energy&fuels

pubs.acs.org/EF

Article

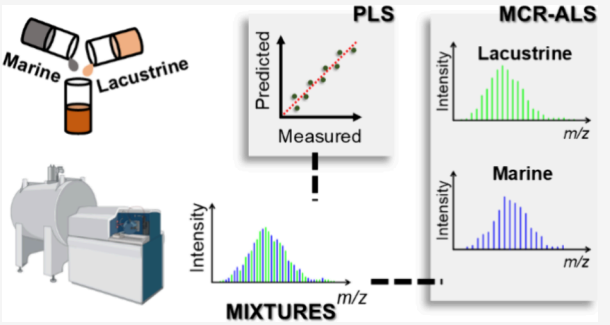
Crude Oil Mixture Resolution by FT-ICR MS and Multivariate Modeling: Quantitative Source Attribution Considering Different Thermal Maturities

João Victor A. Oliveira, Jussara V. Roque,* Danielle M. M. Franco, Leonardo F. Matos, Mario D. Rangel, Joelma P. Lopes, Ygor dos S. Rocha, and Boniek Gontijo*

Cite This: *Energy Fuels* 2025, 39, 15648−15658

Read Online

ABSTRACT: Quantitative determination of the relative contributions from multiple sources in mixed source crude oils has consistently been posed as a significant challenge in petroleum geochemistry. In this context, this work explores the evaluation of crude oil mixtures using Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS) with positive atmospheric pressure photoionization (APPI (+)). Two mixture designs (Plans 1 and 2) were evaluated, involving lacustrine and marine endmembers with moderate and high thermal maturities. Classical petroleomics characterization was combined with multivariate curve resolution-alternating least-squares (MCR-ALS) and partial least-squares (PLS) regression to investigate compositional behavior across these systems. In Plan 1, which involved mixtures of oils with distinct thermal maturity, substantial differences were observed in molecular class distributions. Plan 2, composed of highly mature lacustrine and marine oils, showed more subtle, yet detectable, differences, confirming the role of thermal evolution in shaping class profiles. MCR-ALS successfully recovered pure spectral profiles for both crude oil origins in each design, with Pearson correlation coefficients exceeding 0.97. PLS regression proved to be a robust tool for estimating the lacustrine content in the mixtures. For Plan 1, the model using only hydrocarbon (HC) class variables achieved the best performance, while in Plan 2, the model based on nitrogen-, oxygen-, and sulfur-containing (NOS) compounds provided the most accurate predictions. This study demonstrated the application of different chemometric tools and approaches to understand the behavior of crude oil mixtures with different origins and thermal maturity.



1. INTRODUCTION

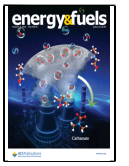
Organic geochemistry has been crucial in enhancing the understanding of the geological history behind the formation and accumulation of organic matter in source rocks.[1] A primary approach in this field is the analysis of biomarkers, compounds linked to precursor molecules from organisms or plants.[2] By utilizing biomarker ratios, researchers can gain insights into the type of organic matter deposited and infer its origin, biodegradation level, and degree of thermal maturity.[2] Biomarkers are commonly analyzed using gas chromatography−mass spectrometry (GC-MS),[3] a widely employed technique in petroleum studies.

In petroleum generation, the characteristics of crude oil and natural gas are primarily influenced by the type of kerogen, an organic material categorized by Tissot et al.[4] into four types (Types I−IV), based on their hydrogen−carbon (H/C) and oxygen−carbon (O/C) ratios.[5] However, in natural settings, petroleum formation results from the contribution of organic matter from multiple source rocks, leading to complex mixtures with diverse organic inputs.[6,7] This variability increases the complexity of the crude oil composition, making its analysis and interpretation more challenging.

Several studies have investigated crude oil mixtures derived from different sources[6] and their thermal maturity[8] and biodegradation.[9,10] These issues have been extensively studied by using hydrocarbon biomarkers analyzed via GC-MS analysis. However, specific processes, such as biodegradation and thermal maturity, can alter the behavior of well-established biomarkers. Consequently, the application of new analytical techniques, such as ultrahigh-resolution mass spectrometry (UHRMS), has emerged as a promising alternative.

Although GC-MS remains a gold standard for crude oil characterization, it has limitations, particularly regarding polar

Received: May 4, 2025
Revised: July 24, 2025
Accepted: July 30, 2025
Published: August 11, 2025



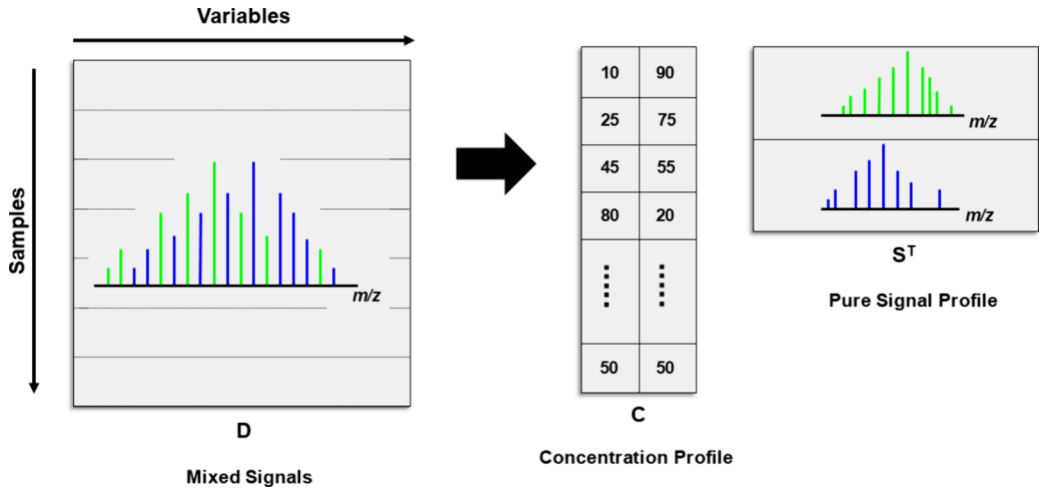


Figure 1. Schematic representation of MCR-ALS (multivariate curve resolution−alternating least squares) decomposition applied to binary mixture samples analyzed by mass spectrometry. The original data matrix (**D**), containing mixed signals across samples and variables (e.g., m/z), is decomposed into two matrices: the concentration profile matrix (**C**), representing the relative contribution of each pure component across samples, and the spectral matrix ($\mathbf{S}^T$), representing the pure signal profiles of the components.

(heteroatomic) compounds. To address these limitations, advances in UHRMS instrumentation have significantly improved the characterization of polar compounds in petroleum. Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS),[11] combined with ionization techniques such as electrospray ionization (ESI), atmospheric pressure chemical ionization (APCI), and atmospheric pressure photoionization (APPI), enables a more detailed analysis of these compounds in crude oils.[12]

Several studies have leveraged FT-ICR MS to investigate depositional environments,[13−16] migration patterns,[17] thermal maturity,[14,18,19] and biodegradation.[20−22] This research field is conventionally referred to as petroleomics.[23] Petroleomics involves the characterization of organic compounds in crude oils to enhance the understanding of their behavior and chemical distribution. This approach employs various tools, such as diagrams of double-bond equivalent (DBE), class distribution, and carbon number (#C).[23−25] Moreover, integrating petroleomics with statistical and chemical methods, collectively known as chemometrics, provides key knowledge about crude oil composition.

Several studies have used different chemometric tools to evaluate the geochemical characteristics. Melendez-Perez et al.[26] described the use of discriminant analysis to differentiate lacustrine and marine organic matter depositional paleoenvironments by ESI (−) FT-ICR MS. Similarly, Rocha et al.[13] conducted an extensive characterization of lacustrine and marine crude oils using ESI (−) FT-ICR MS, combined with unsupervised methods such as principal component analysis (PCA) and hierarchical clustering (HCA), to enhance the understanding of crude oil composition and how the molecular classes differentiate paleoenvironmental conditions.

Despite these advances, few studies have systematically explored crude oil mixtures of distinct origins and thermal histories using advanced chemometric methods. Thus, applying statistical models to analyze crude oil mixture compositions remains an emerging area of research. In this study, we explored the effect of mixing crude oils of distinct origins, such as lacustrine and marine, and thermal maturities using APPI (+) FT-ICR MS. Chemometric methods, including multivariate curve resolution-alternating least-squares (MCR-ALS) and partial least-squares regression (PLS), were employed to model and quantify the compositional differences resulting from oil mixing introducing a novel PLS-based approach for estimating the lacustrine oil fraction within mixed crude oil samples.

2. EXPERIMENTAL SECTION

2.1. Endmember Crude Oil Samples. The Center of Research, Development, and Innovation Leopoldo Américo Miguez de Mello (CENPES, Petrobras, Rio de Janeiro, Brazil) supplied twenty-six non-biodegraded crude oil samples from different basins. Supporting Information Table S1 summarizes the main characteristics of each crude oil sample, while Table S2 presents some biomarker data used for geochemical setting assessment.

2.2. Endmember Blends. For this study, three pools of crude oils of the exact origin were prepared to add variability to the data set. For each blend, equal weight percentages of samples were used to ensure that all samples had the same influence on the blend.

The marine crude oil origin blend (M) was prepared by mixing ten marine crude oil origins with different thermal maturities. Seven crude oils presented high thermal maturity, while the remaining samples presented moderate thermal maturity, as shown in Table S1.

Two lacustrine crude oil origin blends were prepared with different thermal maturity. The first lacustrine blend, designed as L1, was prepared by mixing ten lacustrine crude oils of moderate thermal maturity. The second lacustrine blend was prepared by mixing six lacustrine crude oils of origin with high thermal maturity (L2).

2.3. Design of Mixtures. Two binary mixture designs were generated.[27] Each binary mixture design consists of a calibration and test set, comprising twenty-six design points and nine design points, respectively. Moreover, each binary mixture design combines lacustrine and marine blends. The first binary mixture design (Plan 1) was performed by mixing the M blend and L1 blend, while the second was performed by mixing the M blend and L2 blend. Both mixture designs were prepared by varying the mass content of lacustrine and marine blends from 0 to 100%, as displayed in Table S3. Furthermore, all design points were prepared at a final concentration of 10.0 mg·mL^{-1}, solubilizing the crude oil mixtures in HPLC-grade toluene (Tedia Co., Fairfield, OH, USA). All samples were acquired in a random order.

2.4. FT-ICR MS Analysis. FT-ICR MS analyses were performed using a 7 T SolariX 2xR mass spectrometer (Bruker Daltonics, Bremen, Germany) equipped with an APPI ionization source operated in positive ion mode, utilizing a krypton UV lamp with an energy of 10.6 eV. Samples were prepared by diluting design points in

https://doi.org/10.1021/acs.energyfuels.5c02307
Energy Fuels 2025, 39, 15648−15658

HPLC-grade methanol (Tedia) to a final concentration of 0.50 mg·mL^{-1} in a toluene:methanol mixture (1:1, v/v) and injected at a flow rate of 500 μL·h^{-1}. Nitrogen was used as both the nebulizing gas (2 bar) and the drying gas and maintained at 200 °C. The capillary voltage was set at 4 kV, with ions accumulated in the collision cell for 0.05 s and transferred to the ICR cell within 1 ms. Spectra were acquired in broadband mode using 8 megaword data sets, with 300 scans summed per mass spectrum over a mass range of 150−1200 Da. All parameters can be found in Table S4.

2.5. Data Treatment. *2.5.1. Preprocessing.* The data previously obtained from APPI (+) FT-ICR MS were recalibrated using radical cation hydrocarbons as an internal reference list in DataAnalysis 5.0 SR1 software (ver. 5.0, Bruker Daltonics GmbH). Subsequently, molecular formulas were assigned employing Composer software (Version 1.5.3, Sierra Analytics, USA), with parameters described in Table S5. The composition table, generated from the molecular formula assignment in Composer, was imported into Matlab (The MathWorks, Natick, MA, USA) for multivariate data analyses.

2.5.2. MCR-ALS Analysis. MCR-ALS is a method that decomposes the data matrix into a bilinear model by optimizing a set of initial estimates during successive iterations until a convergence criterion is reached.[28,29] In this way, complex mixtures can be resolved regarding each pure component's contribution to the system.[30,31]

In this study, MCR-ALS analysis was carried out using the MCR-ALS GUI 2.0 toolbox.[32] Two MCR-ALS analyses were performed, one for each mixture design. A matrix $\mathbf{D}(m,n)$ was organized for each MCR-ALS analysis, where the m rows represent the design points analyzed and n columns represent the m/z intensities. Each $\mathbf{D}$ matrix was decomposed into a concentration profile, denoted as $\mathbf{C}(m,k)$ matrix, and a pure-spectral profile, denoted as the $\mathbf{S}^{\mathrm{T}}(k,n)$ matrix, where k is the number of pure components. The $\mathbf{C}$ matrix represents each mixture design's relative proportion of pure components. In contrast, the $\mathbf{S}^{\mathrm{T}}$ matrix represents the spectra obtained by MCR-ALS for each pure component, meaning the contribution of the lacustrine or marine mass spectrum profile, as represented in Figure 1.

The initial estimate of the spectra was performed using the Pure algorithm. The choice of the number of pure components was performed by using the SVD algorithm. Furthermore, the closure constraint was applied to the concentration profile, while the non-negativity constraint was employed in both pure signal and concentration profiles. The closure constraint expresses a mass-balance condition in the system and can be applied to the concentration profile. In turn, the non-negativity constraint expresses the non-existence of negative values in the profile studied.

Additionally, the quality models were evaluated employing the lack of fit (lof) and the variance explained (R^2), described for the equations in eq 1 and eq 2, respectively. Furthermore, Pearson's correlation coefficient (R), eq 3, was exploited for evaluation of the quality of results obtained in MCR-ALS.

$$\mathrm{lof}\,(\%) = 100 \times \sqrt{\frac{\sum_{m=1}^{i}\sum_{n=1}^{j} e_{m,n}{}^{2}}{\sum_{m=1}^{i}\sum_{n=1}^{j} d_{m,n}{}^{2}}} \tag{1}$$

$$R^2\,(\%) = 100 \times \frac{\sum_{m=1}^{i}\sum_{n=1}^{j} d_{m,n}{}^{2} - \sum_{m=1}^{i}\sum_{n=1}^{j} e_{m,n}{}^{2}}{\sum_{m=1}^{i}\sum_{n=1}^{j} d_{m,n}{}^{2}} \tag{2}$$

$$R = \frac{\sum_{i=1}^{N}(\hat{y}_i - \bar{\hat{y}})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(\hat{y}_i - \bar{\hat{y}})^2}\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}} \tag{3}$$

where $d_{m,n}$ is an element of the data matrix $\mathbf{D}$ and $e_{m,n}$ is the related residual error. N represents the number of samples, whereas y, $\hat{y}$, and $\bar{y}$ represent measured and predicted values, and the median value of y, respectively.

2.5.3. PLS Regression. Several strategies were employed to construct regression models based on the classes of variables, as outlined in Table 1. These strategies involved filtering variables according to their heteroatomic class (N, O, and S) or the absence of a heteroatom (HC). In this context, each strategy was organized in a matrix $\mathbf{X}(m,n)$, representing the independent variables, that the m rows represent the number of samples, while the n columns represent the number of variables. The dependent variables, designed by $\mathbf{y}(m,1)$, are associated with the content of the lacustrine blend in each binary mixture.

Table 1. Strategies Adopted for the Building of Partial Least-Squares Regression Models

	No. of variables		
Strategy	1° Plan	2° Plan	Description
$\mathbf{X}_{\mathrm{Full}}$	3777[a]	4005[a]	All variables
$\mathbf{X}_{\mathrm{NOS}}$	2093	1939	Variables from N, O, and S classes
$\mathbf{X}_{\mathrm{HC}}$	1684	2066	Variables from HC class
$\mathbf{X}_{\mathrm{N}}$	903	604	Variables from N class
$\mathbf{X}_{\mathrm{O}}$	531	973	Variables from O class
$\mathbf{X}_{\mathrm{S}}$	659	362	Variables from S class

[a]The number of all variables assessed by APPI (+) FT-ICR MS for HC, N, O, and S classes.

Twelve inverse regression models were built using PLS regression[33] on the open-source software Relyon.[34] This software enabled both the optimization of pretreatments used for model construction and the construction of PLS regression models.

To determine the optimal regression model pretreatment, combinations of two preprocessing methods, autoscale and mean center, and three transformations, norma type I, norma type II, and infinity norm, were evaluated. Random cross-validation with ten splits was employed to select the number of latent variables (LVs).

Quality models were evaluated by two main parameters: Root mean square error (RMSE) and R are designed as eq 4 and eq 3, respectively.

$$\mathrm{RMSE} = \sqrt{\sum_{i}^{N}\frac{(y_i - y_1)^2}{N}} \tag{4}$$

For the calibration build model, N represents the number of samples in the calibration set (C), and the correlation and error are called the correlation coefficient of calibration (R_{C}) and the root-mean-square error of calibration (RMSEC), respectively. For the internal cross-validation (CV), N represents the number of samples in the cross-validation set, and the correlation and error are called the correlation coefficient of cross-validation (R_{CV}) and the root-mean-square error of cross-validation (RMSECV), respectively. For external validation, N represents the number of samples in the prediction (P) set, and the correlation and error are called the correlation coefficient of prediction (R_{P}) and the root-mean-square error of prediction (RMSEP), respectively.

3. RESULTS AND DISCUSSION

3.1. Characterization of Endmember Blends. Figure 2 shows the APPI (+) FT-ICR MS spectra for the three endmember blends evaluated in this study: the marine-derived M blend (Figure 2A) and the two lacustrine-derived blends, L1 (Figure 2B) and L2 (Figure 2C). Molecular formulas were successfully assigned to each mass spectrum, achieving a notably high assignment rate of over 99.4% for all mass spectra analyzed.

A total of 19 molecular classes were identified and assessed in radical and protonated forms using APPI (+) ionization (Figure S1). However, only HC, N, O, and S classes exhibited relative abundances above 1% and thus became the primary focus of this investigation. Figure 3 summarizes the combined abundances of these significant classes across each blend. Although the HC class was the most abundant in all crude oil mixtures, noticeable differences emerged in their relative distributions. For the M blend, relative abundances decreased

https://doi.org/10.1021/acs.energyfuels.5c02307
Energy Fuels 2025, 39, 15648−15658

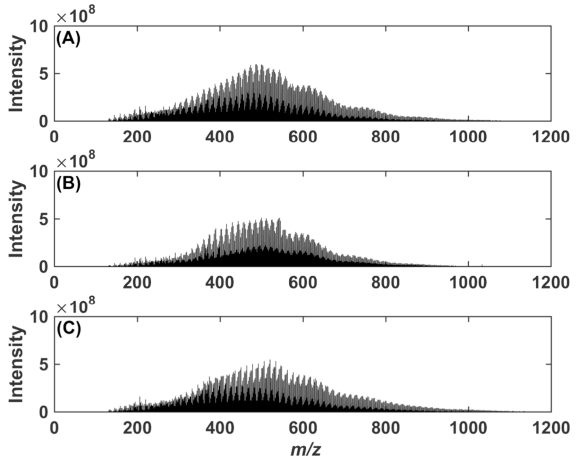


Figure 2. Mass spectra obtained by APPI (+) FT-ICR MS for the endmember blends: (A) marine crude oil (M), (B) lacustrine crude oil with moderate thermal maturity (L1), and (C) lacustrine crude oil with high thermal maturity (L2).

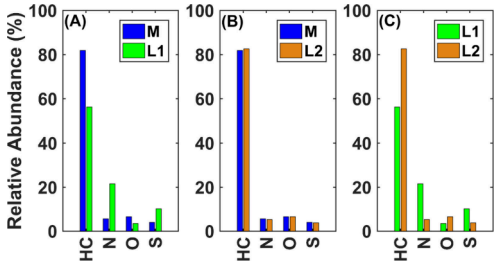


Figure 3. Class distribution of selected molecular classes based on the sum of radical and protonated molecular formulas obtained by APPI (+) FT-ICR MS. Comparisons are shown for (A) marine (M) vs lacustrine with moderate thermal maturity (L1); (B) marine (M) vs lacustrine with high thermal maturity (L2); (C) lacustrine with moderate (L1) vs high thermal maturity (L2).

sequentially as follows: HC (81.99%), O (6.52%), N (5.56%), and S (4.03%). In contrast, the L1 blend showed a distinct pattern with abundances of HC (61.79%), N (20.10%), S (8.88%), and O (3.70%). Meanwhile, the L2 blend presented yet another unique distribution, predominantly composed of HC (84.04%), followed by those of O (8.32%), N (4.01%), and S (1.85%).

Figure 3A illustrates notable differences in class distribution between the M and L1 blends, with only the HC class exceeding 20% relative abundance in both blends. Conversely, the M and L2 blends exhibit greater compositional similarity, as shown in Figure 3B, likely resulting from their comparable thermal maturity. As thermal maturity increases, the relative abundances of heteroatomic classes (such as nitrogen-, oxygen-, and sulfur-containing compounds) typically decrease alongside an increase in hydrocarbon content.[18,19,35] This compositional trend is clearly reflected in the L2 blend, which has undergone greater thermal maturation compared to that of the L1 blend, thus exhibiting a significantly higher hydrocarbon (HC) content.

This pattern is consistent with findings reported by several studies, notably Oldenburg et al.,[18] who employed APPI (+) FT-ICR MS to investigate how thermal maturity influences compound class distribution in crude oils. They observed that increasing thermal maturity results in decreased relative abundances of heteroatomic classes, accompanied by an increased dominance of hydrocarbons. Specifically, Figure 3C illustrates this effect: the L2 blend exhibits significantly higher HC abundance than the less mature L1 blend, effectively demonstrating how thermal evolution shapes the crude oil composition. However, the maturity is not the unique characteristic that can be attributed to variations in the compound distribution. M, L1, and L2 blends were prepared using oils from different basins. In this way, the variations may also result from the different source facies and migration, since all of these factors combined with secondary alterations have significant implications for the compound distribution present in crude oil.[35,36]

Figure 4 presents double bond equivalent (DBE) vs carbon number (#C) diagrams for the HC, N, O, and S classes, which provide insight into the aromaticity and degree of alkylation for each blend. Table S6 shows the ranges of DBE and #C. Notably, all classes exhibit distinct DBE distributions when comparing the L1 and L2 blends, indicating differences in aromatic character. However, when comparing L1 and L2 samples, some discrepancies can be observed in the degree of alkylation as well as the degree of aromatization for HC class.

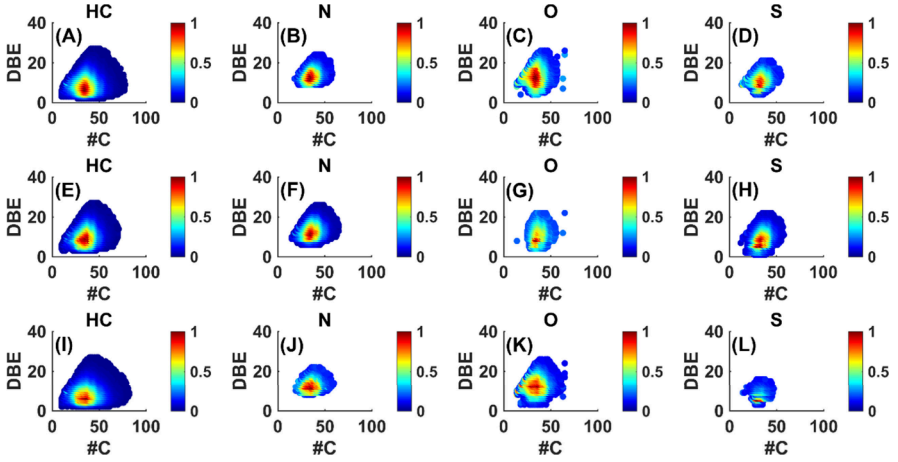


Figure 4. Double bond equivalent (DBE) vs carbon number (#C) diagrams for selected molecular classes obtained by APPI (+) FT-ICR MS. (A–D) Marine (M) endmember; (E–H) lacustrine (L1) endmember with moderate thermal maturity; (I–L) lacustrine (L2) endmember with high thermal maturity. Each row shows the HC, N, O, and S classes, respectively. Color scales represent normalized relative abundances.

https://doi.org/10.1021/acs.energyfuels.5c02307
Energy Fuels 2025, 39, 15648−15658

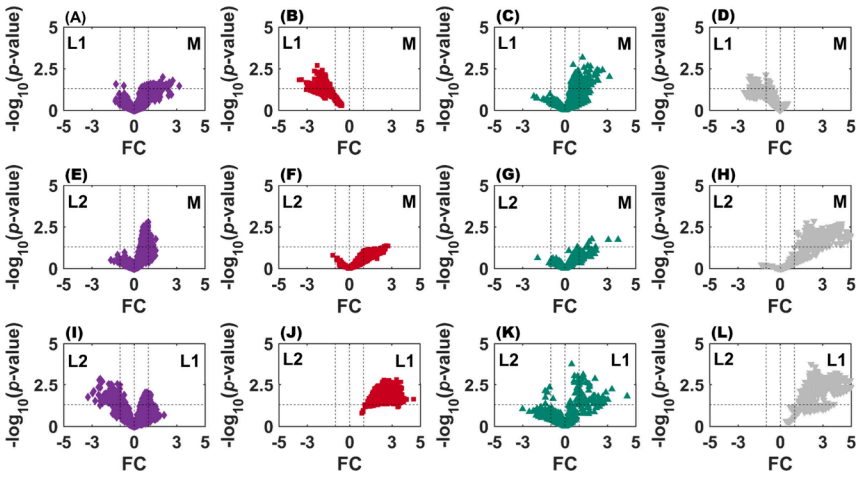


Figure 5. Volcano plot comparing the endmembers based on each heteroatomic class assessed by APPI (+) FT-ICR MS: (A–D) marine crude oils (M) and lacustrine crude oils with moderate thermal maturation (L1); (E–H) marine crude oil and lacustrine crude oils with high thermal maturation (L2); and (I–L) lacustrine crude oils with high thermal maturation and lacustrine crude oil with moderate thermal maturation. Purple diamond symbol indicates hydrocarbon class (HC); red square indicates nitrogen class (N); green upward-pointing triangle indicates oxygen class (O); and gray downward-pointing triangle indicates sulfur class (S).

For both samples, the degrees of alkylation and aromatization appear very similar, indicating that thermal maturity is not the preponderant to distinguishing these samples. Possibly the organic matter input as well as other geological factors such as lithology of each basin and migration must be strongly influencing the compounds distribution.[37]

In addition to classical petroleomics approaches, graphical tools such as volcano plots can provide deeper insights by highlighting statistically significant differences between the conditions or sample groups. Volcano plots are particularly effective in visualizing both the magnitude and significance of variation across multiple molecular classes.[38] In these plots, the *y*-axis represents statistical significance, expressed as the negative logarithm (base 10) of the *p*-value derived from *t* tests between groups. Variables with *p*-values below 0.05, corresponding to a $-\log(p\text{-value})$ greater than 1.3010, are considered statistically significant. The *x*-axis, in turn, displays the fold change (FC), calculated as the logarithm base 2 of the ratio between two sample groups, reflecting the magnitude of the difference.

In this study, volcano plots were applied to compare the endmember samples used in the preparation of the marine and lacustrine blends. As shown in Figure 5, the distribution patterns observed in the volcano plots align well with earlier class distribution findings. When marine-origin samples are compared with lacustrine samples of moderate thermal evolution, the HC (Figure 5A) and O (Figure 5C) classes show a greater number of statistically significant variables in the M endmember. In contrast, the N (Figure 5B) and S (Figure 5D) classes show more significant variables in the L1 endmember, indicating compositional differences in these heteroatomic classes between sample groups.

Further comparisons between the M and L2 samples reveal that the HC (Figure 5E) and S (Figure 5H) classes contain numerous statistically significant variables. Notably, the S class features a substantial number of variables with both high fold change and strong statistical significance, suggesting it is a particularly robust marker for distinguishing between oils with characteristics of the M and L2 groups. When comparing the two lacustrine oils, L1 and L2, the HC class is more abundant in L2, while compounds from other classes are predominantly associated with the L1 blend. These observations reinforce the utility of volcano plots as powerful comparative tools for evaluating differences in molecular composition among the crude oil blends derived from marine and lacustrine origins. The ten most statistically significant compounds for each class were presented in the Supporting Information (Table S7).

3.2. Characterization of Mixtures. Figure 6 presents the class distribution for Plan 1. As the proportion of the L1 blend

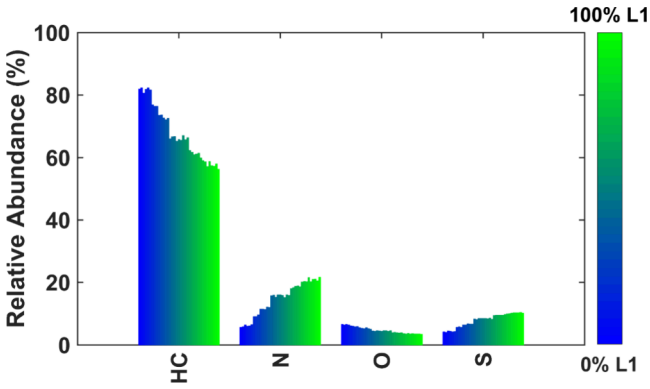


Figure 6. Class distribution for all mixture design points in Plan 1, obtained by APPI (+) FT-ICR MS analysis. Bars represent the relative abundances of HC, N, O, and S classes across the mixtures. The samples are ordered based on increasing marine (M) content, with the corresponding proportion of the lacustrine (L1) blend indicated by the color gradient from 0% to 100%.

increases, a clear trend is observed: the relative abundance of the HC class decreases, while the S and N classes exhibit the opposite behavior with their relative abundances increasing. The abundance of the O class shows a slight decrease as the L1 content rises. These results indicate that increasing the proportion of the L1 blend progressively enhances its molecular signature within the mixtures. Furthermore, additional molecular characteristics, such as carbon number distribution and DBE (double bond equivalent) profiles, shown in Figure S2, follow trends similar to those observed

https://doi.org/10.1021/acs.energyfuels.5c02307
Energy Fuels 2025, 39, 15648–15658

in the class distribution. However, these variations are less pronounced, reinforcing the dominance of class-level compositional changes in distinguishing the influence of the L1 blend.

In Plan 2, we investigated the molecular class behavior in mixtures prepared from the M and L2 endmembers, which are known to exhibit substantial compositional differences. The class distributions for these mixtures are shown in Figure 7.

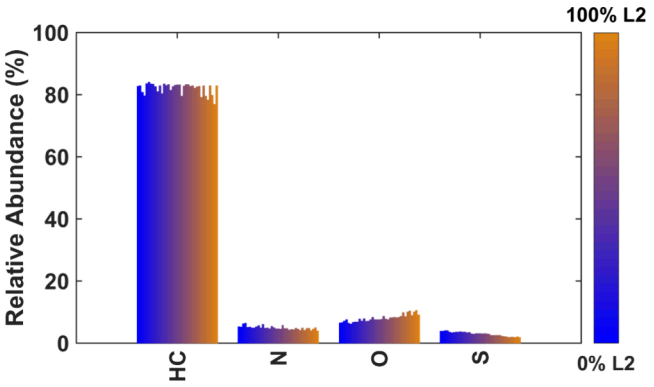


Figure 7. Class distribution for all mixture design points in Plan 2 obtained by APPI (+) FT-ICR MS analysis. The bars represent the relative abundances of HC, N, O, and S classes across mixtures. The samples are ordered according to the increasing proportion of the marine (M) blend. The color gradient indicates the proportion of the L2 blend, from 0% to 100%.

Unlike the patterns observed in Plan 1, blending M and L2 does not produce pronounced or linear trends in the relative abundances of the molecular classes. Heteroatomic S and N classes show a slight decrease in relative intensity as L2 blend proportion increases. O class increases slightly as the L2 blend proportion increases. In contrast, the HC class did not show a clear trend, maybe due to the similarity between the HC content in M and L2 blends.

To further explore the compositional characteristics of these mixtures, we analyzed the DBE and carbon number distributions for each class, as presented in Figure 8.

The DBE and #C profiles appear similar across all mixtures, indicating limited structural variation. However, a subtle trend is observed for the S class: mixtures with a higher proportion of the L2 blend tend to exhibit lower DBE values, while those with a higher proportion of the M blend show elevated DBE values. This suggests that the structural complexity of sulfur-containing compounds is influenced by the relative contribution of each endmember.

In contrast, the #C distributions do not display any meaningful relationship with the blend composition. No evident shift in carbon number correlates with the increasing proportion of either the marine or lacustrine oil nor is there a strong association between #C and the relative abundance trends observed in the molecular classes.

3.3. MCR-ALS. MCR-ALS was applied to deconvolute the spectral data and reveal the underlying components in both mixture designs. The singular value decomposition (SVD) algorithm indicated the presence of two significant pure components in each data set, which correspond directly to the lacustrine and marine endmembers. Figure 9 presents both the original APPI (+) FT-ICR MS spectra and the pure component profiles extracted using MCR-ALS for Plans 1 and 2.

In Plan 1, strong correlations were observed between the original blend spectra and the respective pure components. The L1 blend showed a Pearson correlation of 0.997 with its recovered profile, while the M blend exhibited a correlation of 0.990. For Plan 2, the M blend correlated at 0.995 with its pure component and the L2 blend showed a slightly lower correlation of 0.969. These values indicate that MCR-ALS successfully resolved the primary compositional profiles associated with each blend.

Figure 10 compares the real blend compositions to the concentrations recovered by the MCR-ALS method. These values were presented in Tables S8 and S9 for Plan 1 and Plan 2, respectively. Panels A and B show the results for Plan 1, illustrating the estimated proportions of lacustrine (Figure 10A) and marine (Figure 10B) crude oils in the mixtures. Panels C and D display the corresponding results for Plan 2, with the estimated concentrations of lacustrine (Figure 10C) and marine (Figure 10D) components.

For Plan 1, the recovered concentration profiles are highly consistent with the actual blend compositions. The Pearson correlation coefficient between the first pure component and the L1 content is 0.970. Due to the closure constraint applied in MCR-ALS, the second pure component, associated with the M blend, shows the same correlation with its respective concentration profile. These strong correlations confirm that

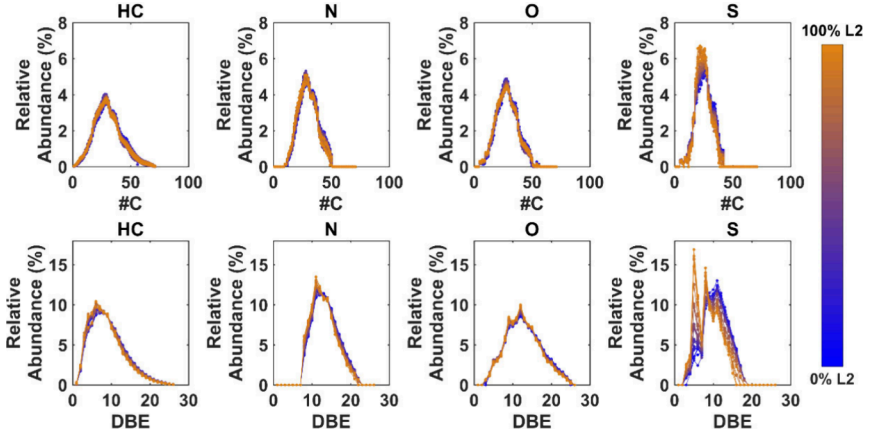


Figure 8. Carbon number (#C) and double bond equivalent (DBE) distributions for Plan 2 crude oil mixtures obtained by APPI (+) FT-ICR MS for the HC, N, and S classes. The color gradient indicates the proportion of the L2 blend in the mixtures, ranging from 0% to 100%.

https://doi.org/10.1021/acs.energyfuels.5c02307
Energy Fuels 2025, 39, 15648−15658

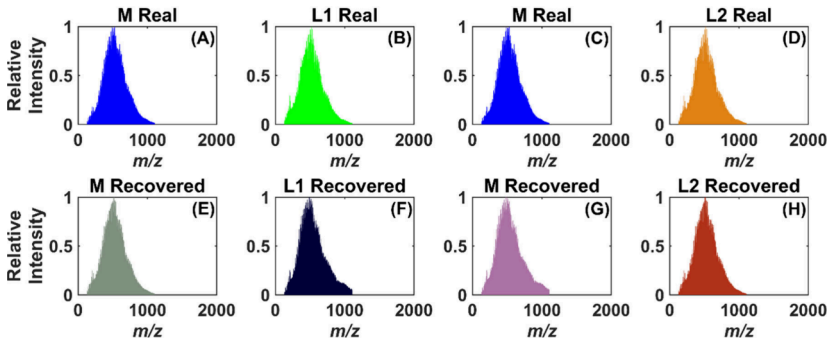


Figure 9. Real and recovered mass spectra for the marine and lacustrine endmember blends obtained by APPI (+) FT-ICR MS. (A–D) Real spectra for the marine (M) and lacustrine blends with moderate (L1) and high (L2) thermal maturity. (E–H) Corresponding pure spectral profiles recovered by MCR-ALS. Plan 1 includes the M (A) and L1 (B) real spectra and their recovered spectra (E and F). Plan 2 consists of the M (C) and L2 (D) real spectra and their recovered spectra (G and H), respectively.

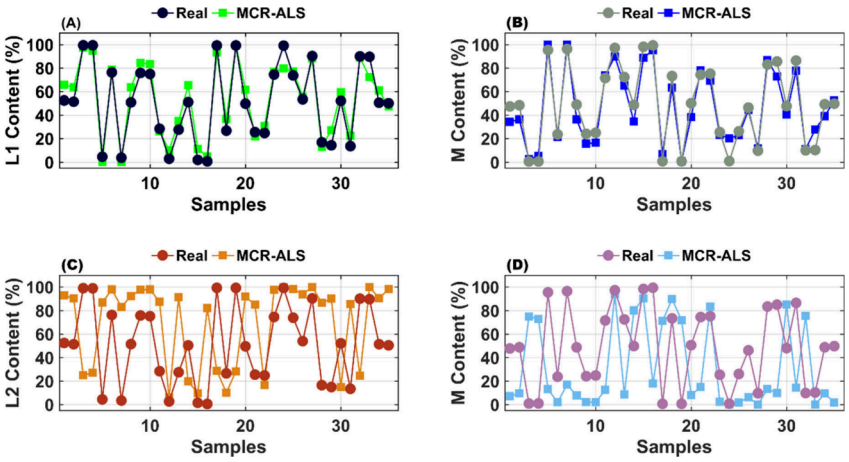


Figure 10. Comparison of values recovered from MCR-ALS with the real percentages of lacustrine and marine contents in crude oil mixtures for Plans 1 (A, B) and 2 (C, D). In Plan 1 (A), midnight blue circles represent the real content of lacustrine with moderate thermal maturation crude oil in each mixture, while gray squares represent the pure concentration profile recovered from MCR-ALS. The content of marine crude oil (B) was expressed using green circles for the real content and blue squares for the pure concentration profile recovered from MCR-ALS. Regarding Plan 2, the content of lacustrine with high thermal maturation (C) was expressed using firebrick circles or the real content and orange squares represent the pure concentration profile recovered from MCR-ALS. For the content of marine (D), purple-pink circles are for the real content and blue squares for the pure concentration profile recovered from MCR-ALS.

MCR-ALS successfully captured the concentration dynamics of the mixture. In contrast, the results for Plan 2 reveal poor agreement between the recovered and the actual concentrations. The correlation between the first pure component and the L2 content is only 0.0244, with an identical correlation observed for the M content. These low values indicate that, for Plan 2, MCR-ALS was not able to accurately resolve the concentration profiles, possibly due to greater compositional similarity between the M and L2 blends.

Therefore, concerning concentration profiles, the MCR-ALS analysis demonstrated satisfactory performance only for Plan 1. This mixture design involved blends that differed both in origin (marine vs lacustrine) and thermal maturity, two factors that enhance chemical contrast between the endmembers. As previously discussed, increasing thermal maturity leads to a decrease in heteroatomic compound content and a corresponding increase in hydrocarbon abundance, which can influence the similarity between crude oils. The successful resolution observed in Plan 1 highlights the utility of MCR-ALS as a chemometric tool for investigating compositional contributions in crude oil mixtures derived from distinct sources. However, as seen in Plan 2, high compositional similarity, such as that between thermally mature marine and lacustrine oils, can pose a challenge for accurate concentration profiling by using this method. Nonetheless, MCR-ALS represents a promising and innovative approach in organic geochemistry, particularly for studies of the blending behavior of oils with varying origins and properties. Its application in this work reinforces its potential for advancing molecular-level interpretations of complex crude oil systems.

3.4. PLS Regression. To improve model robustness, pretreatments were applied to enhance the signal-to-noise ratio and reduce systematic errors in the data. A combination of transformation and preprocessing strategies was tested to identify the optimal configuration. Based on the evaluation of the lowest RMSECV and the highest R_{cv}, Norma Type I (for transformation) and autoscaling (for preprocessing) were selected for both Plan 1 and Plan 2 data sets (see Table S10).

Table 2 presents the statistical performance metrics obtained from PLS regression models built using either all variables (Full) or specific subsets of variables based on compound class: hydrocarbons (HC), nitrogen-containing (N), oxygen-containing (O), sulfur-containing (S), and the combined heteroatomic classes (NOS: N; O, and S). Models were

https://doi.org/10.1021/acs.energyfuels.5c02307
Energy Fuels 2025, 39, 15648−15658

Table 2. Statistical Parameters Obtained in PLS Regression Models for the Prediction of Lacustrine Content in Crude Oil Mixtures[a]

	nVars	LVs	R_C	RMSEC	R_{CV}	RMSECV	R_P	RMSEP
P1Full	3777	5	0.999	1.07	0.994	4.07	0.980	6.94
P1NOS	2093	6	0.998	2.25	0.986	5.91	0.958	9.11
P1HC	1684	5	0.999	1.61	0.994	3.88	0.994	4.35
P1N	903	6	0.993	4.01	0.941	12.06	0.797	18.98
P1O	531	4	0.998	1.79	0.996	3.02	0.986	5.97
P1S	659	6	0.996	3.06	0.953	11.29	0.947	10.07
P2Full	4005	3	0.996	2.92	0.992	4.29	0.990	4.46
P2NOS	1939	5	0.999	0.58	0.989	4.94	0.993	3.71
P2HC	2066	5	0.998	1.85	0.987	5.47	0.982	6.10
P2N	604	5	0.999	0.61	0.982	6.46	0.985	5.69
P2O	973	3	0.996	3.12	0.977	7.47	0.995	3.71
P2S	362	2	0.983	6.26	0.963	9.62	0.985	5.67

[a]nVars, number of variables; LVs, number of latent variables; R_c, correlation coefficient for calibration; RMSEC, root-mean-square error for calibration; R_{cv}, correlation coefficient for cross-validation; RMSECV, root-mean-square error for cross-validation; R_p, correlation coefficient for prediction; RMSEP, root-mean-square error for prediction; P1Full, PLS regression model with all variables for Plan 1; P1NOS, PLS regression model with heteroatomic-containing compounds for Plan 1; P1HC, PLS regression model with hydrocarbon-class for Plan 1; P1N, PLS regression model with nitrogen class for Plan 1; P1O, PLS regression model with oxygen class for Plan 1; P1S, PLS regression model with sulfur class for Plan 1; P2Full, PLS regression model with all variables for Plan 2; P2NOS, PLS regression model with heteroatomic-containing compounds for Plan 2; P2HC, PLS regression model with hydrocarbon class for Plan 2; P2N, PLS regression model with nitrogen class for Plan 2; P2O, PLS regression model with oxygen class for Plan 2; P2S, PLS regression model with sulfur class for Plan 2.

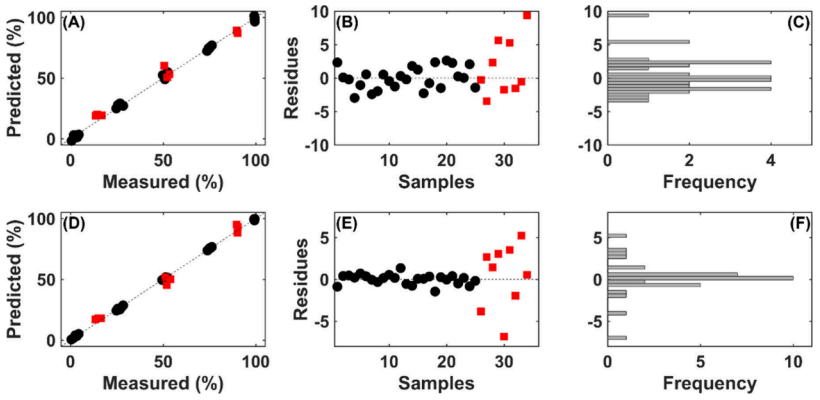


Figure 11. (A) Measured and predicted lacustrine (L1) blend contents for the model using the HC class of Plan 1 (P1HC), (B) residues (black filled circles, calibration samples set; red squares, validation samples set) for the model using the HC class of Plan 1, and (C) histogram of residues. (D) Measured and predicted lacustrine (L2) blend content for the model using nitrogen, oxygen, and sulfur compounds of Plan 2 (P2NOS), (E) residues for the model using nitrogen, oxygen, and sulfur compounds of Plan 2 (P2NOS), and (F) histogram of residues.

evaluated for both Plan 1 (P1) and Plan 2 (P2) mixture designs.

For Plan 1, the full model (P1Full), which includes all molecular classes, achieved the lowest RMSEC (calibration error), indicating good model fit. However, the model based solely on the HC class (P1HC) demonstrated superior predictive performance, as evidenced by the lowest RMSEP (prediction error) and highest R_p (correlation for prediction). The P1O model, based on oxygen-containing compounds, also outperformed the P1Full model in terms of RMSEP. In contrast, models based on N, S, and NOS classes (P1N, P1S, and P1NOS) exhibited relatively high RMSEC, RMSECV, and RMSEP values, indicating a limited predictive utility for estimating the L1 content.

In Plan 2, the models showed different behavior. While the RMSEC values for P2NOS and P2N were both below 1, their predictive performance was moderate with similar RMSEP and R_p values. Interestingly, the HC-based model (P2HC) performed the worst, displaying the highest RMSEP and lowest R_p. These findings suggest that unlike in Plan 1, heteroatomic compound classes (especially N and NOS) are more informative for modeling mixtures of thermally mature crude oils in Plan 2. This aligns with previous findings indicating that heteroatomic species can serve as valuable markers in high maturity oil systems.[39]

Figure 11 presents the measured versus predicted lacustrine crude oil content values and the corresponding residuals for the best-performing PLS regression models in Plans 1 and Plan 2. Panels A and D of Figure 11 show strong linear relationships between the measured and predicted values for both plans, confirming the models' capacity to accurately estimate lacustrine crude oil content in marine−lacustrine mixtures. The fit observed indicates that both models are well-calibrated for their respective data sets. Panels B and E of Figure 11 display the residuals associated with these predictions. They show that the residuals for the calibration samples are randomly distributed around zero and fall mostly within the ±5 range. As expected, the residuals for the validation sets are slightly more dispersed compared with the calibration sets, reflecting the difference between the model training and testing

https://doi.org/10.1021/acs.energyfuels.5c02307
Energy Fuels 2025, 39, 15648−15658

phases. Furthermore, panels C and F of Figure 11 present the histograms of residuals for each plan. The near-normal distribution of residuals in both cases supports the assumption of homoscedasticity and indicates that no significant systematic bias is present in the models. Finally, for completeness, all measured vs predicted values for lacustrine crude oil content in Plan 1 and Plan 2 mixtures are provided in the Supporting Information (Figures S3 and S4, respectively) and the values predicted are present in Tables S11 and S12.

Therefore, PLS regression models proved to be highly effective chemometric tools for predicting the proportion of lacustrine crude oil in mixtures with marine counterparts. The P1HC model yielded the best predictive performance for estimating the L1 content in M and L1 mixtures, while the P2NOS model performed best for predicting the L2 content in M and L2 mixtures. These models demonstrate the potential to investigate crude oil mixtures with varying origins and thermal maturities. This innovative strategy offers a promising solution to ongoing challenges in the petroleum industry, especially those related to the compositional analysis of complex mixtures within petroleum systems. By synthesizing mixtures from known endmembers, it becomes feasible to construct highly accurate predictive models that distinguish crude oils based on origin and thermal history. This work demonstrates that such models provide a robust framework for improving the molecular-level understanding and quantification of oil blending in geochemically diverse systems.

4. CONCLUSIONS

This study established the feasibility of predicting the content of lacustrine crude oil in complex mixtures containing both lacustrine and marine origins through the integration of APPI (+) FT-ICR MS molecular characterization with advanced chemometric modeling. The methodology proved to be effective across different thermal maturities, offering a versatile analytical framework for real-world petroleum systems where compositional variability is inherent.

The strategic selection of molecular classes was shown to significantly influence the model performance. In particular, the use of hydrocarbon- and oxygen-containing species in Plan 1 (P1HC and P1O models) and nitrogen- and sulfur-containing compounds in Plan 2 (P2NOS and P2O models) enhanced the predictive capacity of PLS regression models.

Notably, the results demonstrate the potential of PLS regression as a predictive tool and a diagnostic method for understanding the molecular behavior of crude oil mixtures. The modeling of spectral data using MCR-ALS and PLS provided insights into how source origin and thermal evolution influence blend behavior, supporting compositional differentiation and quantitative estimation. Beyond the immediate context of this study, the approach introduced here offers a new paradigm for integrating petroleomics and chemometrics. By leveraging UHRMS data with multivariate analysis, this framework enables deconvolution of overlapping chemical profiles, predicts mixture composition, and investigates subtle geochemical signatures.

This work lays the groundwork for future applications in reservoir management, forensic oil analysis, and enhanced oil recovery, where rapid and accurate quantification of source contributions is critical. The methods and findings presented here expand the analytical toolbox available to petroleum geochemists and open promising avenues for the molecular-level interpretation of crude oil systems with increasing resolution and confidence.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.energyfuels.5c02307.

More in-depth details; tables describing sample characteristics, biomarkers, percentage values for each design point, acquisition and processing parameters, carbon number and DBE value ranges, top 10 compounds that contribute to sample group differentiation, recovered values from MCR-ALS, predicted PLS values, and types of preprocessing, latent variable numbers, and decision-making statistical parameters for best PLS regression model; class distribution for blends, carbon number distribution, and mixture DBE distributions; graphs of measured versus predicted values (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Jussara V. Roque − *Institute of Chemistry, Federal University of Goiás, 74690-900 Goiânia, Goiás, Brazil;* orcid.org/0000-0002-5220-959X; Email: jussararoque@gmail.com

Boniek Gontijo − *Institute of Chemistry, Federal University of Goiás, 74690-900 Goiânia, Goiás, Brazil;* orcid.org/0000-0003-1197-4284; Email: boniek@ufg.br

Authors

João Victor A. Oliveira − *Institute of Chemistry, Federal University of Goiás, 74690-900 Goiânia, Goiás, Brazil;* orcid.org/0000-0002-3522-2952

Danielle M. M. Franco − *Institute of Chemistry, Federal University of Goiás, 74690-900 Goiânia, Goiás, Brazil;* orcid.org/0000-0002-3691-4328

Leonardo F. Matos − *Institute of Chemistry, Federal University of Goiás, 74690-900 Goiânia, Goiás, Brazil*

Mario D. Rangel − *Institute of Chemistry, Federal University of Goiás, 74690-900 Goiânia, Goiás, Brazil*

Joelma P. Lopes − *Division of Geochemistry, PETROBRAS Research and Development Center (CENPES), 21941-915 Rio de Janeiro, Rio de Janeiro, Brazil*

Ygor dos S. Rocha − *Division of Geochemistry, PETROBRAS Research and Development Center (CENPES), 21941-915 Rio de Janeiro, Rio de Janeiro, Brazil*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.energyfuels.5c02307

Funding

The Article Processing Charge for the publication of this research was funded by the Coordenacao de Aperfeicoamento de Pessoal de Nivel Superior (CAPES), Brazil (ROR identifier: 00x0ma614).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. The authors express gratitude to CENPES/Petrobras and Brazilian Research Council (Conselho Nacional de Desenvolvimento Científico e

Tecnológico—CNPq) for providing the samples and for their financial support (Processo 140732/2024-1).

REFERENCES

(1) Maxwell, J. R.; Pillinger, C. T.; Eglinton, G. Organic Geochemistry. *Q. Rev. Chem. Soc.* **1971**, *25* (4), 571.
(2) Philp, R. P.; Lewis, C. A. Organic Geochemistry of Biomarkers. *Annu. Rev. Earth Planet. Sci.* **1987**, *15* (1), 363−395.
(3) Peters, K. E.; Walters, C. C.; Moldowan, J. M. *The Biomarker Guide: Vol. 1, Biomarkers and Isotopes in the Environment and Human History*, 2nd ed.; Cambridge University Press, 2005. DOI: 10.1017/CBO9780511524868.
(4) Tissot, B.; Durand, B.; Espitalie, J.; Combaz, A. Influence of Nature and Diagenesis of Organic Matter in Formation of Petroleum. *Am. Assoc. Pet. Geol. Bull.* **1974**, *58* (3), 499−506.
(5) Philp, R. P. Formation and Geochemistry of Oil and Gas. In *Treatise on Geochemistry*, 2nd ed.; Elsevier, 2014; pp 233−265. DOI: 10.1016/B978-0-08-095975-7.00709-9.
(6) Jia, W.; Wang, Q.; Peng, P.; Xiao, Z.; Li, B. Isotopic Compositions and Biomarkers in Crude Oils from the Tarim Basin: Oil Maturity and Oil Mixing. *Org. Geochem.* **2013**, *57*, 95−106.
(7) Niu, J.; Huang, H.; Jiang, W. Geochemical Characteristics and Correlation of Continuous Charge Mixing and Biodegradation of Heavy Oil in Southeastern Dongying Sag, Bohai Bay Basin, China. *J. Pet. Sci. Eng.* **2018**, *166*, 1−12.
(8) Wang, Q.; Huang, H.; Li, Z. Mixing Scenario of a Vagarious Oil in the Dongying Depression, Bohai Bay Basin. *Fuel* **2021**, *294*, No. 120589.
(9) Ross, A. S.; Farrimond, P.; Erdmann, M.; Larter, S. R. Geochemical Compositional Gradients in a Mixed Oil Reservoir Indicative of Ongoing Biodegradation. *Org. Geochem.* **2010**, *41* (3), 307−320.
(10) Zhang, S.; Huang, H.; Su, J.; Zhu, G.; Wang, X.; Larter, S. Geochemistry of Paleozoic Marine Oils from the Tarim Basin, NW China. Part 4: Paleobiodegradation and Oil Charge Mixing. *Org. Geochem.* **2014**, *67*, 41−57.
(11) Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. Fourier Transform Ion Cyclotron Resonance Mass Spectrometry: A Primer. *Mass Spectrom. Rev.* **1998**, *17* (1), 1−35.
(12) Cho, Y.; Ahmed, A.; Islam, A.; Kim, S. Developments in FT-ICR MS Instrumentation, Ionization Techniques, and Data Interpretation Methods for Petroleomics. *Mass Spectrom. Rev.* **2015**, *34* (2), 248−263.
(13) Rocha, Y. d. S.; Pereira, R. C. L.; Mendonça Filho, J. G. Geochemical Characterization of Lacustrine and Marine Oils from Off-Shore Brazilian Sedimentary Basins Using Negative-Ion Electrospray Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (ESI FTICR-MS). *Org. Geochem.* **2018**, *124*, 29−45.
(14) Monged, M.; Poetz, S.; Noah, M.; Wilkes, H.; Schulz, H.-M.; Bakr, M. M. Y. Reservoir Oil Geochemistry in the Northern Western Desert of Egypt − High-Resolution Mass Spectrometry and Stable Carbon Isotope Characterization. *Org. Geochem.* **2023**, *176*, No. 104545.
(15) Hughey, C. A.; Rodgers, R. P.; Marshall, A. G.; Qian, K.; Robbins, W. K. Identification of Acidic NSO Compounds in Crude Oils of Different Geochemical Origins by Negative Ion Electrospray Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Org. Geochem.* **2002**, *33* (7), 743−759.
(16) Hosseini, S. H.; Sachsenhofer, R. F.; Snowdon, L. R.; Oldenburg, T. B. P. Characterization of Crude Oils Derived from Carbonate and Siliciclastic Source Rocks Using FTICR-MS. *Org. Geochem.* **2021**, *159*, No. 104286.
(17) Han, Y.; Noah, M.; Lüders, V.; Körmös, S.; Schubert, F.; Poetz, S.; Horsfield, B.; Mangelsdorf, K. Fractionation of Hydrocarbons and NSO-Compounds during Primary Oil Migration Revealed by High Resolution Mass Spectrometry: Insights from Oil Trapped in Fluid Inclusions. *Int. J. Coal Geol.* **2022**, *254*, No. 103974.
(18) Oldenburg, T. B. P.; Brown, M.; Bennett, B.; Larter, S. R. The Impact of Thermal Maturity Level on the Composition of Crude Oils, Assessed Using Ultra-High Resolution Mass Spectrometry. *Org. Geochem.* **2014**, *75*, 151−168.
(19) Poetz, S.; Horsfield, B.; Wilkes, H. Maturity-Driven Generation and Transformation of Acidic Compounds in the Organic-Rich Posidonia Shale as Revealed by Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Energy Fuels* **2014**, *28* (8), 4877−4888.
(20) Martins, L. L.; Pudenzi, M. A.; da Cruz, G. F.; Nascimento, H. D. L.; Eberlin, M. N. Assessing Biodegradation of Brazilian Crude Oils via Characteristic Profiles of O 1 and O 2 Compound Classes: Petroleomics by Negative-Ion Mode Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Energy Fuels* **2017**, *31* (7), 6649−6657.
(21) Oldenburg, T. B. P.; Jones, M.; Huang, H.; Bennett, B.; Shafiee, N. S.; Head, I.; Larter, S. R. The Controls on the Composition of Biodegraded Oils in the Deep Subsurface − Part 4. Destruction and Production of High Molecular Weight Non-Hydrocarbon Species and Destruction of Aromatic Hydrocarbons during Progressive in-Reservoir Biodegradation. *Org. Geochem.* **2017**, *114*, 57−80.
(22) Vaz, B. G.; Silva, R. C.; Klitzke, C. F.; Simas, R. C.; Lopes Nascimento, H. D.; Pereira, R. C. L.; Garcia, D. F.; Eberlin, M. N.; Azevedo, D. A. Assessing Biodegradation in the Llanos Orientales Crude Oils by Electrospray Ionization Ultrahigh Resolution and Accuracy Fourier Transform Mass Spectrometry and Chemometric Analysis. *Energy Fuels* **2013**, *27* (3), 1277−1284.
(23) Marshall, A. G.; Rodgers, R. P. Petroleomics: Chemistry of the Underworld. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (47), 18090−18095.
(24) Angolini, C. F. F. F.; Pudenzi, M. A.; Batezelli, A.; Eberlin, M. N. Comprehensive Petroleomics: Multiple Mass Spectrometry Strategies for Crude Oil Characterization. In *Encyclopedia of Analytical Chemistry*; Major Reference Works; Wiley, 2017; pp 1−16. DOI: 10.1002/9780470027318.a9588.
(25) Borisov, R. S.; Kulikova, L. N.; Zaikin, V. G. Mass Spectrometry in Petroleum Chemistry (Petroleomics) (Review). *Pet. Chem.* **2019**, *59* (10), 1055−1076.
(26) Melendez-Perez, J. J.; Campos Oliveira, L. F.; Miranda, N.; Sussulini, A.; Eberlin, M. N.; Bastos, W. L.; Rangel, M. D.; Rocha, Y. d. S. Lacustrine versus Marine Oils: Fast and Accurate Molecular Discrimination via Electrospray Fourier Transform Ion Cyclotron Resonance Mass Spectrometry and Multivariate Statistics. *Energy Fuels* **2020**, *34* (8), 9222−9230.
(27) Eriksson, L.; Johansson, E.; Wikström, C. Mixture Design—Design Generation, PLS Analysis, and Model Usage. *Chemom. Intell. Lab. Syst.* **1998**, *43* (1), 1−24.
(28) de Juan, A.; Tauler, R. Multivariate Curve Resolution: 50 Years Addressing the Mixture Analysis Problem − A Review. *Anal. Chim. Acta* **2021**, *1145*, 59−78.
(29) Ruckebusch, C.; Blanchet, L. Multivariate Curve Resolution: A Review of Advanced and Tailored Applications and Challenges. *Anal. Chim. Acta* **2013**, *765*, 28−36.
(30) de Juan, A.; Jaumot, J.; Tauler, R. Multivariate Curve Resolution (MCR). Solving the Mixture Analysis Problem. *Anal. Methods* **2014**, *6* (14), 4964−4976.
(31) Ghatee, M. H.; Hemmateenejad, B.; Sedghamiz, T.; Khosousi, T.; Ayatollahi, S.; Seiedi, O.; Sayyad Amin, J. Multivariate Curve Resolution Alternating Least-Squares As a Tool for Analyzing Crude Oil Extracted Asphaltene Samples. *Energy Fuels* **2012**, *26* (9), 5663−5671.
(32) Jaumot, J.; de Juan, A.; Tauler, R. MCR-ALS GUI 2.0: New Features and Applications. *Chemom. Intell. Lab. Syst.* **2015**, *140*, 1−12.
(33) Wold, S.; Sjöström, M.; Eriksson, L. PLS-Regression: A Basic Tool of Chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58* (2), 109−130.
(34) Cardoso, W.; Roque, J. V.; Jansen, J. J.; Teng, S. Y.; Teófilo, R. F. Combinatorial Order Pre-Processing Search (COPS): A New Pre-Processing Strategy for Large-Scale Interpretable Data Analysis in Process Analytical Technologies. *Comput. Chem. Eng.* **2025**, *192*, No. 108892.

(35) Hughey, C. A.; Rodgers, R. P.; Marshall, A. G.; Walters, C. C.; Qian, K.; Mankiewicz, P. Acidic and Neutral Polar NSO Compounds in Smackover Oils of Different Thermal Maturity Revealed by Electrospray High Field Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Org. Geochem.* **2004**, *35* (7), 863−880.
(36) Ji, H.; Li, S.; Greenwood, P.; Zhang, H.; Pang, X.; Xu, T.; He, N.; Shi, Q. Geochemical Characteristics and Significance of Heteroatom Compounds in Lacustrine Oils of the Dongpu Depression (Bohai Bay Basin, China) by Negative-Ion Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Mar. Pet. Geol.* **2018**, *97*, 568−591.
(37) Jiang, B.; Tian, Y.; Zhai, Z.; Zhan, Z.-W.; Liao, Y.; Zou, Y.-R.; Peng, P. Characterisation of Heteroatomic Compounds in Free and Bound Bitumen from Different Source Rocks by ESI FT-ICR MS. *Org. Geochem.* **2021**, *151*, No. 104147.
(38) 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* (2), 1206−1212.
(39) Wang, D.; Li, M.; Chen, J.; Chen, H.; Shi, Q. A New Ternary Diagram to Decipher the Evolution of Maturity and Biodegradation of Crude Oil Using ESI FT-ICR MS. *Fuel* **2024**, *359*, No. 130499.

https://doi.org/10.1021/acs.energyfuels.5c02307
Energy Fuels 2025, 39, 15648−15658