Low-level fusion of FT-ICR MS data sets for the characterization of nitrogen and sulfur compounds in vacuum gas oils

Julie Guillemant, Alexandra Berlioz-Barbier, Florian Albrieux, Luis P de Oliveira, Marion Lacoue-Nègre, Jean-Francois Joly, Ludovic Duponchel

Low-level fusion of FT-ICR MS data sets for the characterization of nitrogen and sulfur compounds in vacuum gas oils

Julie Guillemant†, Alexandra Berlioz-Barbier*†, Florian Albrieux†, Luis P. de Oliveira†, Marion Lacoue-Nègre†, Jean-François Joly†, and Ludovic Duponchel*‡

† IFP Energies nouvelles, Rond-point de l'échangeur de Solaize, BP 3, 69360 Solaize, France

‡ Univ. Lille, CNRS, UMR 8516 - LASIR – Laboratoire de Spectrochimie Infrarouge et Raman, F-59000 Lille, France

ABSTRACT: Eighteen vacuum gas oils have been analyzed by Fourier transform ion cyclotron resonance mass spectrometry considering 6 replicates in three different ionization modes (ESI(+), ESI(-) and APPI(+)) to characterize the nitrogen and sulfur compounds contained in these samples. Classical data analysis has been first performed on generated data sets using DBE (Double Bond Equivalents) versus number of carbon atoms (#C) plots in order to observe similarities and differences within the nitrogen and sulfur-containing molecular classes from samples produced by different industrial processes. In a second step, three-way arrays have been generated for each ionization mode considering three dimensions: DBE related to aromaticity, number of carbon atoms related to alkylation and sample. These three-way arrays have then be concatenated using low-level data fusion strategy to obtain a new tensor with three new modes: aromaticity, alkylation, and sample. The PARAFAC method has then been applied for the first time to this three-way data structure. A two components decomposition has allowed us to highlight unique samples with unexpected reactivity behaviors throughout hydrotreatment. The obtained loadings led to the identification of the variables responsible for this specific character. This original strategy has provided a fast visualization tool able to highlight simultaneously the impact of the three ionization modes in order to explain the differences between the samples and compare them.

Rising demand for light petroleum cuts and middle distillates combined with lower crude oil quality and tougher product specifications impose refiners to modify their refining processes to meet market demands[1]. It is then more necessary than ever to find appropriate conversion processes in order to convert heavy cuts such as vacuum gas oils (VGOs) or vacuum residues (VRs) into commercially more attractive products[1]. Currently, the improvement of refining schemes including hydrotreatment and conversion processes suffers from a lack of analytical detail to describe such petroleum products. Only macroscopic properties are available to characterize the vacuum gas oils such as their nitrogen or sulfur elemental contents, their boiling points ranges or their densities[2]. This level of analytical detail is not sufficient as heavy cuts with similar or even identical macroscopic properties can behave completely different especially regarding hydrotreatment processes where specific compounds can induce catalytic deactivation[3]. Then, detailed knowledge of the exact chemical composition of VGO cuts is of premier importance and we are highly interested in the molecular characterization of the nitrogen and sulfur compounds contained in these complex samples. Hydrotreatment processes rely on sulfide-based catalysts that are unfortunately poisoned by basic nitrogen compounds[3–5]. Moreover, some neutral nitrogen compounds are refractory to hydrotreatment and compete with sulfur compounds for the hydrotreatment of the latter[6,7]. Some sulfur compounds are also known to be refractory whereas the amount of total sulfur in converted products is lower than 10 ppm as it is strictly regulated[8,9]. As a consequence, the efficiency of hydrodenitrogenation (hydrotreatment of nitrogen compounds, often noted HDN) and hydrodesulfurization (hydrotreatment of sulfur compounds, often noted HDS) processes is very dependent on both nitrogen and sulfur compounds compositions.

To get access to these compositions, ultra-high-resolution mass spectrometry such as Fourier transform ion cyclotron resonance mass spectrometry (noted FT-ICR MS) has been shown to be a very helpful tool providing fast and accurate chemical fingerprinting of heteroatomic compounds contained in crude oils[10–14]. Especially, the distributions of heteroatoms, rings, and double bonds can be obtained with this technique. Electrospray ionization (ESI) source is used for the characterization of basic and neutral nitrogen compounds as it is particularly useful for the analysis of polar compounds[15]. Atmospheric pressure photo-ionization (APPI) source is preferred for the characterization of aromatic sulfur compounds because APPI source is mainly efficient for aromatic compounds like sulfur-containing molecules[16]. However, the main benefit of FT-ICR MS analysis is also in some ways its main drawback: numerous peaks are detected and identified then providing thousands of molecular formulas of compounds present in the considered sample[17]. When complex samples have to be compared, standard data processing methods are often not sufficient to visualize simultaneously all generated data and explore fully the resulting FT-ICR MS analyses.

As a consequence, more and more studies using multivariate analysis are reported in the literature but none combined spectroscopic data acquired from different ionization sources

to go deeper in sample exploration[18–20]. In a previous study, a data formatting strategy has been introduced in order to simultaneously manage DBE values, number of carbon atoms and sample for the exploration of a single ionization mode (APPI(+)) focusing only on sulfur compounds[21]. In order to explore simultaneously the data obtained for basic/neutral nitrogen and sulfur compounds from several ionization modes, a low-level data fusion strategy will be applied to generate a new tensor from the three-way arrays obtained for each ionization mode. Multi-way approaches will be then required for this tensor exploration such as Parallel Factor Analysis (PARAFAC). It is often used to explore multi-dimensional data matrices mainly for excitation-emission fluorescence spectroscopy[22–24]. It has also been applied to GC/MS datasets considering three-way arrays with retention times, abundances of *m/z* ratios and samples being the investigated modes[25]. This method allows decomposing data into several components in order to identify the main variables that express much the characteristics of the samples. Thus, the aim of this study is to assess the relevance of PARAFAC to decompose FT-ICR MS data into three modes that are DBE (aromaticity degree), number of carbon atoms (alkylation degree) and sample which are the most used variables in petroleomics studies[26]. To do this, 18 vacuum gas oils were analyzed considering 6 technical replicates in three ionization modes. The molecular compositions of selected samples were compared through a classical data analysis thanks to DBE versus number of carbon atoms plots. A low-level data fusion method was performed to get a three-way array on which PARAFAC method was then applied to decompose data focusing on nitrogen and sulfur compounds characterization.

MATERIAL AND METHODS

18 vacuum gas oils samples were analyzed including five Straight Run Vacuum Gas oils (SRVGO), one Heavy Coker Gas Oil (HCGO), six vacuum gas oils from ebullating bed reactor (EBVGO), three Hydrotreated Vacuum Gas Oil (HDT) and three blends (i.e. mixtures of previous vacuum gas oils in different ratios). More details about these samples are given in Table 1. Different cuts of EBVGO samples have been considered such as short-cut (350-540°C) or long-cut (480-540°C) vacuum gas oils. Hydrotreated samples (HDT) originating from the same feed (VGO 8 in Table 1) but obtained from different catalysts have also been selected in this work.

Sample preparation and FT-ICR MS measurements

The physical characteristics, macroscopic properties and elementary compositions of the samples are given in Table 1. All samples have been first solubilized in toluene to 1% w/w and then diluted according to conditions presented in Table 2. Six technical replicates have been prepared and analyzed for each sample. Mass spectrometry analyses have been carried out using an LTQ FT Ultra FT-ICR MS instrument (ThermoFisher Scientific, Bremen, Germany) equipped with a 7 T superconducting magnet (Oxford Instruments). The ionization process has been performed using both the ESI source (ThermoFisher Scientific) in positive and negative modes and the APPI source (Syagen Technology, Tustin CA, USA) in positive mode. The mass range has been set to *m/z* 98-1000. Mass spectra have been acquired considering 4 µscans, 70 scans, resolution set to 200,000 (transient length of 1.6 s) at *m/z*=500 (center of average VGO samples mass distribution) and time-domain signals (transients) have been recorded. Ionization and ion transfer conditions are also available in Table 2. External mass calibration has been performed with a custom sodium formate clusters solution (sodium formiate from VWR, Fontenay-sous-Bois, France) covering the entire selected mass range (90-1000 Da).

Table 2: Ionization and ion transfer conditions for each ionization mode. Tol=Toluene, MeOH=Methanol, FA=Formic Acid, AmHy=Ammonium Hydroxide.

Parameter	ESI(+)	ESI(-)	APPI(+)
% dilution	0.05	0.05	0.05
% solvents mix	75%-25% Tol-MeOH	75%-25% Tol-MeOH	90%-10% Tol-MeOH
% additive	1% FA	1% AmHy	-
Spray voltage (kV)	3.5	3.3	-
Tube lens (V)	140	-140	70
Capillary voltage (V)	40	-40	30
Capillary temperature (°C)	375	375	275
Vaporization temperature (°C)	-	-	250
Sheath gas (a.u)	-	5	20
Auxiliary gas (a.u)	-	-	5
Flow rate (µL/min)	5	5	5

Data Processing

Data have been processed using several softwares. The full workflow has already been described in a previous work[15]. After using absorption mode, the observed resolutions were increased to about 400,000-500,000 allowing sulfur and nitrogen speciation. Molecular formula assignment conditions have been the following ones: $C_{0-100}H_{0-200}O_{0-5}N_{0-5}S_{0-5}$ with maximum content of heteroatoms in one molecular formula set to 5. Radical ions are identified and denoted as X families while protonated or deprotonated ions are identified and denoted as X[H] families[14]. Iterative mass recalibration has then been processed on the mass spectra with Peak-by-Peak® software (SpectroSwiss SARL, Lausanne, Switzerland) considering N1[H] family for analyses acquired with ESI ionization mode and S1 family for analyses performed using APPI ionization mode with a maximum mass error set to 1 ppm for all assignments[27]. Relative intensities for each N1[H] or S1 compound have been calculated as the peak intensity divided by the sum of intensities from all N1[H] or S1 peaks. Nitrogen and sulfur families have been attributed according to DBE values[15,21].

Chemometric approach

Neutral nitrogen compounds are identified in N1[H] class in ESI(-) mode, basic nitrogen compounds are identified in N1[H] class in ESI(+) mode and sulfur compounds are identified in S1 class in APPI(+) mode. Numerous peaks from other classes (HC, N1O1[H], N2[H]...) were also identified within each dataset but this work only focused on N1[H] and S1 classes which are highly problematic throughout hydrotreatment. The global methodology of data rearrangement used in this work is given in Figure 1. In a first step, a DBE vs carbon number plot (i.e. a 2D representation DBE=f(#C)) has been generated for each MS spectrum and each ionization mode considering relative intensities of the peaks. In a second step, for a given ionization mode, all

Table 1: Properties of vacuum gas oil samples used in this study. The ASTM standard methods used are mentioned for each property.

Sample	Type (*)	Density at 15°C *Ref. method: ASTM D4052*	Total sulfur (ppm) *Ref. method: ASTM D2622*	Total nitrogen (ppm) *Ref. method: ASTM D4629*	Basic nitrogen (ppm) *Ref. method: ASTM D2896*	Boiling point range (°C) *Ref. method: ASTM D86*
VGO 1	SRVGO	0.9375	28743	1300	316	384 – 570
VGO 2	SRVGO	0.9234	17711	1745	455	340 – 555
VGO 3	EBVGO	0.9375	3932	2800	910	348 – 545
VGO 4	HCGO	0.9551	28217	4190	1335	301 – 546
VGO 5	SRVGO	0.9355	22375	1755	501	375 – 555
VGO 6	MIX *50% VGO 2 + 50 % VGO 3*	0.9315	10776	2610	688	346 – 551
VGO 7	SRVGO	0.9120	3213	1191	337	431 – 559
VGO 8	SRVGO	0.9284	18921	1335	347	394 – 581
VGO 9	MIX *60% VGO 8 + 40% VGO 4*	0.9459	23131	2750	749	363 – 579
VGO 10	MIX *80% VGO 1 + 20% Safaniya GO*	0.9204	25739	1015	270	266 – 567
VGO 11	HDT *Effluent from VGO 8 Catalyst 1 – Low T°C*	0.9011	979	350	55	337 – 570
VGO 12	EBVGO	0.9318	4978	2740	741	362 – 519
VGO 13	EBVGO	0.906	5534	3122		421 – 551
VGO 14	EBVGO	0.9284	4691	3295	836	361 – 522
VGO 15	EBVGO	0.9215	5813	3627		448 – 547
VGO 16	EBVGO	0.9322	4725	2943	771	359 – 520
VGO 17	HDT *Effluent from VGO 8 Catalyst 2 – Low T°C*	0.9026	708	660	89	333 – 564
VGO 18	HDT *Effluent from VGO 8 Catalyst 2 – High T°C*	0.8985	358	675	71	269 – 557

(*): SRVGO = Straight Run Vacuum Gas Oil; HCGO = Heavy Coker Gas Oil; EBVGO = Vacuum Gas Oil from ebullating-bed reactor; MIX = blended Vacuum Gas Oil, HDT = Hydrotreated Vacuum Gas Oil, T°C = Temperature.

available DBE=f(#C) plots have been concatenated in order to form a 3D cube of size 100x25x108 containing the DBE vs carbon number plots of all samples. Here, the value of 100 stands for the range of the number of carbon atoms (from 1 to 100), 25 for the range of aromaticity values (from 1 to 25) and 108 for the number of acquired spectra on the whole data set (6 replicates for each of the 18 vacuum gas oil samples). If nothing is detected for a given DBE/number of carbon atoms pair, a zero value is assigned in the matrix. In the last step, prior to the multi-way analysis, a low-level data fusion strategy has been applied on datasets from each ionization mode including block scaling of each cube[28]. In other words, one hypercube of size 100x75x108 has been generated from the concatenation of the three available three-way arrays in order to explore them simultaneously. In order to facilitate data interpretation and variable identification, the three-way arrays have been merged along the DBE dimension as it is used for family assignments. PARAFAC (PARallel FACtor analysis) has been used to explore the generated hypercube[29]. This method is a generalization of PCA to higher-order arrays. As a consequence, it is particularly well-suited for the analysis of the generated hypercube[30]. The key principle of PARAFAC is similar to PCA when decomposing data into contributions along several components. However PARAFAC decomposes data into trilinear components. A PARAFAC model of a three-way array **X** is given by three loading matrices, **A**, **B**, and **C** with elements a_{if}, b_{jf}, and c_{kf}. The trilinear decomposition is then obtained from minimizing the sum of squares of the residuals, e_{ijk} in the model following Equation 1.

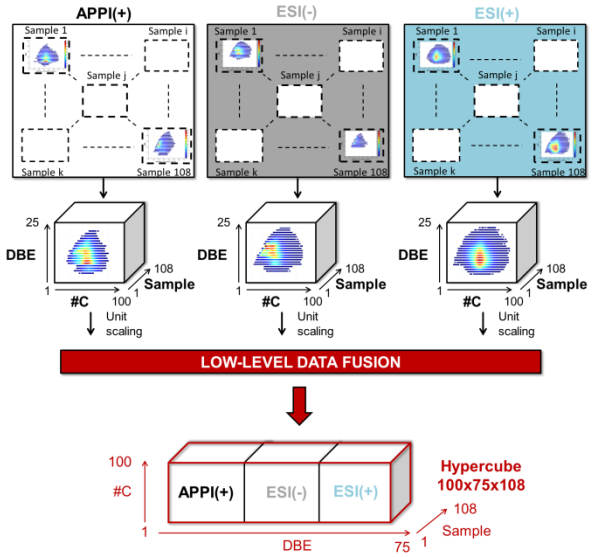

Figure 1: Low-level data fusion methodology

$$x_{i,j,k} = \sum_{p=1}^{P} a_{i,p}\, b_{j,p} c_{k,p} + e_{i,j,k} \quad (1)$$

with $x_{i,j,k}$ the ijk^{th} element of **X** and p the number of selected components. Traditionally, this model can also be written using tensor notations such as (Eq. 2):

$$\boldsymbol{X} = \sum_{p=1}^{P} \boldsymbol{a}_p \otimes \boldsymbol{b}_p \otimes \boldsymbol{c}_p \quad (2)$$

with $\boldsymbol{a}_p$, $\boldsymbol{b}_p$ and $\boldsymbol{c}_p$ being the p^{th} columns of the loading matrices **A**, **B,** and **C** respectively. Figure 2 provides a graphical

representation of the decomposition considering 2 components (i.e. p=2).

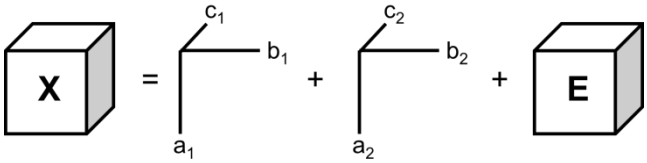

Figure 2: Illustration of a PARAFAC model considering 2 components.

In our case, the three modes of the hypercube correspond to the number of carbon atoms, DBE values, and samples. Moreover, a non-negativity constraint has been applied to each matrix as all their elements should be strictly positive. To evaluate the optimal number of components to be used in the PARAFAC decomposition, the well-known core consistency criterion has been calculated based on CONCORDIA parameter[30]. This parameter allows us to check if data are correctly modeled by the trilinear PARAFAC decomposition for a given number of components, 100% being the maximum core consistency value. All chemometric calculations in this work have been done on a desktop computer under the Matlab environment.

RESULTS AND DISCUSSION

Classical data exploration

In this specific section, one sample of the 4 different VGO types has been selected for further comparisons: sample VGO 4 (HCGO type), sample VGO 8 (SRVGO type), sample VGO 11 (HDT type), and sample VGO 13 (EBVGO type). The mass spectra acquired from these samples in ESI(+), ESI(-), and APPI(+) modes are available in S1 in Supporting Information. The compositions of nitrogen (N1[H]) and sulfur (S1) classes have been investigated by the study of DBE versus carbon number plots presented in Figure 3. It is quite usual to generate this kind of representation in mass spectrometry in order to summarize chemical information[26]. From a global point of view, some differences are spotted for a given ionization mode between the different samples, and more especially between ESI(-) and APPI(+) modes.

ESI(+) data. Even though the plots appear similar at first glance, some differences can be highlighted. As an example, very aromatic compounds (DBE>22) are identified in the HCGO sample which is directly related to its very aromatic character. Indeed, total aromatics contents determined by UV spectroscopy [ASTM D2269] are equal to 64.22 % w/w and 46.99 % w/w for the HCGO and SRVGO samples respectively. For the SRVGO sample, the intensity is mainly focused over specific DBE values such as 9, 10, 11 corresponding to quinolines and acridines families. For the other samples, the intensity is spread over a larger range of DBE values with the highest contributions at lower DBE (tetrahydroquinolines anilines pyridines family). Considering the HDT sample, all families are intense meaning that hydrotreatment efficiency seems to be evenly distributed over the different basic nitrogen families with no particularly refractory intense species. The distributions of the number of carbon atoms in considered samples are intense over C33 whereas it is more spread over the C25-C35 range for the HCGO sample. Very alkylated compounds (C>55) are not identified in the HDT sample but some low-alkylated compounds (C<20) are found in comparison to the SRVGO sample (feed of the HDT process). This could mean that the hydrotreatment of basic nitrogen compounds is more efficient towards very alkylated species that react faster than poorly alkylated species. As there are no significant changes in aromaticity, it can be assumed that these molecules may be in an island-type configuration with some big aromatic cores and some surrounding alkyl chains[31]. Thus, the hydrotreatment would decrease the average alkylation of the molecules without impacting their aromatic content, even though it ought to be the case for an archipelago configuration. To validate this assumption, collision energy has been applied in the ion source for another similar SRVGO sample (VGO 1). It has therefore been confirmed that no significant changes in DBE were observed compared to the original distribution either at 35 eV or 50 eV (see S2 in Supporting Information).

ESI(-) data. Poorly alkylated and very aromatic compounds (C22, DBE 15 corresponding to dibenzocarbazoles [DBC]) are especially present in the SRVGO sample. Some medium alkylated and less aromatic compounds (C32, DBE 12 corresponding to benzocarbazoles [BC]) were also fairly high in this sample. In comparison, only poorly alkylated and very aromatic compounds (C22, DBE 15) have intense contributions to the HCGO sample. For the EBVGO sample, even more aromatic compounds (DBE 16 and 17 corresponding to more aromatic dibenzocarbazoles) have very intense contributions over a larger carbon distribution. When considering the HDT sample, less aromatic compounds are present (DBE 12 to 14) but the whole distribution is much more alkylated. Thus, it has been assumed that these less aromatic compounds have been produced along hydrotreatment process as a consequence of hydrogenation mechanism of aromatic compounds such as dibenzocarbazoles. Indeed, the hydrogenation of dibenzocarbazoles (DBE 15) could lead to benzocarbazoles species with DBE equal to 12, 13 or 14 depending on the hydrogenation level. In this case, an island configuration could also be suspected. However, fragmentation should be performed to verify this hypothesis. It remains surprising that a lot of compounds with very high alkylation degrees (C>55) are identified in the HDT sample. N1O1 and N2 compounds hydrogenation could lead to N1 production[32]. As N2 compounds and N1O1 are completely or partly removed in the HDT sample (see S3 in Supporting Information), part of these species might actually have been converted into very-alkylated N1 species during hydrotreatment. An increase of N1 class is actually observed in the HDT sample (see S3 in Supporting Information).

APPI(+) data. Lastly, some differences have also been spotted within the different vacuum gas oils focusing on sulfur compounds. Again, the SRVGO sample contains a relatively high amount of aromatic and alkylated compounds (C<60) with a very intense contribution for a DBE value equal to 6 corresponding to benzothiophenes (BT) compounds. As regards the HCGO sample, compounds are less alkylated (C<50) and a significant amount of identified compounds are very aromatic (DBE>20) with a high contribution over a DBE value of 12 corresponding to naphtobenzothiophenes (NBT) with a very low alkylation level (C20). As observed for the other ionization modes, the contributions of the EBVGO

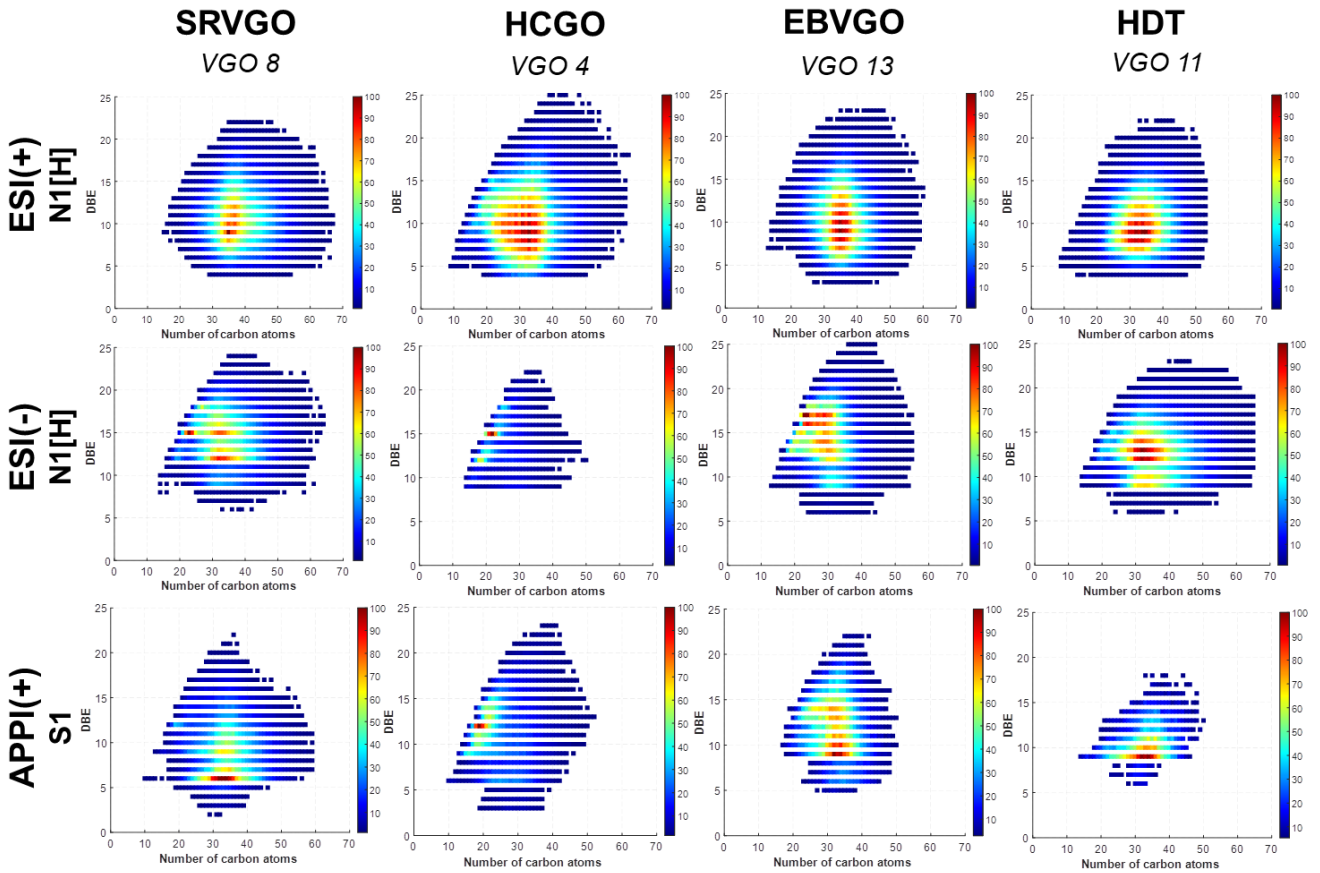

Figure 3: DBE versus number of carbon plots for the 4 selected gas oil samples (VGO 4, 8, 11 and 13) in ESI(+) mode focusing on N1[H] class, ESI(-) mode focusing on N1[H] class, and APPI(+) mode focusing on S1 class.

sample are more distributed over several families including DBE 9 to 11 (Dibenzothiophenes [DBT]) and 12 to 14 (Naphtobenzothiophenes). The alkylation levels of these molecules are quite similar to those of the HCGO sample (C<50). When considering the HDT sample, the maximum observed aromaticity degree is 18 when the other samples reach a value of 23 with an alkylation

range similar to the EBVGO and HCGO samples. The distribution is centered over a DBE equal to 9 (corresponding to dibenzothiophenes family) with a medium alkylation level (C33). It has been shown that dibenzothiophenes compounds are very refractory to hydrotreatment whereas benzothiophenes compounds are easily converted in gas oils samples[33]. These reactivity differences are also clearly observed here as benzothiophenes compounds are mainly present in the SRVGO sample while almost completely removed in the HDT sample. On the other side, dibenzothiophenes compounds remain very abundant in the HDT sample. In this first part, we can see that a comparison of different vacuum gas oil samples is possible considering a visual inspection of several DBE versus number of carbon atoms plots in parallel. However a systematic and exhaustive exploration remains difficult, which needs a new data analysis strategy developed in the next section.

Low-level data fusion and PARAFAC exploration

In this section, data have been formatted into a three-way array with three different modes (DBE, number of carbon atoms and samples) as explained earlier. The main benefit of PARAFAC in this study is the simultaneous visualization of the impact of each ionization mode on the decomposition of data. Besides, the projection of all samples over both components allows getting direct information rather than analyzing individually each DBE=f(#C) as previously done. Thus, PARAFAC method has been applied to this dataset in order to extract more information about the different samples. Two components have been selected with a core consistency equal to 83 as a strong decrease of this criterion has been observed when considering more components. In these conditions, 75% of the total variance has been explained by these two components PARAFAC decomposition. The projection of the samples over the two components and the resulting loadings for each mode are shown in Figures 4 and 5 respectively.

For easier reading, Figure 4 gives the projection of the different samples over both components. This projection allows validating the repeatability of the different analysis as some replicates clusters are mostly observed for each individual sample. Different groupings are observed according to the type of vacuum gas oil (SRVGO, HDT, EBVGO....) as well as their intrinsic characteristics (cuts, hydrotreatment level...). Indeed, the EBVGO samples are separated with respect to their initial cuts ranges (see Table 1) i.e. 350-540°C for samples VGO 12, 14 and 16 or 480-540°C for samples VGO 13 and 15. As regards the hydrotreated samples, the separation between them is due to the hydrodesulfurization level of sulfur. The sulfur content of the VGO sample 18 (i.e. 358 ppm) is lower than those of the VGO samples 17 and 11

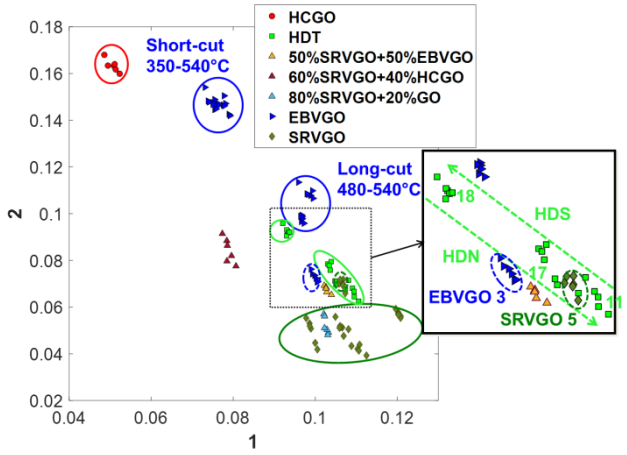

Figure 4: Projection of the samples over the 1st and 2nd components using PARAFAC approach.

(i.e. 708 and 979 ppm respectively). As a comparison, the nitrogen contents of these samples are respectively 675, 660 and 350 ppm. It is well known that nitrogen compounds are competing with sulfur compounds for the hydrotreatment of the latter[9]. For a low-pressure HDT process, high working temperature (i.e conditions used for HDT 18) increases hydrodesulfurization efficiency but does not affect much hydrodenitrogenation efficiency as these conditions are close to the thermodynamic limits of the hydrotreatment of nitrogen compounds (HDT 17 and 18 have similar nitrogen contents, see Table 1). The operating conditions (middle temperature and high pressure) used for the sample HDT 11 favor hydrodenitrogenation efficiency which could explain the poor efficiency of hydrodesulfurization. The 3 different mixed blends are projected with respect to their original content. As an example, the 60% SRVGO + 40% HCGO mix (VGO 9) is actually projected between the SRVGO and HCGO clusters and a little bit closer to the SRVGO cluster. Additionally, it is of interest to discuss the projection of the samples VGO 5 (SRVGO type) and VGO 3 (EBVGO type). Indeed, these two samples are projected quite far away from their respective clusters (SRVGO or EBVGO) to which they should belong. Looking at Table 1, we can see first that the macroscopic properties of the sample VGO 3 are very close to those of the other EBVGO samples. The same statement can be made for the VGO 5 sample and the other SRVGO samples. However, these two specific samples exhibit some unexpected behaviors when processed within several units (HDN, HDS, hydrocracking...). Especially, their reactivity to these processes is low compared to feeds with similar macroscopic properties. As a consequence, their unique projections using this chemometric approach is particularly interesting in order to find some explanations about their reactivity. Mix sample VGO 9 is partly composed of sample VGO 3 (40%) and it can be noted that the mixed blend also retains its unique character.

In order to identify variables responsible for such projection, one should consider the obtained loadings in each mode. From the loadings of DBE mode shown in Figure 5A, the chosen data formatting gives access to the identification of important families within each mode. From a general point of view, it can be noted that both electrospray modes and especially the negative one mainly describe each component.

The first component is mainly driven by ESI(+) data. More precisely, the contributions from ESI(+) mode are equally

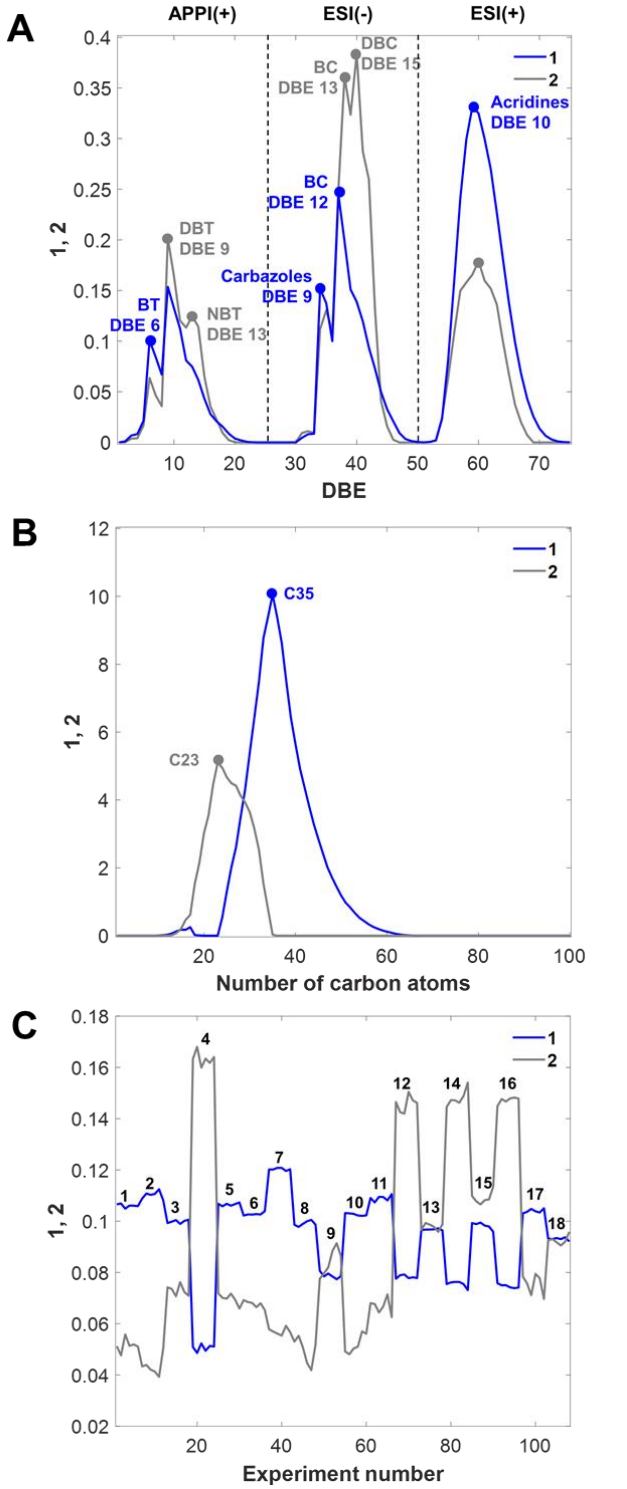

Figure 5: PARAFAC loadings. (A) Loadings from DBE mode. (B) Loadings from Number of carbon atoms mode. (C) Loadings from Sample mode (Experiment number 1 to 6 correspond to the 6 replicates of sample VGO 1, 7 to 12 to the 6 replicates of sample VGO 2 and so on).

distributed over the whole DBE range with a maximum for the acridines family (DBE 10, 11 and 12). This is consistent with previous observations in the classical exploration part as the DBE=f(#C) plots are quite similar. As regards APPI(+) and ESI(-) data, some specific families have significant contributions. Considering APPI(+) mode, the first component

mainly highlights contributions of benzothiophenes (DBE 6) and dibenzothiophenes compounds (DBE 9). For ESI(-) mode, the first component has contributions of carbazoles (DBE 9) and benzocarbazoles compounds (DBE 12).

The second component is driven by ESI(-) data. The contributions from ESI(+) mode are also equally distributed over the whole DBE range. As regards sulfur compounds, there is also a contribution of the BT family but to a lesser extent than the first component. We also observe a higher contribution of the DBT family compared with component 1. A small contribution of the NBT family is also observed in the second component. The second component has strong contributions for ESI(-) data of more aromatic compounds such as benzocarbazoles (DBE 13) and dibenzocarbazoles (DBE 15).

Thus, the second component would be more likely to describe samples containing very aromatic compounds (DBT, NBT, BC, DBC...) than the first component for which relatively low aromatic compounds would be expected (BT, Carbazoles, BC...). This is actually the case as the HCGO sample has a high score along with the second component with very aromatic compounds (see *classical data exploration*) in all ionization modes (see Figure 5C). On the opposite, most SRVGO samples show higher contributions along the first component (see Figure 5C) due to their high content in less aromatics compounds and especially BT compounds. One unique SRVGO sample (VGO 5) has a slightly higher contribution along the second component than the other SRVGO samples. Thus, its unique character could come from a higher content in very aromatic species such as BC with DBE equal to 12 or 13. The relative intensities of these species in VGO 5 are respectively equal to 18.5% and 15% while equal to 14% and 12% in the SRVGO 8 for example (see Figure S4 in Supporting Information). Regarding acridines contribution, it has also been shown earlier that the SRVGO intensity distribution was fully focused on acridines family while the distributions of HCGO and EBVGO samples were more spread over different families. The contribution along the first component is then logically very important for this family as it describes largely the SRVGO samples compositions. The sample VGO 3 (EBVGO type) should be closer to the 350-540°C EBVGO cluster according to its boiling points which induce a difference of composition over both components. Indeed, the VGO 3 sample shows very high contents in less aromatic neutral nitrogen compounds such as carbazoles with DBE equal to 9 (9.5%) and 10 (10.5%) or dibenzocarbazoles with DBE equal to 12 (14%) while in comparison, these compounds represent 6.5%, 8%, and 9.5% respectively of the relative intensity in the VGO 12 sample (see Figure S5 in Supporting Information). As carbazoles and benzocarbazoles are refractory to hydrotreatment[4], an increasing amount of these species compared to other samples could logically reduce the efficiency of HDT of the samples VGO 3 and VGO 5 (SRVGO type). This could then give us some clues on their different reactivity behavior and explain why there are closer to HDT cluster rather than EBVGO and SRVGO clusters. Regarding the different EBVGO cuts, it is expected to find a higher amount of less aromatics compounds (BT, DBT, Carbazoles, BC...) in the short cuts (350-540°C) and consequently a smaller amount of these compounds in the heavier cuts (480-540°C). For the HDT samples, the slight difference between the deep hydrodesulfurized sample (VGO 18) and the moderate hydrodesulfurized sample (VGO 11) could be explained by a larger projection along the second component and thus a smaller projection along the first component. This is consistent with previous observations regarding the very refractory character of DBT compounds and the easy conversion of BT compounds. On the opposite, a larger projection of deep hydrodenitrogenated sample (VGO 11) along the first component would indicate higher contents in less aromatic compounds such as benzocarbazoles or carbazoles which, again, are known to be refractory. Moreover, the intensity of acridines family seems to be consistent as these compounds would also remain in the HDT sample.

When focusing on alkylation degree loadings in Figure 5B, the average carbon number distribution of the first component is centered on C35 while the distribution of the second component is centered on C23. It means that samples more described by the first component would be more alkylated than those best described by the second component. This is actually the case as the HCGO sample has been shown to present some relatively low alkylation degrees, as well as short-cut EBVGO samples which logically have a smaller alkylation degree than heavier long-cuts. The SRVGO samples have relatively high levels of alkylation as shown for the VGO 8 sample in previous part. The unique character of the VGO 3 sample (EBVGO type) could also be explained by a higher shift of the degree of alkylation compared to other EBVGO samples. Indeed, its overall number of carbon distribution is centered on the C31-C32 range while it is centered on the C28-C30 range in ESI(-) mode for the sample VGO 12 (EBVGO type with similar boiling range). This is observed in all ionization modes with a very slight difference for APPI(+) data (see Figure S6 in Supporting Information). Thus, its smaller projection along the second component is consistent with an increase in alkylation level. A very small shift is also observed for the sample VGO 5 (SR type) compared to the sample VGO 8 (1 carbon atom more for the sample VGO 8). For the HDT samples, hydrodesulfurization would produce less alkylated compounds while hydrodenitrogenation would produce very alkylated compounds (see Figure 3).

The loadings from the sample mode are presented in Figure 5C. They describe how each experiment (and thus each replicate for each sample) is individually decomposed by each component which is actually a 1D projection of Figure 4. As a consequence, it will not be further discussed.

In summary, SRVGO samples are found to be very alkylated but not so aromatic while the HCGO sample is revealed to be very aromatic but poorly alkylated. The aromaticity and alkylation levels of the EBVGO samples are directly related to their boiling cuts which are in agreement with Boduszynski principle[34]. The HDT samples show two opposed characters whether deep hydrodenitrogenation or deep hydrodesulfurization were targeted to produce them. Among hydrotreated samples compositions, benzocarbazoles and dibenzothiophenes were found to be the main refractory compounds.

CONCLUSION

In this study, a combined approach using ultra-high resolution mass spectrometry (FT-ICR MS), low-level data fusion and PARAFAC method has been applied for the characterization of nitrogen and sulfur compounds in vacuum gas oils samples. 18 vacuum gas oil samples have been

analyzed in three ionization modes. Firstly, classical data exploration has shown differences in composition for the different vacuum gas oils within nitrogen and sulfur families thanks to DBE=f(#C) plots. One feed and its effluent have also been compared to identify refractory species. In a second step, data sets from the three ionization modes have been concatenated using a low-level data fusion strategy. Then, the PARAFAC method has been applied to the previously generated hypercube for simultaneous exploration of the different experimental conditions. To the best of our knowledge, it was the first time a multiway analysis has been considered for the exploration of FT-ICR MS data sets. Three different modes have then been considered - DBE, number of carbon atoms, and samples. Moreover, also for the first time, high-resolution MS data from three ionization modes have been simultaneously considered for an extended chemometric study of nitrogen and sulfur compounds. The decomposition of the data cube considering two significant components has led to the identification of the variables explaining the characteristics of the samples including details about their aromaticity and alkylation degrees. The PARAFAC method was able to extract significant information while the classical data exploration method was less exhaustive and more fastidious. To conclude, this multi-way approach is a very helpful and promising tool to explore FT-ICR MS datasets when considering numerous samples and different experimental conditions such as the use of different ionization sources. To go further, future works could consider the simultaneous analysis of previous classes with other heteroatomic one's such as N2 or N1O1 in an extended n-way exploration framework, hence considering an even more complex hypercube in order to assess their reactivity to hydrotreatment.

ASSOCIATED CONTENT

Supporting Information

FT-ICR MS mass spectra from samples VGO 4, 8, 11, and 13 in all ionization modes. Comparison of DBE distribution before and after application of collision energy. Relative heteroatomic abundances distributions of SRVGO and HDT samples in ESI(-) mode. Distributions of the relative intensities according to DBE values for SRVGO 5 and 8 and EBVGO 3 and 12 in ESI(-) mode and distribution of the relative intensities according to the number of carbon atoms in all ionization modes for EBVGO 3 and 12 (PDF)

AUTHOR INFORMATION

Corresponding Author

*alexandra.berlioz-barbier@ifpen.fr and ludovic.duponchel@univ-lille.fr

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank warmly Fabien Chainet, Jérémie Barbier and Laurent Simon from IFP Energies Nouvelles for providing the samples and Raffaele Vitale from the LASIR lab at the University of Lille for fruitful discussions about chemometric strategies.

REFERENCES

(1) Mullins, O. C.; Sheu, E.; Hammami, A.; Marshall, A. G. *Asphaltenes, heavy oils, and petroleomics*; Springer: New-York, 2007.

(2) Boduszynski, M. M. Composition of heavy petroleums. 2. Molecular characterization. *Energy Fuels* **1988**, *2*, 597–613.

(3) Sau, M.; Basak, K.; Manna, U.; Santra, M.; Verma, R. P. Effects of organic nitrogen compounds on hydrotreating and hydrocracking reactions. *Catalysis Today* **2005**, *109*, 112–119.

(4) Shin, S.; Sakanishi, K.; Mochida, I.; Grudoski, D. A.; Shinn, J. H. Identification and Reactivity of Nitrogen Molecular Species in Gas Oils. *Energy Fuels* **2000**, *14*, 539–544.

(5) Nguyen, M.T.; Pirngruber, G.; Chainet, F.; Albrieux, F.; Tayakout-Fayolle, M.; Geantet, C. Molecular level insights into straight run/coker gas oil hydrodenitrogenation by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Energy Fuels* **2019,** DOI: 10.1021/acs.energyfuels.8b04432.

(6) Wiwel, P.; Knudsen, K.; Zeuthen, P.; Whitehurst, D. Assessing Compositional Changes of Nitrogen Compounds during Hydrotreating of Typical Diesel Range Gas Oils Using a Novel Preconcentration Technique Coupled with Gas Chromatography and Atomic Emission Detection. *Ind. Eng. Chem. Res.* **2000**, *39*, 533–540.

(7) Nguyen, M. T. Support acidity effects of NiMo sulfide catalysts in hydrodenitrogenation of quinoline, indole and Coker Gas Oil: Thesis manuscript, 2016.

(8) Ma, X.; Sakanishi, K.; Mochida, I. Hydrodesulfurization Reactivities of Various Sulfur Compounds in Vacuum Gas Oil. *Ind. Eng. Chem. Res.* **1996**, *35*, 2487–2494.

(9) Rabarihoela-Rakotovao, V.; Diehl, F.; Brunet, S. Deep HDS of Diesel Fuel: Inhibiting Effect of Nitrogen Compounds on the Transformation of the Refractory 4,6-Dimethyldibenzothiophene Over a NiMoP/Al2O3 Catalyst. *Catal Lett* **2009**, *129*, 50–60.

(10) Liu, M.; Wang, M.; Zhang, L.; Xu, Z.; Chen, Y.; Guo, X.; Zhao, S. Transformation of Sulfur Compounds in the Hydrotreatment of Supercritical Fluid Extraction Subfractions of Saudi Arabia Atmospheric Residua. *Energy Fuels* **2015**, *29*, 702–710.

(11) McKenna, A. M.; Purcell, J. M.; Rodgers, R. P.; Marshall, A. G. Heavy Petroleum Composition. 1. Exhaustive Compositional Analysis of Athabasca Bitumen HVGO Distillates by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry: A Definitive Test of the Boduszynski Model. *Energy Fuels* **2010**, *24*, 2929–2938.

(12) McKenna, A. M.; Blakney, G. T.; Xian, F.; Glaser, P. B.; Rodgers, R. P.; Marshall, A. G. Heavy Petroleum Composition. 2. Progression of the Boduszynski Model to the Limit of Distillation by Ultrahigh-Resolution FT-ICR Mass Spectrometry. *Energy Fuels* **2010**, *24*, 2939–2946.

(13) Rodgers, R. P.; Schaub, T. M.; Marshall, A. G. Petroleomics: MS returns to its roots. *Analytical chemistry* **2005**, *27*.

(14) Palacio Lozano, D. C.; Gavard, R.; Arenas-Diaz, J. P.; Thomas, M. J.; Stranz, D. D.; Mejía-Ospino, E.; Guzman, A.; Spencer, S. E. F.; Rossell, D.; Barrow, M. P. Pushing the analytical limits: New insights into complex mixtures using mass spectra segments of constant ultrahigh resolving power. *Chemical science* **2019**, *10*, 6966–6978.

(15) Guillemant, J.; Albrieux, F.; Oliveira, L. P. de; Lacoue-Nègre, M.; Duponchel, L.; Joly, J.-F. Insights from Nitrogen Compounds in Gas Oils Highlighted by High-Resolution Fourier Transform Mass Spectrometry. *Analytical chemistry* **2019**, *91*, 12644–12652.

(16) Purcell, J. M.; Juyal, P.; Kim, D.-G.; Rodgers, R. P.; Hendrickson, C. L.; Marshall, A. G. Sulfur Speciation in Petroleum: Atmospheric Pressure Photoionization or Chemical Derivatization and Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Energy Fuels* **2007**, *21*, 2869–2874.

(17) Marshall, A. G.; Rodgers, R. P. Petroleomics: The next grand challenge for chemical analysis. *Accounts of chemical research* **2004**, *37*, 53–59.

(18) Hur, M.; Yeo, I.; Park, E.; Kim, Y. H.; Yoo, J.; Kim, E.; No, M.-h.; Koh, J.; Kim, S. Combination of statistical methods and Fourier transform ion cyclotron resonance mass spectrometry for more comprehensive, molecular-level interpretations of petroleum samples. *Analytical chemistry* **2010**, *82*, 211–218.

(19) Hur, M.; Yeo, I.; Kim, E.; No, M.-h.; Koh, J.; Cho, Y. J.; Lee, J. W.; Kim, S. Correlation of FT-ICR Mass Spectra with the Chemical and Physical Properties of Associated Crude Oils. *Energy Fuels* **2010**, *24*, 5524–5532.

(20) Hur, M.; Ware, R. L.; Park, J.; McKenna, A. M.; Rodgers, R. P.; Nikolau, B. J.; Wurtele, E. S.; Marshall, A. G. Statistically Significant Differences in Composition of Petroleum Crude Oils Revealed by Volcano Plots Generated from Ultrahigh Resolution Fourier Transform Ion Cyclotron Resonance Mass Spectra. *Energy Fuels* **2018**, *32*, 1206–1212.

(21) Guillemant, J.; Albrieux, F.; Lacoue-Nègre, M.; Pereira de Oliveira, L.; Joly, J.-F.; Duponchel, L. Chemometric Exploration of APPI(+)-FT-ICR MS Data Sets for a Comprehensive Study of Aromatic Sulfur Compounds in Gas Oils. *Analytical chemistry* **2019**, *91*, 11785–11793.
(22) Ohno, T.; Amirbahman, A.; Bro, R. Parallel factor analysis of excitation-emission matrix fluorescence spectra of water soluble soil organic matter as basis for the determination of conditional metal binding parameters. *Environmental science & technology* **2008**, *42*, 186–192.
(23) Mirnaghi, F. S.; Soucy, N.; Hollebone, B. P.; Brown, C. E. Rapid fingerprinting of spilled petroleum products using fluorescence spectroscopy coupled with parallel factor and principal component analysis. *Chemosphere* **2018**, *208*, 185–195.
(24) Teglia, C. M.; Azcarate, S. M.; Alcaráz, M. R.; Goicoechea, H. C.; Culzoni, M. J. Exploiting the synergistic effect of concurrent data signals: Low-level fusion of liquid chromatographic with dual detection data. *Talanta* **2018**, *186*, 481–488.
(25) Rubio, L.; Valverde-Som, L.; Sarabia, L. A.; Ortiz, M. C. The behaviour of Tenax as food simulant in the migration of polymer additives from food contact materials by means of gas chromatography/mass spectrometry and PARAFAC. *Journal of chromatography. A* **2019**, *1589*, 18–29.
(26) Marshall, A. G.; Rodgers, R. P. Petroleomics: Chemistry of the underworld. *PNAS*, *105*, 18090–18095.
(27) Kozhinov, A. N.; Zhurov, K. O.; Tsybin, Y. O. Iterative method for mass spectra recalibration via empirical estimation of the mass calibration function for Fourier transform mass spectrometry-based petroleomics. *Analytical chemistry* **2013**, *85*, 6437–6445.
(28) *Data Fusion Methodology and Applications;* Cocchi, M., Ed.; Elsevier: Cambridge, 2019.
(29) Bro, R. PARAFAC. Tutorial and applications. *Chemometrics and Intelligent Laboratory Systems* **1997**, *38*, 149–171.
(30) *Fundamentals and Analytical Applications of Multiway Calibration;* Muñoz De la Peña, Arsenio; Goicoechea, H. C.; Escandar, G.; Olivieri, A., Eds.; Elsevier: Oxford, 2015.
(31) Chacón-Patiño, M. L.; Rowland, S. M.; Rodgers, R. P. Advances in Asphaltene Petroleomics. Part 1: Asphaltenes Are Composed of Abundant Island and Archipelago Structural Motifs. *Energy Fuels* **2017**, *31*, 13509–13518.
(32) Chen, X.; Shen, B.; Sun, J.; Wang, C.; Shan, H.; Yang, C.; Li, C. Characterization and Comparison of Nitrogen Compounds in Hydrotreated and Untreated Shale Oil by Electrospray Ionization (ESI) Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (FT-ICR MS). *Energy Fuels* **2012**, *26*, 1707–1714.
(33) Valencia, D.; García-Cruz, I.; Uc, V. H.; Ramírez-Verduzco, L. F.; Aburto, J. Refractory Character of 4,6-Dialkyldibenzothiophenes: Structural and Electronic Instabilities Reign Deep Hydrodesulfurization. *ChemistrySelect* **2018**, *3*, 8849–8856.
(34) Boduszynski, M. M. Composition of heavy petroleums. 1. Molecular weight, hydrogen deficiency, and heteroatom concentration as a function of atmospheric equivalent boiling point up to 1400.degree.F (760.degree.C). *Energy Fuels* **1987**, *1*, 2–11.