility to biodegradation based on chemical structure), the retention time model could be used to predict trends of compositional change that would be observed in the GCxGC chromatogram upon biodegradation of the PS. This approach could provide predictions of the extent to which PBT properties vary throughout different regions of the GCxGC chromatogram. Such predictions could be compared to the regions that are delineated by the HCBM, showing visually and quantitatively how the same analytes are expected to fall into hydrocarbon blocks.

## 5. GLOSSARY

Acronym	Definition
FID	Flame Ionization Detector
GC	gas chromatography
GCxGC	Comprehensive two-dimensional gas chromatography
HCB	Hydrocarbon block
HCBs	hydrocarbon blocks
HCBM	hydrocarbon block method
LSERs	linear solvation energy relationships
OV-17	50% Diphenyl, 50% dimethylpolysiloxane (Ohio Valley)
PAHs	Polynuclear Aromatic Hydrocarbons
PBT	persistence, bioaccumulation, toxicity
PS	petroleum substances
QSAR	quantitative structure activity relationship
SE-30	100% dimethylpolysiloxane (Agilent)
SGE BPX50	50% Phenyl Polysilphenylene-siloxane (Trajan)
SMILES	Simplified molecular-input line-entry system
TOFMS	time-of-flight mass spectrometry
UFZ	Helmholtz Centre for Environmental Research

## 6. REFERENCES

1. Michael H. Abraham, Colin F. Poole, Salwa K. Poole. Classification of stationary phases and other materials by gas chromatography, *Journal of Chromatography A* 1999, 842 (1-2), 79-114.  
DOI: 10.1016/S0021-9673(98)00930-3
2. J. Samuel Arey, Robert K. Nelson, Li Xu, and Christopher M. Reddy. Using Comprehensive Two-Dimensional Gas Chromatography Retention Indices To Estimate Environmental Partitioning Properties for a Complete Set of Diesel Fuel Hydrocarbons, *Analytical Chemistry* 2005, 77 (22), 7172-7182.  
DOI: 10.1021/ac051051n
3. J. Samuel Arey, Robert K. Nelson, and Christopher M. Reddy. Disentangling Oil Weathering Using GC×GC. 1. Chromatogram Analysis, *Environmental Science & Technology* 2007a, 41 (16), 5738-5746.  
DOI: 10.1021/es070005x
4. J. Samuel Arey, Robert K. Nelson, Desiree L. Plata, and Christopher M. Reddy. Disentangling Oil Weathering Using GC×GC. 2. Mass Transfer Calculations, *Environmental Science & Technology* 2007b, 41 (16), 5747-5755.  
DOI: 10.1021/es070006p
5. J. Samuel Arey, Deedar Nabi. Documentation for Matlab code to estimate partitioning properties of non-polar chemicals based on GC×GC retention time data. Version 1.2., 2015. Available from: <https://github.com/jsarey/GCxGC-property-estimation/blob/master/Documentation%20for%20users.pdf>
6. Stuart Forbes. Comments on the hydrocarbon structures of petroleum components generated by the Laboratory of Mathematical Chemistry using G-Graph Software, Nov. 2, 2015.
7. Jonas Gros, Deedar Nabi, Petros Dimitriou-Christidis, Rebecca Rutler, and J. Samuel Arey. Robust Algorithm for Aligning Two-Dimensional Chromatograms, *Analytical Chemistry* 2012, 84 (21), 9033-9040.  
DOI: 10.1021/ac301367s
8. Jonas Gros, Christopher M. Reddy, Christoph Aeppli, Robert K. Nelson, Catherine A. Carmichael, and J. Samuel Arey. Resolving Biodegradation Patterns of Persistent Saturated Hydrocarbons in Weathered Oil Samples from the Deepwater Horizon Disaster, *Environmental Science & Technology* 2014a, 48 (3), 1628-1637.  
DOI: 10.1021/es4042836
9. Jonas Gros, Deedar Nabi, Birgit Würz, Lukas Y. Wick, Corina P. D. Brussaard, Johannes Huisman, Jan R. van der Meer, Christopher M. Reddy, and J. Samuel Arey. First Day of an Oil Spill on the Open Sea: Early Mass Transfers of Hydrocarbons to Air and Water, *Environmental Science & Technology* 2014b, 48 (16), 9400-9411.  
DOI: 10.1021/es502437e
10. Jonas Gros, Christopher M. Reddy, Robert K. Nelson, Scott A. Socolofsky, and J. Samuel Arey. Simulating Gas-Liquid-Water Partitioning and Fluid Properties of Petroleum under Pressure: Implications for Deep-Sea Blowouts, *Environmental Science & Technology* 2016, 50 (14), 7397-7408.  
DOI: 10.1021/acs.est.5b04617

11. D. L. King, R. L. Lyne, A. Girling, D. R. Peterson, R. Stephenson, D. Short. Environmental Risk Assessment of Petroleum Substances: The Hydrocarbon Block Method. Concawe's Petroleum Products Ecology Group, Report No. 96/52, Brussels, February 1996.
12. Ferry de Kruijff. Concawe Backbone project: Eluents analysis by GCxGC, Shell Global Solutions Analysis Report, Oct. 9, 2018.
13. Miriam Leon Paumen. The hydrocarbon block method for environmental risk assessment of petroleum substances. SETAC Europe 25th Annual Meeting, Barcelona, May 3-7, 2015.
14. Miriam Leon Paumen, Mark A. Lampi, Mike H. I. Comber, Nadia Djemel, S. Linington, S. A. Villalobos, Charles V. Eadsforth, Klaas den Haan, Thomas F. Parkerton, Aaron D. Redman. An evaluation of the Persistence, Bioaccumulation, and Toxicity of Petroleum Hydrocarbons (revised), Concawe's Ecology Group, Brussels, Belgium, Dec. 2016.
15. LMC. ECOTOX assessment of petroleum substances - Project 1. Laboratory of Mathematical Chemistry. Burgas, Bulgaria, Aug. 2014.
16. Debin Mao, Richard Lookman, Hendrik Van De Weghe, Dirk Van Look, Guido Vanermen, Nicole De Brucker, Ludo Diels. Detailed analysis of petroleum hydrocarbon attenuation in biopiles by high-performance liquid chromatography followed by comprehensive two-dimensional gas chromatography, *Journal of Chromatography A* 2009, 1216 (9), 1524-1527.  
DOI: 10.1016/j.chroma.2008.12.087
17. Deedar Nabi, Jonas Gros, Petros Dimitriou-Christidis, and J. Samuel Arey. Mapping Environmental Partitioning Properties of Nonpolar Complex Mixtures by Use of GC × GC, *Environmental Science & Technology* 2014, 48 (12), 6814-6826.  
DOI: 10.1021/es501674p
18. Deedar Nabi and J. Samuel Arey. Predicting Partitioning and Diffusion Properties of Nonpolar Chemicals in Biotic Media and Passive Sampler Phases by GC × GC, *Environmental Science & Technology* 2017, 51 (5), 3001-3011.  
DOI: 10.1021/acs.est.6b05071
19. Robert K. Nelson, Christoph Aeppli, J. Samuel Arey, Huan Chen, André H.B. de Oliveira, Christiane Eiserbeck, Glenn S. Frysinger, Richard B. Gaines, Kliti Grice, Jonas Gros, Gregory J. Hall, Hector H.F. Koolen, Karin L. Lemkau, Amy M. McKenna, Christopher M. Reddy, Ryan P. Rodgers, Robert F. Swarthout, David L. Valentine, Helen K. White. Chapter 8 - Applications of comprehensive two-dimensional gas chromatography (GC × GC) in studying the source, transport, and fate of petroleum hydrocarbons in the environment. In: Scott A. Stout, Zhendi Wang (eds.) *Standard Handbook Oil Spill Environmental Forensics* (2nd ed.), 2016, Academic Press, pp 399-448.  
DOI: 10.1016/B978-0-12-803832-1.00008-8
20. Aaron D. Redman, Thomas F. Parkerton, Joy A. McGrath, Dominic M. Di Toro. PETROTOX: An aquatic toxicity model for petroleum substances. *Environmental Toxicology and Chemistry* 2012, 31 (11), 2498-2506.  
DOI: 10.1002/etc.1982
21. Aaron D. Redman, Thomas F. Parkerton, Mike H. I. Comber, Miriam Leon Paumen, Charles V. Eadsforth, Bhodan Dymtrasz, Duncan King, Christopher S. Warren, Klaas

- den Haan, Nadia Djemel. PETRORISK: A risk assessment framework for petroleum substances, *Integrated Environmental Assessment and Management* 2014, 10 (3), 437-448.  
DOI: 10.1002/ieam.1536
22. Aaron D. Redman, Thomas F. Parkerton, Miriam Leon Paumen, Josh D. Butler, Daniel J. Letinski, and Klaas den Haan. A re-evaluation of PETROTOX for predicting acute and chronic toxicity of petroleum substances. *Environmental Toxicology and Chemistry* 2017, 36 (8), 2245-2252.  
DOI: 10.1002/etc.3744
23. Stephen E. Reichenbach, Mingtian Ni, Dongmin Zhang, Edward B. Ledford. Image background removal in comprehensive two-dimensional gas chromatography, *Journal of Chromatography A* 2003, 985 (1-2), 47-56.  
DOI: 10.1016/S0021-9673(02)01498-X
24. Saer Samanipour, Petros Dimitriou-Christidis, Jonas Gros, Aureline Grange, J. Samuel Arey. Analyte quantification with comprehensive two-dimensional gas chromatography: Assessment of methods for baseline correction, peak delineation, and matrix effect elimination for real samples, *Journal of Chromatography A* 2015, 1375, 123-139.  
DOI: 10.1016/j.chroma.2014.11.049
25. Robert F. Swarthout, Jonas Gros, J. Samuel Arey, Robert K. Nelson, David L. Valentine, Christopher M. Reddy. Comprehensive Two-Dimensional Gas Chromatography to Assess Petroleum Product Weathering. In: Timothy McGenity, K. Timmis, B. Nogales (eds) *Hydrocarbon and Lipid Microbiology Protocols*, 2016, Springer Protocols Handbooks. Springer, Berlin, Heidelberg.
26. H. Y. Tong and F. W. Karasek. Flame ionization detector response factors for compound classes in quantitative analysis of complex organic mixtures, *Analytical Chemistry* 1984, 56 (12), 2124-2128.  
DOI: 10.1021/ac00276a033
27. Nadin Ulrich, Satoshi Endo, Trever N. Brown, N. Watanabe, G. Bronner, Michael H. Abraham, Kai-Uwe Goss. UFZ-LSER database v 3.2.1 [Internet], Leipzig, Germany, Helmholtz Centre for Environmental Research-UFZ. 2017 [accessed on 26.07.2018]. Available from <http://www.ufz.de/lserd>
28. Frank C.-Y. Wang. GCxGC principles - GC to GCxGC for complex hydrocarbon substances. GCxGC Analytical Seminar, Shell Technology Centre Amsterdam, Netherlands, Sept. 5, 2018.
29. George D. Wardlaw, J. Samuel Arey, Christopher M. Reddy, Robert K. Nelson, G. Todd Ventura, and David L. Valentine. Disentangling Oil Weathering at a Marine Seep Using GCxGC: Broad Metabolic Specificity Accompanies Subsurface Petroleum Biodegradation, *Environmental Science & Technology* 2008, 42 (19), 7166-7173.  
DOI: 10.1021/es8013908



**Concawe**  
Boulevard du Souverain 165  
B-1160 Brussels  
Belgium

Tel: +32-2-566 91 60  
Fax: +32-2-566 91 81  
e-mail: [info@concawe.org](mailto:info@concawe.org)  
<http://www.concawe.eu>

ISBN 978-2-87567-107-3



9 782875 671073 >