Modeling the GCxGC Elution Patterns of a Hydrocarbon Structure Library To Innovate Environmental Risk Assessments of Petroleum Substances

J. Samuel Arey, Alberto Martin Aparicio, Eleni Vaiopoulou, Stuart Forbes, and Delina Lyon*

Read Online

ACCESS | Metrics & More | Article Recommendations | SI Supporting Information

ABSTRACT: Comprehensive two-dimensional gas chromatography (GCxGC) offers unrivaled separation of petroleum substances, which can contain thousands of constituents or more. However, interpreting substance compositions from GCxGC data is costly and requires expertise. To facilitate environmental risk assessments, industries provide aggregated compositional information known as "hydrocarbon blocks" (HCBs), but these proprietary methods do not transparently associate the HCBs with GCxGC chromatogram data. These obstacles frustrate efforts to study the environmental risks of petroleum substances and associated environmental samples. To address this problem, we developed a GCxGC elution model for user-defined petroleum substance compositions. We calibrated the elution model to experimental GCxGC retention times of 56 known hydrocarbons by fitting three tunable model parameters to two candidate instrument methods. With the calibrated model, we simulated retention times for a library of 15,447−15,455 hydrocarbon structures (plus 40−48 predicted as chromatographically unretained) spanning 11 classes of petroleum substance constituents in the C_{10}−C_{30} range. The resulting simulation data reveal that GCxGC retention times are quantitatively associated with hydrocarbon class and carbon number information throughout the GCxGC chromatogram. These innovations enable the development of transparent and efficient technical methods to investigate the chemical compositions and environmental properties of petroleum substances, including in environmental and lab-weathered samples.
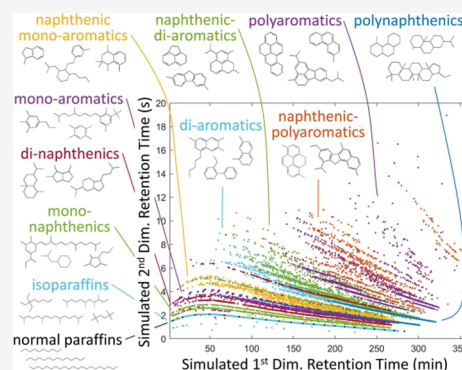
KEYWORDS: GCxGC, retention time, hydrocarbon block, petroleum, naphthenic, aromatic, UVCB

INTRODUCTION

Petroleum substances are unknown or variable composition, complex reaction products, or biological materials (UVCBs) that contain numerous constituents, ranging from hundreds of constituents in naphthas to millions in bitumens / residual aromatic extracts. To investigate the environmental impacts of these extremely complex substances, an understanding of the composition of the petroleum substance is needed; many researchers have turned to comprehensive two-dimensional gas chromatography (GCxGC).[1−15] GCxGC can resolve thousands of petroleum substance constituents in the C_6−C_{40} elution window and beyond, separating analytes into structured elution patterns that form rows and clusters of peaks according to their chemical class and carbon number.[2,8,9,16−21] When coupled to a flame ionization detector, GCxGC permits consistent quantitations of petroleum hydrocarbon analytes individually and as groups, due to this detector's excellent signal-to-noise ratio, dynamic linear response over a wide concentration range, and a response factor, which is proportional to carbon number for hydrocarbons.[11,22−24]

Despite the advantages of GCxGC for studying petroleum substances, interpreting chromatogram data is costly and requires deep expertise. For example, investigators have reported that the naphthenic, branched saturate and substituted aromatic fractions exhibit slow degradability compared to other GCxGC-amenable fractions,[1,2,4,5,11] bringing attention to the environmental relevance of these petroleum hydrocarbon classes. However, these fractions exhibit crowded and overlapping elution patterns in the GCxGC chromatogram,[2,11] which complicates their compositional interpretation, and few chemical standards are available to identify the analytes. Yet GCxGC-FID and GCxGC coupled to time-of-flight mass spectrometry[8,11,20,25,26] are among the most effective analytical tools available to separate and interpret these challenging petroleum fractions, which are even less accessible with conventional analyses by gas chromatography coupled to mass spectrometry.
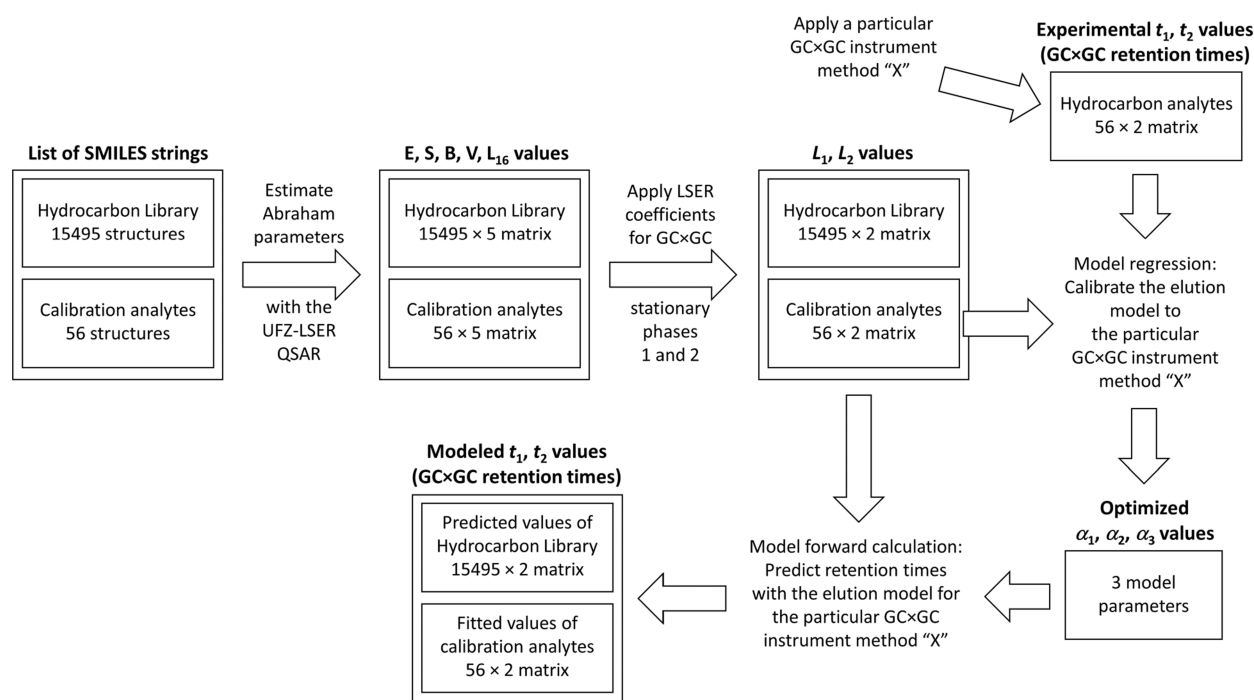
Figure 1. Schematic overview of the calibration of the elution model to a particular GCxGC instrument method and application of the elution model to the hydrocarbon library.

Further exacerbating these challenges, the existing data analysis methods for GCxGC do not effectively address the information needs of environmental and human health risk assessments. Petroleum substances are not amenable to conventional risk assessment approaches for chemical mixtures, because many constituents have not been positively identified.[27] To adapt petroleum substances to the rubric of risk assessment,[28−31] experts simplify the composition information by aggregating the constituents in groups called "hydrocarbon blocks" (HCBs),[32] which can be derived from GCxGC-FID data.[33] Each HCB consists of a specified range of carbon numbers within a class of structurally related constituents (e.g., C_{14} dicyclic aromatics), which are expected to exhibit similar toxicities and environmental behaviors. GCxGC-FID is well-suited to support the derivation of HCBs: industry experts can delineate the GCxGC-FID chromatogram into segregated regions that represent individual HCBs, each assigned to a particular chemical class and carbon number range.[33] However, the existing HCB implementations are proprietary methods, originally designed for internal industry purposes, and they cannot be applied to environmental samples nor reproduced externally. Consequently, environmental and human health experts have not established representative constituent compositions for many HCBs, and the available HCB data are not transparently associated with the GCxGC chromatogram data of the analyzed petroleum substances.

These obstacles frustrate efforts to evaluate the risks of petroleum substances and their constituents. For example, previous works have employed GCxGC to determine the primary biodegradation of thousands of petroleum substance analytes,[7,34,35] but the investigators lacked methods to interpret the compositional changes in these rich data sets, which complicates the evaluation of constituent degradability properties. Although methods exist to estimate partitioning and bioavailability properties of petroleum substance constituents by their retention position on the GCxGC chromato-

gram,[3,15,36] experts still struggle to characterize the biodegradation and biotransformation properties of most constituents or sub-fractions (e.g., HCBs). These properties are key inputs to assessments of environmental persistence and bioaccumulation. The abovementioned information gaps also stymie attempts to evaluate the similarities among the constituents within petroleum substances or HCBs, which limits industry's ability to address regulatory requirements designed for constituent-based testing approaches.[37]

To address these challenges, here we develop an elution model of GCxGC retention times. We calibrate the elution model with previously reported retention time data of 56 structurally diverse hydrocarbon compounds ("calibration analytes"). We then apply the calibrated model to an in-house library of 15,495 individual hydrocarbon structures, which enables a theoretical visualization of the elution clusters of 217 HCBs that originate from 11 different hydrocarbon classes spanning C_{10} to C_{30}. Taken together, the GCxGC elution model and hydrocarbon library provide an unprecedented theoretical basis to simulate the hydrocarbon class and carbon number information of GCxGC-analyzed petroleum substances. These methods elucidate quantitative associations between GCxGC elution patterns and petroleum constituent composition, which in turn forges a path to develop transparent, reproducible methods to delineate and quantify HCBs for risk assessment of petroleum substances. These advances can address essential needs in the existing framework to assess the environmental risks of petroleum substances, and they represent a significant step toward unlocking the full potential of GCxGC to innovate risk assessment of petroleum substances.

■ METHODS

Development of a GCxGC Elution Model for Petroleum Substances. We developed and tested a model

of GCxGC retention times (Figure 1). The model framework takes as input a list of chemical structures, which could represent a petroleum substance composition, defined with the simplified molecular-input line-entry system (SMILES). The elution model produces as output a set of predicted two-dimensional GCxGC retention times for these constituents. The model output depends on the GCxGC instrument conditions. The elution model can flexibly represent a useful range of possible GCxGC instrument methods through three tunable model parameters: α_1, α_2, and α_3. These parameters are further explained with eqs 3 and 4 below. To adapt the model to a particular instrument method, the user must optimize the three parameters by regression of the model equations to measured two-dimensional retention time data for a training set of known calibration analytes. The user can then employ the calibrated model to predict GCxGC retention times of constituents outside of the training set, assuming the same instrument method conditions. The GCxGC elution model can simulate diverse chemical structures containing a wide variety of functionalities. Section S5 describes the assumptions, limitations, and domain of the elution model. The elution model code is freely distributed under an open-source license.[38]

We describe the development and application of the GCxGC elution model (Figure 1) as five sequential steps:

Development of a Hydrocarbon Library. Anticipating the need to represent realistic petroleum substance compositions, Concawe assembled a library of 15,495 individual hydrocarbon structures designed to aid environmental risk assessment of petroleum substances (Section S6).[39,40] The hydrocarbon library consists of known (i.e., experimentally verified) constituents that occur in crude oils and refinery distillation streams and hypothetical (i.e., plausible) hydrocarbon structures containing 10 to 30 carbon atoms. The library does not encompass the conversion products of secondary refinery processes such as cracking and hydroprocessing. We categorized the library constituents according to 11 hydrocarbon classes: normal paraffins (nP), isoparaffins (iP), monocyclic naphthenes (mN), dicyclic naphthenes (dN), polycyclic naphthenes having ≥3 rings (polyN), monocyclic aromatics (mAr), dicyclic aromatics (dAr), polycyclic aromatics having ≥3 rings (polyAr), naphthenic monocyclic aromatics (NmAr), naphthenic dicyclic aromatics (NdAr), and naphthenic polycyclic aromatics (NpolyAr). The nomenclature of these classes follows terminology that is familiar to petrochemical industry rather than academic conventions (e.g., "paraffins" rather than "alkanes") in order to facilitate application by industry. The hydrocarbon library does not represent a comprehensive listing of all chemical structures that could occur in petroleum substances. More detailed information is provided in Section S6. For this work, the hydrocarbon library provides a reasonable basis to investigate simulated GCxGC elution patterns of petroleum substances within these limitations.

Concawe assigned the 15,495 hydrocarbon structures to 217 HCBs, defined by splitting each of the 11 hydrocarbon classes into 21 bins according to carbon number (C_{10} to C_{30}). Each HCB represents the subset of structures sharing a common hydrocarbon class and carbon number (e.g., C_{20} polycyclic naphthenes). Some classes do not span the entire range of 21 carbon numbers, because they have a minimum size of >10 carbons (e.g., polyAr, NdAr, and NpolyAr).

Estimation of Chromatographic Retention Properties of Each Hydrocarbon in the Library. To characterize the chromatographic retention properties of the hydrocarbons, we estimated their gas-stationary phase partition coefficients, L_1 and L_2, at a reference temperature (121 °C) in each of the two respective stationary phases employed in the first- and second-dimension columns of the GCxGC, as follows. Using the SMILES string of each hydrocarbon as input, we employed the online UFZ-LSER tool[41,42] (access dates: 08/25/2018, 01/04/2021) to estimate the following four Abraham solvation parameters: E, S, V, and L_{16}. These solute parameters represent, respectively: E, extent of electronic polarizability exceeding that of an identically-sized *n*-alkane; S, electrostatic polarity and electronic polarizability; V, molecular volume/surface area; and L_{16}, the gas-hexadecane partition coefficient.[43,44] We estimated the L_1 and L_2 parameters of each library constituent, i, by the linear solvation energy relationship (LSER) equations of Abraham:

$$\log L_{1,i}(121°C) = 0.024E_i + 0.190S_i + 0.498L_{16,i} - 0.194 \tag{1}$$

$$\log L_{2,i}(121°C) = 0.071E_i + 0.653S_i + 0.518L_{16,i} - 0.372 \tag{2}$$

To obtain the coefficients for eqs 1 and 2 (Table S2), we represented the first- and second-dimension stationary phases of the GCxGC instrument, 100% dimethylpolysiloxane (Restek Rtx-1) and 50% phenyl polysilphenylene-siloxane (SGE BPX50), with published LSER coefficients for the SE-30[43] and OV-17 stationary phases,[43] respectively. As written above, eqs 1 and 2 omit the reported coefficient[43] to the hydrogen-bonding parameter (A), because this parameter is not relevant for hydrocarbon structures.

The reliance on LSER chromatographic parameters differentiates the GCxGC elution model from other GCxGC retention time models.[45−51] Due to the broad predictive capability of the UFZ-LSER tool,[41,42] users can apply the GCxGC elution model to diverse chemical structures containing a wide variety of functionalities, including all of the hydrocarbon structures in the library.

Derivation of the GCxGC Elution Model. Based on chromatographic relationships in previous work,[3,36] we assumed that $L_{1,i}(121°C)$, obtained from eq 1, can be related to the observed first-dimension retention time for each analyte i, through:

$$\log L_{1,i}(121°C) = \alpha_1 N_i^* + \alpha_2 + \epsilon_{i,1} \tag{3}$$

where α_1 (having units of carbon number^{-1}) and α_2 (unitless) represent tunable model parameters that can be adjusted to enable a best-fit of eq 3 with observed first-dimension retention time values for a given GCxGC instrument program, $\epsilon_{i,1}$ represents the discrepancy of the fitted $\log L_{1,i}(121°C)$ value by eq 3 for analyte i, and N_i^* represents the fractional number of carbon atoms of a hypothetical *n*-alkane (nP class) that would elute with the same first-dimension retention time as analyte i, interpolated from the two nearest *n*-alkanes (eq S1). Eqs 3 and S1 assume that the GCxGC instrument employs a linear temperature program for the first-dimension column (primary oven).[36]

We assumed that the $L_{2,i}(121°C)$ value, obtained by eq 2, can be related to the observed second-dimension retention time for each analyte i, through:[36]

Table 1. Calibration Parameters and Statistical Properties of the Elution Model for Two Different GCxGC Instrument Methods

GCxGC instrument method	α_1	α_2	α_3	t_1 statistics for 56 calibration analytes		t_2 statistics for 56 calibration analytes		number of library structures with predicted t_1 $\geq$ 2 min (of 15,495)
				RMSE (min)	r^2	RMSE (s)	r^2	
A	0.245 ± 0.004^a	-0.21 ± 0.07^a	0 ± 0.1^a	4.3^b	0.994^b	0.63^b	0.95^b	15,447
B	0.243 ± 0.005^a	-0.19 ± 0.08^a	1.0 ± 0.1^a	2.6^b	0.994^b	0.50^b	0.96^b	15,455

[a]Reported uncertainties for α_1, α_2, and α_3 represent an estimated 95% probability interval as determined by a bootstrap of the model regression equations. [b]Reported root-mean-squared-error (RMSE) and squared correlation coefficient (r^2) values indicate the goodness-of-fit between modeled values and measured values for the first-dimension retention times (t_1) and the second-dimension retention times (t_2) of the 56 calibration analytes.

$$\log L_{2,i}(121°C) = 0.2613N_i^* + \log\left(\frac{t_{2,i} - \alpha_3}{t_{2,i}^* - \alpha_3}\right) - 0.557$$

$$+ \epsilon_{i,2} \tag{4}$$

where N_i^* represents the fractional number of carbon atoms of a hypothetical n-alkane (nP class) that would elute with the same first-dimension retention time as analyte i (eq S1), α_3 (s) represents a tunable model parameter that can be adjusted to enable a best-fit with observed second-dimension retention time values for a given GCxGC instrument program, $\epsilon_{i,2}$ represents the discrepancy of the fitted $\log L_{2,i}(121°C)$ value by eq 4 for analyte i, $t_{2,i}$ (s) is the second-dimension retention time of analyte i, and $t_{2,i}^*$ represents the interpolated second-dimension retention time of the hypothetical n-alkane that would elute with the same first-dimension retention time as analyte i (eq S2).

Two numerical constants, 0.2613 and −0.557, appear in eq 4. These constants result from the linear free energy relationship:

$$\log L_2^*(121°C) = 0.2613N^* - 0.557 \tag{5}$$

where $L_2^*(121°C)$ represents the gas−liquid partition coefficient of an n-alkane having N^* carbons on stationary phase 2 (50% phenyl polysilphenylene-siloxane) at 121 °C. The numerical slope and intercept of eq 5 were determined based on the application of eq 2 to a homologous series of n-alkanes.[36] The constants in eq 5 must be re-derived if the second-dimension column would contain a different type of stationary phase than that considered here.

In practice, the values of $\log L_{2,i}$ exhibit high correlation with $\log L_{1,i}$ for hydrocarbon structures that span a wide range of carbon numbers (Section S9). To facilitate robust calibrations of the tunable model parameters, α_1, α_2, and α_3, we defined a new quantity, $\Delta \log L_{21,i}$, which expresses the variability of $\log L_{2,i}$ independently of $\log L_{1,i}$:

$$\Delta \log L_{21,i} = \log L_{2,i} - \beta \log L_{1,i} \tag{6}$$

where β was determined by Gram-Schmidt orthogonalization of the 15,495 structures of the hydrocarbon library. The resulting β value of 1.07 is comparable to the value of 1.14 reported previously[36] for a set of nonpolar halogenated and non-halogenated hydrocarbon compounds. Use of eq 6 improves the robustness of the implementation of the non-linear regression of eq 4. However, the value of β does not otherwise affect the results of the elution model.

Fitting the GCxGC Elution Model with Measured Retention Times of "Calibration Analytes". To adapt the

elution model to a particular GCxGC instrument program, the user optimizes the three calibration parameters (α_1, α_2, and α_3) by a non-linear regression of eqs 3 and 4 that minimizes the sum-of-square of the deviations ($\epsilon_{i,1}$ and $\epsilon_{i,2}$) between the fitted and theoretical values of $L_{1,i}$ and $L_{2,i}$ for a set of known analytes. By tuning α_1, α_2, and α_3 to optimized values, the user calibrates the elution model to a particular GCxGC instrument program (i.e., oven temperature program, inlet pressure program, and column dimensions). To test the calibration procedure, we employed previously reported[3] two-dimensional retention time data for 56 hydrocarbon standards which had been measured experimentally using two different sets of GCxGC instrument conditions, referred to as method A and method B (see Sections S7 and S8). We refer to these known chemical structures as "calibration analytes" (Table S1), which spanned the n-C$_{10}$ to n-C$_{24}$ elution window and encompassed: 15 n-alkanes (nP class), 10 linear alkyl cyclohexanes (mN class), 11 linear alkyl benzenes (mAr class), biphenyl and 5 alkyl-substituted naphthalenes (dAr class), 11 parent and alkyl-substituted ≥3-ring polycyclic aromatics (polyAr class), tetralin (NmAr class), decalin (dN class), and acenaphthene and fluorene (NdAr class).

Simulation of the GCxGC Retention Times of Hydrocarbons in the Library with the Calibrated Elution Model. We applied the theoretical L_1 and L_2 values (obtained in Step 2) of the 15,495 hydrocarbons to the elution model that was calibrated in Step 4. This enabled a prediction of two-dimensional retention times for the hydrocarbons, assuming the GCxGC instrument conditions used in the trial calibration. We predicted the first-dimension retention time of each analyte i by rearranging eq S1 to obtain the following:

$$t_{1,i}^{pred} = t_{1,N}^* + (N_i^* - N^*)(t_{1,N+1}^* - t_{1,N}^*) \tag{7}$$

where N_i^* is determined by inverting eq 3 (shown by eq S3), N^* represents the (integer) carbon number of the n-alkane that would elute just prior to the analyte, and $t_{1,N}^*$ and $t_{1,N+1}^*$ represent the measured retention times of the n-alkanes that elute immediately before and after analyte i, respectively. We predicted the second-dimension retention time of each analyte i by rearranging eq 4:

$$t_{2,i}^{pred} = \alpha_3 + (t_{2,i}^* - \alpha_3)10^{(\log L_{2,i}(121°C) - 0.2613N_i^* + 0.557)} \tag{8}$$

where $t_{2,i}^*$ is determined by eq S2, using values of $t_{1,i}$ obtained by eq 7.

RESULTS

Trial Calibrations of the Elution Model with Two GCxGC Instrument Methods. In two-candidate trial
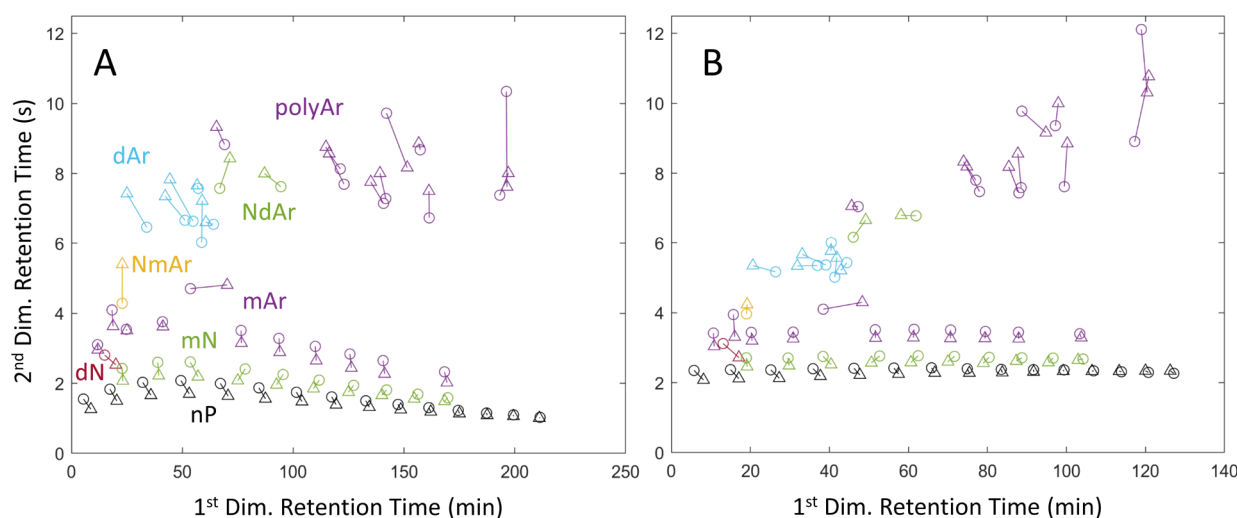
Figure 2. Fit of the elution model with measured GCxGC retention time data for 56 calibration analytes spanning the n-C_{10}−n-C_{24} elution window. Triangles depict measured retention time data. Circles show elution model retention times after the calibration procedure. Line segments connect the measured and modeled values of each individual calibration analyte. Symbol color corresponds to chemical class (Table S1), annotated in Panel A. See Methods for chemical class abbreviations. (A) Instrument method A. (B) Instrument method B.

calibrations, we adapted the elution model to each of two different instrument methods (A and B) based on previously reported experimental two-dimensional retention time data (t_1 and t_2) for 56 calibration analytes (Table S1). By applying non-linear regression to the elution model eqs 3 and 4, we successfully optimized the calibration parameters, α_1, α_2, and α_3 (Table 1), finding robust model calibrations for both instrument methods A and B (see Section S9). The calibrated elution model largely captures the variability in the experimental two-dimensional retention time data for the 56 calibration analytes, producing squared correlation coefficients of $r^2 > 0.99$ for first-dimension retention times and $r^2 > 0.95$ for second-dimension retention times (Table 1). The calibration results further confirm that the model can flexibly mimic the experimentally observed distortion in the overall elution pattern that arises in instrument method A versus method B (Figure 2).

The elution model correctly predicts the experimentally observed separation of the calibration analytes into clusters of different hydrocarbon classes (Figure 2). The model also correctly assigns the region in which each analyte is expected to elute: the modeled retention time values of calibration analytes exhibited root-mean-squared-errors (RMSEs) of 2.6−4.3 min in the first dimension and 0.50−0.63 s in the second dimension, with respect to the corresponding experimental values (Table 1). These RMSE values represent relative errors of 2.1−2.2% in the first-dimension retention time and 5.8−7.6% in the second-dimension retention time. The model accuracy does not match the resolving power of the GCxGC instrument: for example, the model does not always correctly predict the elution order of well-resolved but near-neighboring analytes. However, the trial calibrations suggest that the elution model may successfully differentiate groups of analytes by their class and carbon number.

Among the calibration analytes, the elution model goodness-of-fit varied according to the type of chemical structure: the lowest deviations in second-dimension retention time (t_2) were found for analytes of the nP, mN, and dN classes and linear alkyl benzene (mAr class), whereas larger deviations arose for analytes of the dAr, polyAr, NmAr, NdAr classes, and

hexamethylbenzene (mAr class). This outcome arises partly because the model regression of α_3 minimizes the residuals of log L_2 values, which are related to t_2 values by the transformation shown in eqs 4 and 8. Consequently, the best-fit model tolerates larger residuals (on average) for analytes having higher t_2 values than those with lower t_2 values. We justify this implementation by the fact that we desire better accuracy for the t_2 values of the less-polar classes (which exhibit low t_2 values) than the more-polar classes (which exhibit higher t_2 values), since the less-polar classes exhibit tighter separation in practice. In addition, we anticipate that the elution model may predict the retention times of structurally simple compounds (e.g., mono-cyclic, single-substituted) more accurately than those of structurally complex compounds (e.g., multi-cyclic aromatic, multiple-substituted) based on the limitations of the UFZ-LSER tool, which was used to estimate the model input parameters L_1 and L_2. Members of the iP, polyN, and NpolyAr classes were not tested in these calibration trials due to lack of available data. The elution model also has not been validated for C_{25}-C_{30} hydrocarbons.

Finally, we can verify that the model-fitted values of the calibration parameters, α_1, α_2, and α_3, appear physically reasonable (Table 1). The values of α_1 and α_2 are well-constrained, and they represent the slope and intercept of the expected positive linear relationship between the log L_1 value of the analyte and the carbon number of a hypothetical n-alkane (N_i^*) that would elute with the same first-dimension retention time (eq 3). The GCxGC instrument method A produced calibrated values of α_1 and α_2 that are similar to those found for instrument method B, consistent with the expectation that the relationship between log L_1 and N_i^* should not depend strongly on the choice of first-dimension temperature ramp (which differs by ×2.3 between the two instrument methods).

The third elution model parameter, α_3, represents the elution model's estimate of the second-dimension hold-up time, which is the duration of time required for the carrier gas to travel from the thermal modulator to the detector. Previous work has suggested that the observable column bleed line
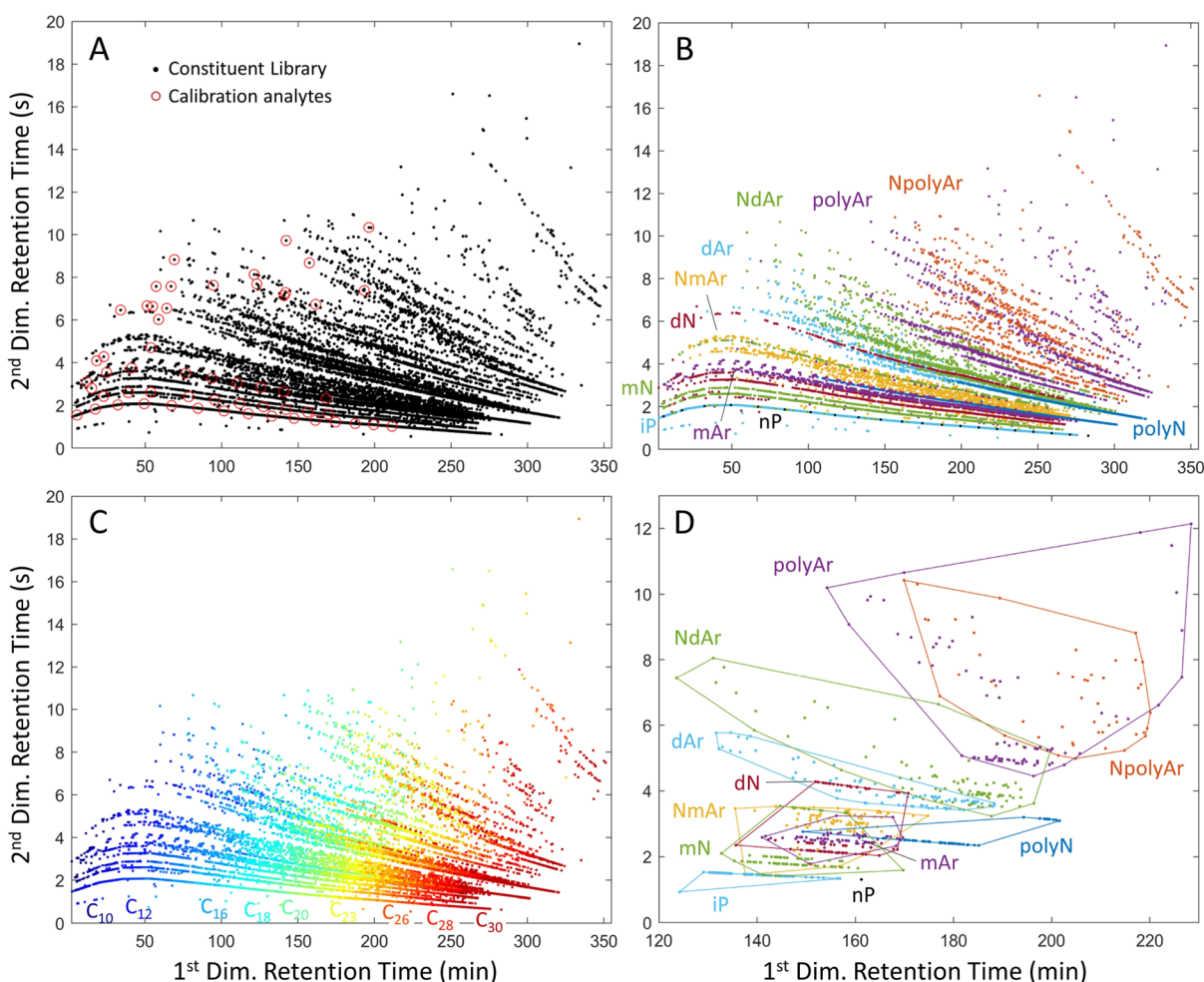
Figure 3. Simulated retention times of the hydrocarbon library with GCxGC instrument method A. (A) Black points depict predicted retention times of the 15,447 hydrocarbons in the constituent library that elute after 2 min, forming a characteristic triangular elution pattern. Red circles show model-fitted retention times of the chemical structures representing the 56 calibration analytes. (B) Predicted retention times of the 15,447 hydrocarbons with simulated first-dimension elution times >2 min, color-coded by hydrocarbon class and annotated. See Methods for chemical class abbreviations. (C) Same as panel B, color-coded by carbon number and annotated. (D) Predicted retention times of the hydrocarbons that represent the C_{20} constituents of the library, color-coded by hydrocarbon class and overlaid with polygons, which illustrate the elution regions of the 10 HCBs containing ≥3 structures (Sections S10 and S11).

provides a reasonable indicator of the second-dimension hold-up time,[3] which is found to be valid for the region of the chromatogram possessing second-dimension temperatures exceeding perhaps 100−110 °C, for an apolar/polar column combination.[52] Therefore, we can refer to the original chromatogram data and verify whether the observed location of the column bleed line is consistent with the model-fitted values of α_3. For GCxGC instrument method A, the elution model constrains an α_3 value of 0.0 ± 0.1, which appears physically unreasonable. However, this result is precisely consistent with the fact that the analyst chose to artificially offset the second-dimension retention times of the entire chromatogram by a constant value (also called "rotating" the second-dimension retention times), intentionally setting the observed column bleed line at $t_2 = 0$, after the chromatogram data were obtained with instrument method A.[3] For instrument method B, the analyst did not intervene with the chromatogram, and the originally produced second-dimension retention time data were available. In this case, the elution model calibration finds $\alpha_3 = 1.0 ± 0.1$, which is reasonably consistent with the observed location of the column bleed line

of 1.2 s throughout the part of the chromatogram having first-dimension retention times of >30 min (we neglect consideration of the initial 30 min of analysis, during which second-dimension elution temperatures ranged from 70 to 108 °C and the column bleed exhibited longer retention times of 1.3−1.7 s). Although the current version of the GCxGC elution model represents the hold-up time as a constant parameter, it is also possible to implement the hold-up time as a time-varying function based on experimental data or other considerations,[3] if such flexibility becomes needed.

Simulated GCxGC Chromatograms of the Hydrocarbon Library. After calibration, we employed the elution model to extrapolate the retention times of the entire hydrocarbon library, by applying the UFZ-LSER-estimated L_1 and L_2 values to the calibrated model equations (eqs 7 and 8, Figure 3, and Sections S3 and S10). In this way, we obtained simulated GCxGC retention time values for 15,447 of the 15,495 available hydrocarbons with instrument method A, and for 15,455 hydrocarbons with method B. The remaining 40−48 hydrocarbons were predicted as chromatographically unretained because they exhibited predicted first-dimension
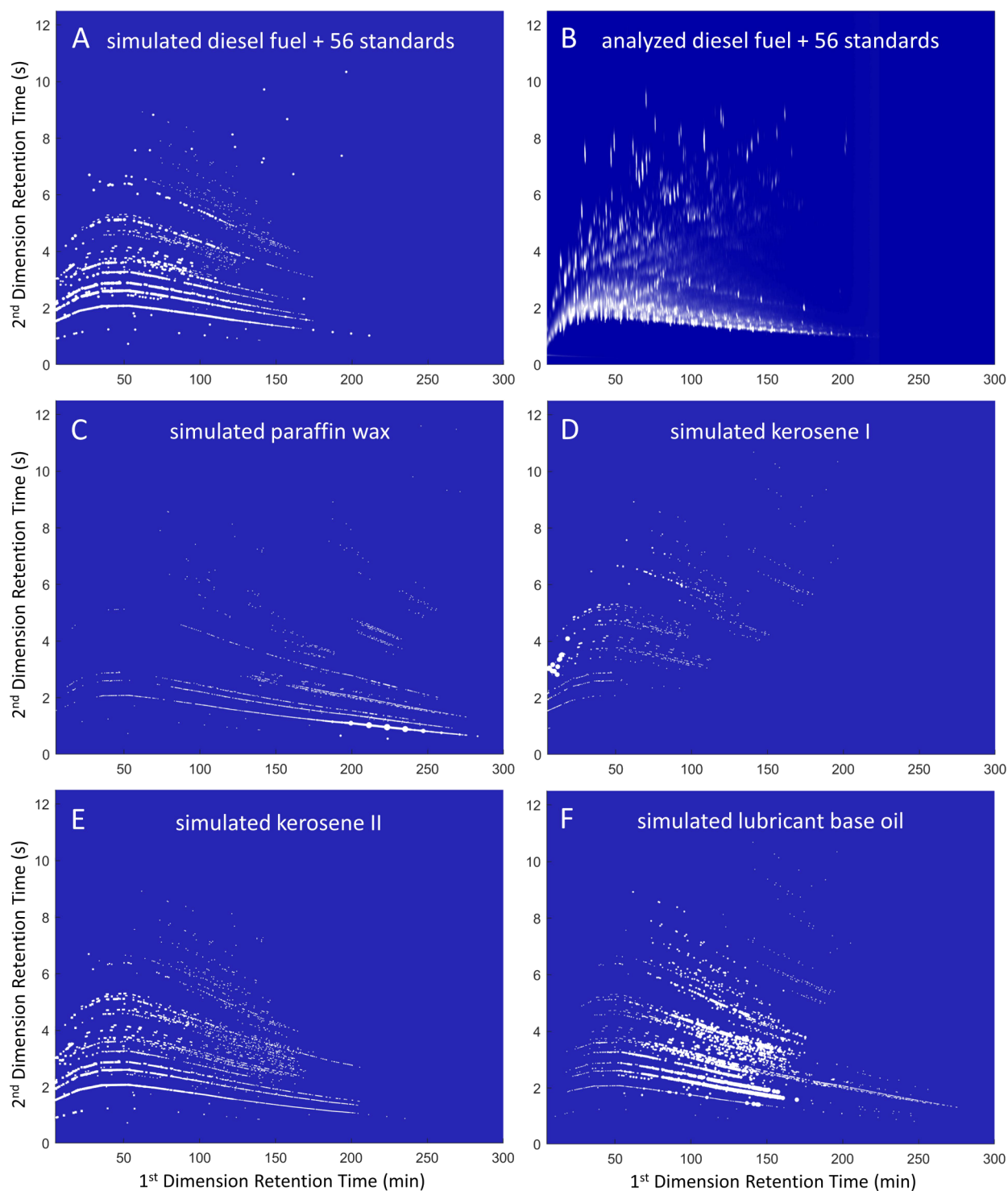
Figure 4. Simulated and observed GCxGC elution patterns of individual petroleum products using instrument method A. (A) Simulation of a diesel fuel amended with the 56 calibration analytes, based on reported Concawe HCB data for a diesel fuel.[54] The retention time of each simulated constituent is plotted as a white filled circle; the size of the symbol is approximately scaled by the reported concentration of the HCB containing that constituent. (B) Experimental GCxGC-FID chromatogram of a different diesel fuel sample amended with the same 56 standards, reported previously,[3] which can be compared with panel A. (C) Simulation of a paraffin wax, based on Concawe HCB data.[54] (D,E) Simulations of two different kerosene products, based on Concawe HCB data.[54] (F) Simulation of a lubricant base oil, based on Concawe HCB data.[54]

retention times of <2 min, which would typically imply poor separation and overlap with injected carrier solvent during a real GCxGC analysis. These unretained constituents were composed entirely of C_{10} isoparaffins, which would exhibit a low affinity to the first-dimension stationary phase relative to *n*-decane, according to the predicted L_1 values of the UFZ-LSER tool (the experimental $t_{1, \text{decane}}$ value was 8.75 min for GCxGC instrument method A). It is possible to devise other instrument programs that successfully retain and separate these compounds.

The elution model enables an unprecedented theoretical prediction of the GCxGC chromatogram for hydrocarbons spanning C_{10} to C_{30} (Figure 3 and S3). The simulated hydrocarbon library produces a characteristic triangle-shaped elution pattern resembling the GCxGC chromatogram of a typical light crude oil or middle distillate. Each class of hydrocarbons leads to a banded distribution of simulated retention times in the second dimension, consistent with the well-known observation that GCxGC produces structured chromatographic separations arranged by the chemical class. The simulated members within each class are further separated by carbon number along the first dimension, which is also consistent with experimental observations. The simulations of the hydrocarbon library suggest that the elution model could reasonably differentiate groups of hydrocarbon analytes by their class and carbon number, which is a precondition for allocating petroleum analytes to HCBs in real samples.

The simulated elution bands qualitatively resemble the trends and ordering of reported experimental elution bands of the nP, iP, mN, mAr, NmAr, dN, dAr, polyN, and polyAr classes, in the GCxGC chromatogram. However, these banded distributions are more broadly dispersed for some classes than for others. The elution model predicts that acyclic and mono-cyclic saturate classes (nP, iP, and mN) form successive tight elution bands in the early second dimension, comparable to experimentally observed elution bands for these classes.[2,4,11,18,21] Noisier bands of the dN, mAr, and polyN classes appear later in the second dimension, overlapping with broadly dispersed bands of the NmAr, dAr, NdAr, polyAr, and NpolyAr classes, according to simulations. These trends appear qualitatively consistent with the reported ordering of experimental GCxGC elution patterns of saturated, naphthenic, and aromatic hydrocarbon classes,[1,2,4,8,11,21] including the iP and polyN classes (which encompass many biomarkers), which were not represented in the model training set. Exceptionally, we did not find literature data on the observed elution behaviors of members of the NpolyAr class. Finally, we observe that polyaromatic hydrocarbon structures possess widely varying magnitudes of molecular polarizability and polarity,[53] which causes the polyaromatic hydrocarbons to become widely dispersed in the simulated GCxGC chromatogram, consistent with experimental observations. By comparison, the elution model predicts that hydrocarbon classes lacking aromaticity form much tighter elution bands, also consistent with experimental results. Overall, these findings inspire confidence that the elution model may improve efforts to locate and quantify poorly characterized classes, such as naphthenic constituents that are difficult to separate and identify.

Modeling GCxGC Elution Patterns of Individual Petroleum Products. We simulated the GCxGC elution patterns of individual petroleum product samples, by projecting previously reported HCB compositional data[54] onto the hydrocarbon library. We compared a simulated diesel fuel (which we amended with the 56 calibration analytes; Figure 4A) to a previously reported GCxGC chromatogram[3] of an experimentally-analyzed diesel fuel amended with the same 56 standards (Figure 4B). The HCBM data used to populate the simulated diesel fuel composition are derived from a diesel product that was different from the experimentally analyzed diesel sample, yet the simulated and experimental elution patterns exhibit reasonable visual agreement with one another. This example illustrates how we can use the elution model to predict the GCxGC chromatogram of a hypothetical product in the absence of available experimental GCxGC data, based on known or inferred HCB data. In further simulations, we employ the reported HCBM data from other product samples[54] to predict and visualize the GCxGC elution patterns of a paraffin wax, two different kerosene products, and a lubricant base oil (Figures 4C−F). The simulated elution patterns of the latter four products could not be compared to the original experimental GCxGC data, which were unavailable. These examples illustrate the remarkably diverse compositions of these petroleum products, as judged by their simulated chromatograms. The results also support the expectation that analysts could employ the elution model to meaningfully interpret the compositional information in experimental GCxGC chromatograms of real petroleum products.

■ DISCUSSION AND IMPLICATIONS

The risk assessment community lacks cost-effective and systematic tools to interpret petroleum substance compositions in commercial products and environmental samples. The present work addresses this need by establishing a relationship between GCxGC retention time and hydrocarbon structure. The members of each hydrocarbon class separate into distinct bands along the second dimension of the simulated chromatogram, and members within each class are further separated by carbon number along the first dimension. Hydrocarbon class and carbon number information exhibit pronounced associations with retention time in the simulated GCxGC chromatogram. These relationships could be empirically encoded with nonparametric data analysis methods such as kernel density or nearest-neighbor algorithms.[55,56] The present work thus offers a means to broadly estimate the chemical class and carbon number of the analyzed petroleum substance constituents based on their GCxGC retention times. When further employed with GCxGC coupled to time-of-flight mass spectrometry, the elution model could enable a powerful three-pronged approach to interpret analyzed constituent structures based on two-dimensional retention times and mass spectral information. These innovations can broadly advance in silico capabilities to interpret petroleum substance compositions by GCxGC analysis, which is otherwise costly and reliant on expert judgment.

By creating a path to estimate hydrocarbon class and carbon number information throughout the GCxGC chromatogram of a petroleum substance, the present work provides a technical basis to allocate experimental analytes to HCBs. This capability would enable experts to apply the HCB method to environmental samples, which is currently not possible, thereby opening opportunities to evaluate the environmental properties of HCBs. A transparent HCB method would also permit experts to develop representative constituent compositions for HCBs and assess similarities among constituents within HCBs, which are needed to address regulatory requirements. Such efforts are currently confounded by the non-transparency of the existing proprietary HCB methods. By contrast, the elution model forges a transparent and reproducible path to HCB delineation, which is better suited to risk assessment.

These new capabilities would prove especially useful to investigate the poorly characterized fractions that degrade slowly in the environment, such as the naphthenic, naphthenic-aromatic, and substituted-aromatic constituents. Previous

studies have characterized the biodegradation behaviors of thousands of petroleum substance analytes by GCxGC,[7,34,35] but the investigators lacked methods to determine the analyte identities. With a systematic basis to interpret the compositional information in the GCxGC chromatogram, we can better elucidate the broad spectrum of compositional changes affecting petroleum substances during biodegradation. This would address a critical information gap in ongoing efforts to assess the environmental persistence properties of petroleum substances, including the development and testing of biodegradation prediction models.[57]

The present work formalizes the relationship between UVCB constituent libraries and GCxGC analysis. The elution model provides a unique capability to predict the GCxGC chromatogram data for hypothesized constituents or subfractions of petroleum substances. Comparable laboratory experiments would be infeasible, because currently it is impossible to physically separate or concoct most of the individual constituents found in petroleum substances. By enabling comparisons between simulated compositions and experimental GCxGC chromatogram data for petroleum substances, the elution model provides a unique avenue to support the development and validation of UVCB constituent libraries. This represents an opportunity to advance risk assessment technology for UVCBs: regulatory agencies and industry are developing large libraries that contain thousands of hypothetical constituents to support UVCB risk assessment, but limited means are available to validate the realism of these libraries. In this regard, the GCxGC elution model is not restricted to hydrocarbons: it can be applied to any GCxGC-amenable chemical structure that falls within the model domain of the UFZ-LSER tool,[41,42] including structures that contain heteroatoms.[26,58]

Finally, this initial version of the elution model is an intentionally parsimonious implementation, aiming to provide robustness at the cost of predictive accuracy. Future efforts to upgrade the model could include improved methods to estimate the LSER parameters, expressions to represent changes in the heat capacity gas-stationary phase transfer, and expressions or data to represent possible changes in the gas hold-up time[3] and improvements to the hydrocarbon library.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.est.2c06922.

Overview of GCxGC; operational definition and practical precision of experimental GCxGC-FID retention times; overview of using GCxGC data for HCB methods; auxiliary equations of the GCxGC elution model; assumptions, limitations, and domain of the GCxGC elution model; description of the development of the hydrocarbon library; GCxGC instrument methods; parameters used for construction and calibration of the GCxGC elution model; fitted values of L_1 and L_2, and determination of β during trial calibrations of the elution model; further explanation of application of the calibrated elution model to the hydrocarbon library structures; and frequency of overlaps among HCBs and among hydrocarbon class regions in the simulated GCxGC chromatograms (PDF)

Number of hydrocarbon library structures representing each class and carbon number. The complete hydrocarbon library can be obtained from Concawe at the provided correspondence address (XLSX)

AUTHOR INFORMATION

Corresponding Author

Delina Lyon — *Concawe, B-1160 Brussels, Belgium;* orcid.org/0000-0002-2291-9048; Email: environment@concawe.eu

Authors

J. Samuel Arey — *ExxonMobil Biomedical Sciences Inc., Annandale, New Jersey 08801, United States;* Present Address: Oleolytics LLC, State College, Pennsylvania 16801, United States

Alberto Martin Aparicio — *Concawe, B-1160 Brussels, Belgium;* Present Address: Penman Consulting, B-1040 Brussels, Belgium

Eleni Vaiopoulou — *Concawe, B-1160 Brussels, Belgium;* orcid.org/0000-0001-5498-7400

Stuart Forbes — *SF Analyticon Limited, Hartford CW8 2AQ, U.K.*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.est.2c06922

ABBREVIATIONS

UVCB	unknown or variable composition complex reaction products or biological materials
HCB	hydrocarbon block
GCxGC	comprehensive two-dimensional gas chromatography
FID	flame ionization detector
LSER	linear solvation energy relationships
UFZ	Helmholtz Centre for Environmental Research
SMILES	simplified molecular-input line-entry system
nP	normal paraffins
iP	iso-paraffins
mN	mono-cyclic naphthenes
dN	di-cyclic naphthenes
polyN	polycyclic naphthenes having ≥3 rings
mAr	mono-cyclic aromatics
dAr	di-cyclic aromatics
polyAr	polycyclic aromatics having ≥3 rings
NmAr	naphthenic mono-cyclic aromatics

NdAr	naphthenic di-cyclic aromatics
NpolyAr	naphthenic polycyclic aromatics
SE-30	100% dimethylpolysiloxane (Agilent)
OV-17	50% diphenyl 50% dimethylpolysiloxane (Ohio Valley)
SGE BPX50	50% phenyl polysilphenylene-siloxane (Trajan)
RMSE	root mean squared error

■ REFERENCES

(1) Reddy, C. M.; Eglinton, T. I.; Hounshell, A.; White, H. K.; Xu, L.; Gaines, R. B.; Frysinger, G. S. The West Falmouth oil spill after thirty years: the persistence of petroleum hydrocarbons in marsh sediments. *Environ. Sci. Technol.* **2002**, *36*, 4754−4760.

(2) Frysinger, G. S.; Gaines, R. B.; Xu, L.; Reddy, C. M. Resolving the unresolved complex mixture in petroleum-contaminated sediments. *Environ. Sci. Technol.* **2003**, *37*, 1653−1662.

(3) Arey, J. S.; Nelson, R. K.; Xu, L.; Reddy, C. M. Using comprehensive two-dimensional gas chromatography retention indices to estimate environmental partitioning properties for a complete set of diesel fuel hydrocarbons. *Anal. Chem.* **2005**, *77*, 7172−7182.

(4) Nelson, R. K.; Kile, B. M.; Plata, D. L.; Sylva, S. P.; Xu, L.; Reddy, C. M.; Gaines, R. B.; Frysinger, G. S.; Reichenbach, S. E. Tracking the weathering of an oil spill with comprehensive two-dimensional gas chromatography. *Environ. Forensics* **2006**, *7*, 33−44.

(5) Arey, J. S.; Nelson, R. K.; Reddy, C. M. Disentangling oil weathering using GC× GC. 1. Chromatogram analysis. *Environ. Sci. Technol.* **2007**, *41*, 5738−5746.

(6) Arey, J. S.; Nelson, R. K.; Plata, D. L.; Reddy, C. M. Disentangling oil weathering using GC× GC. 2. Mass transfer calculations. *Environ. Sci. Technol.* **2007**, *41*, 5747−5755.

(7) Wardlaw, G. D.; Arey, J. S.; Reddy, C. M.; Nelson, R. K.; Ventura, G. T.; Valentine, D. L. Disentangling oil weathering at a marine seep using GC× GC: Broad metabolic specificity accompanies subsurface petroleum biodegradation. *Environ. Sci. Technol.* **2008**, *42*, 7166−7173.

(8) Mao, D.; Lookman, R.; Van De Weghe, H.; Vanermen, G.; De Brucker, N.; Diels, L. Aqueous solubility calculation for petroleum mixtures in soil using comprehensive two-dimensional gas chromatography analysis data. *J. Chromatogr. A* **2009**, *1216*, 2873−2880.

(9) Mao, D.; Lookman, R.; Van De Weghe, H.; Van Look, D.; Vanermen, G.; De Brucker, N.; Diels, L. Detailed analysis of petroleum hydrocarbon attenuation in biopiles by high-performance liquid chromatography followed by comprehensive two-dimensional gas chromatography. *J. Chromatogr. A* **2009**, *1216*, 1524−1527.

(10) Gros, J.; Nabi, D.; Würz, B.; Wick, L. Y.; Brussaard, C. P.; Huisman, J.; van der Meer, J. R.; Reddy, C. M.; Arey, J. S. First day of an oil spill on the open sea: Early mass transfers of hydrocarbons to air and water. *Environ. Sci. Technol.* **2014**, *48*, 9400−9411.

(11) Gros, J.; Reddy, C. M.; Aeppli, C.; Nelson, R. K.; Carmichael, C. A.; Arey, J. S. Resolving biodegradation patterns of persistent saturated hydrocarbons in weathered oil samples from the Deepwater Horizon disaster. *Environ. Sci. Technol.* **2014**, *48*, 1628−1637.

(12) Nelson, R. K.; Aeppli, C.; Arey, J. S.; Chen, H.; de Oliveira, A. H.; Eiserbeck, C.; Frysinger, G. S.; Gaines, R. B.; Grice, K.; Gros, J.; Hall, G. J.; Koolen, H. H. F.; Lemkau, K. L.; McKenna, A. M.; Reddy, C. M.; Rodgers, R. P.; Swarthout, R. F.; Valentine, D. L.; White, H. K.Applications of comprehensive two-dimensional gas chromatography (GC× GC) in studying the source, transport, and fate of petroleum hydrocarbons in the environment. In *Standard Handbook Oil Spill Environmental Forensics*; Elsevier, 2016; pp 399−448.

(13) Swarthout, R. F.; Gros, J.; Arey, J. S.; Nelson, R. K.; Valentine, D. L.; Reddy, C. M.Comprehensive two-dimensional gas chromatography to assess petroleum product weathering. In *Hydrocarbon and Lipid Microbiology Protocols*; Springer, 2016; pp 129−149.

(14) Gros, J.; Reddy, C. M.; Nelson, R. K.; Socolofsky, S. A.; Arey, J. S. Simulating gas−liquid− water partitioning and fluid properties of petroleum under pressure: implications for deep-sea blowouts. *Environ. Sci. Technol.* **2016**, *50*, 7397−7408.

(15) Nabi, D.; Arey, J. S. Predicting partitioning and diffusion properties of nonpolar chemicals in biotic media and passive sampler phases by GC× GC. *Environ. Sci. Technol.* **2017**, *51*, 3001−3011.

(16) Liu, Z.; Phillips, J. B. Comprehensive two-dimensional gas chromatography using an on-column thermal modulator interface. *J. Chromatogr. Sci.* **1991**, *29*, 227−231.

(17) Blomberg, J.; Schoenmakers, P. J.; Beens, J.; Tijssen, R. Compehensive two-dimensional gas chromatography (GC× GC) and its applicability to the characterization of complex (petrochemical) mixtures. *J. High Resolut. Chromatogr.* **1997**, *20*, 539−544.

(18) Phillips, J. B.; Beens, J. Comprehensive two-dimensional gas chromatography: a hyphenated method with strong coupling between the two dimensions. *J. Chromatogr. A* **1999**, *856*, 331−347.

(19) Frysinger, G. S.; Gaines, R. B. Separation and identification of petroleum biomarkers by comprehensive two-dimensional gas chromatography. *J. Sep. Sci.* **2001**, *24*, 87−96.

(20) Dallüge, J.; Beens, J.; Udo, A. Comprehensive two-dimensional gas chromatography: a powerful and versatile analytical tool. *J. Chromatogr. A* **2003**, *1000*, 69−108.

(21) Wang, F. C.-Y.; Qian, K.; Green, L. A. GC× MS of diesel: a two-dimensional separation approach. *Anal. Chem.* **2005**, *77*, 2777−2785.

(22) Tong, H.; Karasek, F. Flame ionization detector response factors for compound classes in quantitative analysis of complex organic mixtures. *Anal. Chem.* **1984**, *56*, 2124−2128.

(23) Scanlon, J. T.; Willis, D. E. Calculation of flame ionization detector relative response factors using the effective carbon number concept. *J. Chromatogr. Sci.* **1985**, *23*, 333−340.

(24) Reichenbach, S. E.; Ni, M.; Zhang, D.; Ledford, E. B., Jr. Image background removal in comprehensive two-dimensional gas chromatography. *J. Chromatogr. A* **2003**, *985*, 47−56.

(25) Koolen, H. H.; Swarthout, R. F.; Nelson, R. K.; Chen, H.; Krajewski, L. C.; Aeppli, C.; McKenna, A. M.; Rodgers, R. P.; Reddy, C. M. Unprecedented insights into the chemical complexity of coal tar from comprehensive two-dimensional gas chromatography mass spectrometry and direct infusion fourier transform ion cyclotron resonance mass spectrometry. *Energy Fuels* **2015**, *29*, 641−648.

(26) Nelson, R. K.; Gosselin, K. M.; Hollander, D. J.; Murawski, S. A.; Gracia, A.; Reddy, C. M.; Radović, J. R. Exploring the complexity of two iconic crude oil spills in the gulf of Mexico (Ixtoc I and Deepwater Horizon) Using comprehensive two-dimensional gas chromatography (GC× GC). *Energy Fuels* **2019**, *33*, 3925−3933.

(27) Salvito, D.; Fernandez, M.; Jenner, K.; Lyon, D. Y.; de Knecht, J.; Mayer, P.; MacLeod, M.; Eisenreich, K.; Leonards, P.; Cesnaitis, R.; León-Paumen, M.; Embry, M.; Déglin, S. E. Improving the environmental risk assessment of substances of unknown or variable composition, complex reaction products, or biological materials. *Environ. Toxicol. Chem.* **2020**, *39*, 2097−2108.

(28) Redman, A. D.; Parkerton, T. F.; Comber, M. H.; Paumen, M. L.; Eadsforth, C. V.; Dmytrasz, B.; King, D.; Warren, C. S.; den Haan, K.; Djemel, N. PETRORISK: A risk assessment framework for petroleum substances. *Integr. Environ. Assess. Manage.* **2014**, *10*, 437−448.

(29) Redman, A. D.; Parkerton, T. F.; Leon Paumen, M.; Butler, J. D.; Letinski, D. J.; den Haan, K. A re-evaluation of PETROTOX for predicting acute and chronic toxicity of petroleum substances. *Environ. Toxicol. Chem.* **2017**, *36*, 2245−2252.

(30) McGrath, J. A.; Fanelli, C. J.; Di Toro, D. M.; Parkerton, T. F.; Redman, A. D.; Paumen, M. L.; Comber, M.; Eadsforth, C. V.; den Haan, K. Re-evaluation of target lipid model−derived HC5 predictions for hydrocarbons. *Environ. Toxicol. Chem.* **2018**, *37*, 1579−1593.

(31) Paumen, M. L.; Lampi, M. A.; Comber, M. H. I.; Djemel, N.; Linington, S.; Villalobos, S. A.; Eadsforth, C. V.; den Haan, K.; Parkerton, T. F.; Redman, A. D.*An evaluation of the Persistence, Bioaccumulation, and Toxicity of Petroleum Hydrocarbons (revised)*; Brussels, Belgium, 2019.

(32) King, D.; Lyne, R.; Girling, A.; Peterson, D.; Stephenson, R.; Short, D. Environmental risk assessment of petroleum substances: The hydrocarbon block method. *Report* **1996**, *96*, 52.

(33) Paumen, M. L.; Comber, M.; Eadsforth, C. V.; den Haan, K.; Linington, S.; Redman, A. D.; Parkerton, T.The hydrocarbon block method for environmental risk assessment of petroleum substances. In *SETAC Europe 25th Annual Meeting*; Barcelona, May 3−7, 2015.

(34) Arey, J. S.; Gros, J.; Redman, A. D.; Bleich, A.; Letinski, D. J.; Conelly, M.; Nelson, R. K.Quantifying biodegradation rates of thousands of petroleum constituents in seawater by GCxGC-FID. In *SETAC North America 41st Annual Meeting (SciCon2 Virtual Meeting)*, Nov 15−19, 2020.

(35) Sørenson, L.; Creese, M.; Arey, J. S.; Brakstad, O. G.; Ribicic, D.; Gros, J.; Lyon, D.; Redman, A. D.; Camenzuli, L.; Wang, N.; Martin, A. A.; Booth, A. M.Biodegradation assessment of UVCBs: Two-dimensional gas chromatography with peak tracking of two petroleum substances for mapping biodegradation half-lives. In *SETAC Europe 32nd Annual Meeting*; Copenhagen, Denmark, May 15−19, 2022.

(36) Nabi, D.; Gros, J.; Dimitriou-Christidis, P.; Arey, J. S. Mapping environmental partitioning properties of nonpolar complex mixtures by use of GC× GC. *Environ. Sci. Technol.* **2014**, *48*, 6814−6826.

(37) *Advice on using read-across for UVCB substances: Obligations arising from Commission Regulation 2021/979, amending REACH annexes*; European Chemicals Agency (ECHA): Helsinki, Finland, 2022.

(38) Arey, J. S., *GC×GC elution model*; 2022. https://github.com/concawe/GCxGC-Elution-Model. (accessed November 2, 2022).

(39) Dimitrov, S. D.; Georgieva, D. G.; Pavlov, T. S.; Karakolev, Y. H.; Karamertzanis, P. G.; Rasenberg, M.; Mekenyan, O. G. UVCB substances: methodology for structural description and application to fate and hazard assessment. *Environ. Toxicol. Chem.* **2015**, *34*, 2450−2462.

(40) Kutsarova, S. S.; Yordanova, D. G.; Karakolev, Y. H.; Stoeva, S.; Comber, M.; Hughes, C. B.; Vaiopoulou, E.; Dimitrov, S. D.; Mekenyan, O. G. UVCB substances II: Development of an endpoint-nonspecific procedure for selection of computationally generated representative constituents. *Environ. Toxicol. Chem.* **2019**, *38*, 682−694.

(41) Ulrich, N., Endo, S., Brown, T. N., Watanabe, N., Bronner, G., Abraham, M. H., Goss, K.-U.*UFZ-LSER database v 3.2.1*; Helmholtz Centre for Environmental Research-UFZ; Leipzig, Germany, 2017. http://www.ufz.de/lserd. (accessed November 2, 2022).

(42) Brown, T. N.; Arnot, J. A.; Wania, F. Iterative fragment selection: A group contribution approach to predicting fish biotransformation half-lives. *Environ. Sci. Technol.* **2012**, *46*, 8253−8260.

(43) Abraham, M. H.; Poole, C. F.; Poole, S. K. Classification of stationary phases and other materials by gas chromatography. *J. Chromatogr. A* **1999**, *842*, 79−114.

(44) Arey, J. S.; Green, W. H.; Gschwend, P. M. The electrostatic origin of Abraham's solute polarity parameter. *J. Phys. Chem. B* **2005**, *109*, 7564−7573.

(45) Seeley, J. V.; Seeley, S. K. Model for predicting comprehensive two-dimensional gas chromatography retention times. *J. Chromatogr. A* **2007**, *1172*, 72−83.

(46) Karolat, B.; Harynuk, J. Prediction of gas chromatographic retention time via an additive thermodynamic model. *J. Chromatogr. A* **2010**, *1217*, 4862−4867.

(47) McGinitie, T. M.; Harynuk, J. J. Prediction of retention times in comprehensive two-dimensional gas chromatography using thermo-dynamic models. *J. Chromatogr. A* **2012**, *1255*, 184−189.

(48) Barcaru, A.; Anroedh-Sampat, A.; Janssen, H.-G.; Vivó-Truyols, G. Retention time prediction in temperature-programmed, comprehensive two-dimensional gas chromatography: Modeling and error assessment. *J. Chromatogr. A* **2014**, *1368*, 190−198.

(49) Burel, A.; Vaccaro, M.; Cartigny, Y.; Tisse, S.; Coquerel, G.; Cardinael, P. Retention modeling and retention time prediction in gas chromatography and flow-modulation comprehensive two-dimen-sional gas chromatography: The contribution of pressure on solute partition. *J. Chromatogr. A* **2017**, *1485*, 101−119.

(50) Jaramillo, R.; Dorman, F. L. Retention time prediction in thermally modulated comprehensive two-dimensional gas chromatog-raphy: Correcting second dimension retention time modeling error. *J. Chromatogr. A* **2018**, *1581*, 116−124.

(51) Stevenson, K. A.; Harynuk, J. J. Thermodynamics-based modelling of gas chromatography separations across column geo-metries and systems, including the prediction of peak widths. *J. Sep. Sci.* **2019**, *42*, 2013−2022.

(52) Klee, M. S.; Blumberg, L. M. Measurement of retention in comprehensive two-dimensional gas chromatography using flow modulation with methane dopant. *J. Chromatogr. A* **2010**, *1217*, 1830−1837.

(53) Abraham, M. H.; Chadha, H. S.; Whiting, G. S.; Mitchell, R. C. Hydrogen bonding. 32. An analysis of water-octanol and water-alkane partitioning and the Δlog P parameter of Seiler. *J. Pharm. Sci.* **1994**, *83*, 1085−1100.

(54) Forbes, S.*Concawe Substance Identification Group Analytical Program Report (Abridged Version)*; Brussels, Belgium, 2019. https://www.concawe.eu/wp-content/uploads/Concawe-Substance-Identification-Group-Analytical-Program-Report-Abridged-Version.pdf. (accessed November 2, 2022).

(55) Parzen, E. On estimation of a probability density function and mode. *Ann. Math. Stat.* **1962**, *33*, 1065−1076.

(56) Altman, N. S. An introduction to kernel and nearest-neighbor nonparametric regression. *Am. Stat.* **1992**, *46*, 175−185.

(57) Davis, C. W.; Camenzuli, L.; Redman, A. D. Predicting Primary Biodegradation of Petroleum Hydrocarbons in Aquatic Systems: Integrating System and Molecular Structure Parameters using a Novel Machine-Learning Framework. *Environ. Toxicol. Chem.* **2022**, *41*, 1359−1369.

(58) Blomberg, J.; Riemersma, T.; van Zuijlen, M.; Chaabani, H. Comprehensive two-dimensional gas chromatography coupled with fast sulphur-chemiluminescence detection: implications of detector electronics. *J. Chromatogr. A* **2004**, *1050*, 77−84.