Machine learning approach for predicting crude oil stability based on NMR spectroscopy

Dubravka Raljević[a], Jelena Parlov Vuković[a,*], Vilko Smrečki[b,*], Ljiljana Marinić Pajc[a], Predrag Novak[c,*], Tomica Hrenar[c,*], Tomislav Jednačak[c,*], Lucija Konjević[a], Bruno Pinević[c], Tonka Gašparac[c]

[a] *INA-Industrija nafte d.d., Refining & Marketing, Central Testing Laboratory, Lovinčićeva 4, HR-10002 Zagreb, Croatia*

[b] *Ruđer Bošković Institute, NMR Centre, Bijenička 54, HR-10000 Zagreb, Croatia*

[c] *University of Zagreb, Faculty of Science, Department of Chemistry, Horvatovac 102a, HR-10000, Zagreb, Croatia*

* Corresponding authors.

E-mail addresses: jelena.parlov-vukovic@ina.hr (J.P.V.); smrecki@irb.hr (V.S.); pnovak@chem.pmf.hr (P.N.); hrenar@chem.pmf.hr (T.H.); tjednacak@chem.pmf.hr (T.J.)

Keywords: machine learning, crude oil, NMR spectroscopy, stability

Abstract

Crude oils are extremely complex organic mixtures, composed of various constituents ranging in size, shape and polarity. Obtaining a detailed insight into the petroleum composition is of highest priority for quality evaluation of crude oils and crude oil product performances. The stability of crude oils and their components represents one of the major challenges in petroleum industry, since there is no existing single method to determine the stability of all fractions. In this study, statistical multi-way analysis (MWA) and machine learning (ML) methods were coupled with diffusion-ordered NMR spectroscopy (DOSY) and compared to different crude oil stability affecting parameters in order to explore possibilities to predict crude oil stability. The potential of this approach was explored to identify and classify the crude oils of different origin according to their composition, stability, density and diffusion properties. With the application of MWA using the TUCKER3 decomposition model for a set of DOSY NMR spectra, the principal components were determined for the model (5,5,5), which described 99.89% of the total variance. The reduced space of the first 3 principal components was used for the sample classification. Similar samples were identified, and reduced space was further utilized for the regression of measured stabilities. Extensive ML multivariate linear regression was carried out for modeling crude oil stability in relation to DOSY NMR spectra and other measured properties, such as aromaticity, API gravity, percentage of aliphatic chains, asphaltene content and relative diffusivities. In both MWA and ML cases the best predictive models were determined. For such complex mixtures as crude oils are, exceptionally good correlations were obtained, proving that this new and robust model can accurately predict crude oil stability and other important parameters relevant for petroleum industry thus showing a great potential for practical applications.

1. Introduction

Crude oil is a highly complex organic mixture composed of various aliphatic and aromatic hydrocarbons ranging in size, shape and polarity. Obtaining a detailed insight into chemical composition is of highest priority for quality evaluation of crude oils and crude oil product performances [1,2]. According to their polarity, the components of crude oil are often divided into four main groups: asphaltenes, saturates, aromatics and resins. Asphaltenes are the heaviest and the most polar crude oil components, composed of aromatic and saturated rings, aliphatic moieties, some heteroatoms, such as nitrogen, oxygen and sulfur, and traces of transition metals [3–5]. During petroleum processing, asphaltenes may form aggregates and precipitates, leading to serious problems in production, transportation and storage. The stability of asphaltenes in crude oils and petroleum products is one of the major challenges in petroleum industry, since there is no existing single method to determine stability of all oil fractions [6–9].

Nuclear magnetic resonance (NMR) spectroscopy has emerged as a valuable tool for studying crude oils and their derivatives [10–18]. However, proton and carbon NMR spectra of petroleum samples are characterized by severely overlapping signals, which are difficult to straightforwardly assign and analyze. Further insight into the nature and structure of crude oils can be obtained by diffusion ordered NMR spectroscopy (DOSY) [14–18]. This approach can be applied to measure translational diffusion properties of individual components in complex mixtures without their physical separation. DOSY NMR spectra are pseudo-two-dimensional, where one dimension is represented by chemical shifts and the other by translational diffusion coefficients, which depend on the shape and size of a molecule or an aggregate in the sample. Crude oils originating from different geographical regions contain various types of compounds that can be separated and identified according to their diffusion coefficients. Nevertheless, even with the state-of-the-art NMR techniques one is still not able to perform a complete differentiation among crude oil samples based on spectral inspection only (Figs. 2 and 3). Hence, further evaluation and spectral processing by statistical methods are required to explore the correlation between the origin and physical properties of crude oils. Recently, it has been shown that petroleum samples of different origin can be identified, clustered and well-separated by employing a combination of DOSY NMR spectroscopy and multi-way analysis [17]. Moreover, an advanced statistical model based on trilinear decomposition algorithm has been developed, validated and applied to evaluate DOSY NMR spectra. In a similar study, proton NMR spectra have been processed by principal component analysis to reveal characteristic spectral areas responsible for sample differentiation and classification [18].

In this study, multi-way analysis and machine learning methods are combined to predict the crude oil stability based on NMR spectroscopy. For this purpose, capabilities of DOSY NMR coupled with both multi-way and machine learning multivariate linear regression analyses have been explored to identify and classify crude oils of different origin according to their content, stability, density and diffusion properties.

2. Experimental

Samples

All crude oil samples were obtained from geographical regions designated in Tables 1, 2 and S1.

2.1. Asphaltene content analysis

Asphaltenes were extracted from the crude oil samples by employing the standard ASTM D 6560-17 method to determine the content of insoluble asphaltenes in heptane [19]. The crude oil samples were refluxed in heptane and mixed with the precipitate. Subsequently, asphaltenes, waxy substances and inorganic material were collected on a filter paper. In the next step, the waxy substances were removed by washing with hot heptane in an extractor, while the asphaltenes were separated from the inorganic material by dissolving in hot toluene. The extraction solvent was evaporated.

2.2. *Stability testing*

Stability testing was based on the ASTM D 7157-18 standard method [20]. The sample solutions were prepared in toluene at three different concentrations and analyzed by ROFA France automated stability analyzer equipped with an optical probe for detecting the asphaltene flocculation. Stability parameters S_{total} (overall stability of the sample), S_{asph} (peptizability or ability of asphaltenes to remain in a dispersed state) and S_{resin} (aromaticity of the resins and their capability to maintain asphaltenes in solution) were calculated as well as intrinsic stability from volumes of toluene and *n*-heptane and mass of the samples.

2.3. *NMR measurements*

NMR experiments were performed at 298 K and chemical shifts were reported relative to tetramethylsilane (TMS) internal standard. The samples (100 μL) were dissolved in 500 μL of a deuterated solvent. One-dimensional ^{1}H NMR spectra were recorded on a Bruker Avance Neo 300 NMR spectrometer in chloroform-d (99.8%, Aldrich) using a C/H dual 5 mm probe with 32 scans, 10 s recycle delay, 7.6 μs π/2 pulse length and 16 K time domain. ^{1}H NMR DOSY spectra were acquired in toluene-d$_8$ (99.5%, CIL) on a Bruker Avance 600 NMR spectrometer using a 5 mm TBI probe equipped with z-gradients. Triplicate measurements were carried out using a dstebpgp3s pulse sequence with convection compensation, 16 scans, 6.0 kHz spectral width, 600 μs spoil gradients, 16 K time domain, 150 μs gradient recovery and 5 ms eddy current delays. The gradient strength was varied from 2 % to 95 % in 16 steps, while both the small (2.0 ms) and the big (70.0 ms) delta were kept constant.

2.4. *Multi-way analysis*

Multi-way analysis (MWA) presents decomposition of multidimensional datasets represented as multidimensional numerical arrays (or a higher order data tensor). It could be considered as an extension of principal component analysis [21]. Data tensor is composed from sequences of numbers dependent on different physical dimensions or *ways*. In this case, the 3rd order tensor consists of two-dimensional DOSY NMR spectra for different crude oil samples.

Each DOSY NMR spectrum was extracted with 128×2192 records providing the total dimensions of the 3rd order tensor: 18×128×2192. The data in this 3rd order tensor depend on three independent variables: chemical shift, magnetic gradient pulse amplitude and sample diversity [17]. To extract the quantitative classification information, MWA was used as a tool that allows detection of variabilities among all investigated samples based on their 2-dimensional DOSY NMR datasets. After tensor decomposition, each DOSY NMR spectrum was finally represented as one point in reduced space.

MWA on the set of DOSY NMR spectra placed in the 3rd order tensor was carried out using the 3-way decomposition model TUCKER3 [22]:

$$\boldsymbol{X} = \boldsymbol{AG}(\boldsymbol{C} \otimes \boldsymbol{B})^{\tau} + \boldsymbol{E} \qquad (1)$$

where $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ are the 1st-way, 2nd-way, and 3rd-way loadings matrices, respectively (symbol $\otimes$ represents Kronecker matrix product) (Fig. 1).

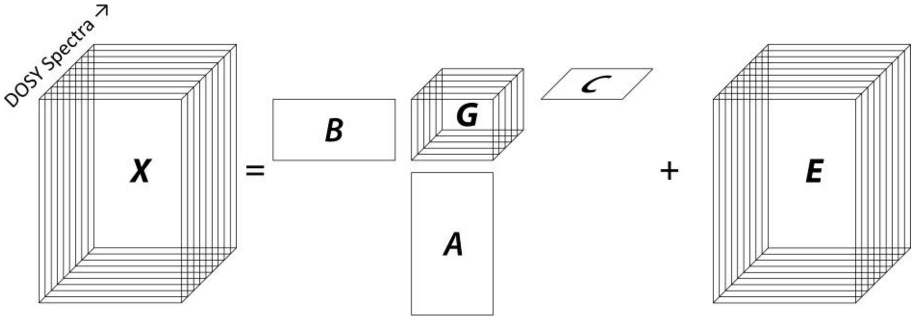

Fig. 1. Graphical representation of the TUCKER3 model.

The $\boldsymbol{G}$ matrix is the *core-array* and is associated with the amount of variation explained by loadings in the different modes. MWA was performed by using the code ***moonee*** [23–26] developed *in-house*.

2.5. Machine learning multivariate linear regression

Measured stability data was used as dependent variable in two cases. Firstly, this data was regressed on the first three principal components of the reduced space of DOSY NMR spectra. In the second case, stability data was regressed to 6 other measured properties. Extensive machine learning (ML) procedure was applied for generation of all possible multivariate linear regression (MLR) models with any possible linear combination of original variables as well as their higher-order polynomial terms (up to the 4th order in the first case and up to the 2nd order in the second case). Total numbers of generated different models for each dependent variable were 1 717 869 184 and 134 217 728 for the first and the second case, respectively.

MLR was performed using the following expression for matrices of coefficients B calculated by singular value decomposition:

$$\boldsymbol{B} = (\boldsymbol{X}^{\tau}\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}^{\tau}\boldsymbol{Y} \qquad (2)$$

where $\boldsymbol{X}$ and $\boldsymbol{Y}$ are the matrices of independent and dependent variables, respectively. For each model, *leave-one-out cross-validation* (LOO-CV) was performed, and various statistical parameters were computed. Methodical validation of models by LOO-CV provided an optimal representation selected on the basis of adjusted and predicted R^2 values as well as LOO-CV mean squared error.

3. Results and Discussion

3.1. Evaluation of crude oil properties by standard testing methods

Chosen crude oil properties, such as the asphaltene content (w_{asph}), stability parameters (S_{asph}, S_{resin}, S_{total}) and API gravity values (ρ_{API}) of analyzed crude oil samples are shown in Table 1. API values indicated that all analyzed crude oils belong to lighter and medium crude oil categories.

The overall stability of crude oils decreases with the increase in asphaltene content. However, a comparison of some test samples show that the crude oil stability does not exclusively depend on the content of asphaltenes, especially in light crude oils. Parameters, such as the composition and structure of resins, aromatics and other components affect the total stability. Furthermore, the stability of crude oils depends on various processes that involve blending, dilution, temperature and pressure changes.

Table 1. Comparison of crude oil properties determined by standard methods.

Sample No.	Designation	w_{asph} / % [a]	S_{asph} [b]	S_{resin} [c]	S_{total} [d]	ρ_{API} [e]
1	North Africa 1	2.50	0.75	0.58	2.33	29.93
2	Southwest Asia 1	0.15	0.41	1.88	3.18	38.15
3	Southwest Asia 2	2.25	0.6	1.15	2.92	31.21
4	Southwest Asia 3	2.74	0.76	0.57	2.39	30.13
5	Central Europe 1	1.19	0.77	0.69	3.02	30.11
6	Eastern Europe 1	0.80	0.77	1.07	4.61	29.58
7	Eastern Europe 2	0.26	0.78	1.01	4.54	36.39
8	Central Europe 2	1.01	0.77	0.73	3.22	30.24
9	West Africa	0.01	0.44	1.63	2.94	32.42
10	Southwest Asia 4	3.12	0.74	0.59	2.25	29.94
11	Southwest Asia 5	1.58	0.77	0.48	2.12	35.24
12	Eastern Europe 3	0.64	0.65	0.95	2.72	35.05
13	NorthEast Asia	1.45	0.77	0.68	2.97	29.58
14	North Africa 2	0.37	0.65	0.81	2.33	37.15
15	North Asia	0.71	0.75	1.15	4.68	33.51
16	Central Europe 3	0.14	0.64	0.97	2.71	37.51
17	Central America	0.62	0.70	1.43	4.69	41.70

[a] asphaltene content according to gravimetric analysis
[b] peptizability or ability of asphaltenes to remain in a dispersed state
[c] aromaticity of the resins and their capability to maintain asphaltenes in solution
[d] total stability or overall stability of the sample
[e] gravity according to American Petroleum Institute (API)

3.2. NMR spectroscopy

Typical proton NMR spectra of crude oil samples are displayed in Fig 2. Severe peak overlapping makes these spectra difficult to analyze and only information on different classes of hydrocarbons can be obtained. One of the features that can be determined from ^{1}H NMR spectra is the aromaticity (H_{ar}), usually expressed as the content (in percentage) of aromatic hydrogen atoms. It can be calculated as the ratio between the sum of all aromatic hydrogen integrals ($I_{\mathrm{H_{ar}}}$) and the total amount of hydrogen atoms (consisting of the sum of all aliphatic and aromatic hydrogen integrals, $I_{\mathrm{H_{aliph}}}$ and $I_{\mathrm{H_{ar}}}$), using the previously described procedure [18,27,28]:

$$H_{\mathrm{ar}}[\%] = \frac{\sum I_{\mathrm{H_{ar}}}}{\sum I_{\mathrm{H_{aliph}}} + \sum I_{\mathrm{H_{ar}}}} \qquad (3)$$

In order to assure more accurate H_{ar} calculation by avoiding overlapping signals of crude oil aromatic hydrogens with toluene aromatic hydrogens, corresponding ^{1}H NMR spectra were measured in deuterated chloroform.

Characteristic signals in the ^{1}H NMR spectra of crude oil samples corresponding to aromatic and aliphatic protons were found in the chemical shift regions 6.5–9.0 and 0.5–4.0 ppm, respectively (Fig. 2). As shown in Table 2, the aromaticity depends on the sample origin, having values in the range of 2.10 % – 7.29 %. If compared with data summarized in Table 1, H_{ar} is well correlated with the asphaltene content, which is in agreement with the presence of condensed aromatic rings in the asphaltene structure. On the other hand, no correlations were observed as expected between the percentage of aliphatic chains calculated from the integral at 1.3 ppm ($I_{1,3\mathrm{ppm}}$) and the asphaltene content, since aliphatic chains are present in all major crude oil components.

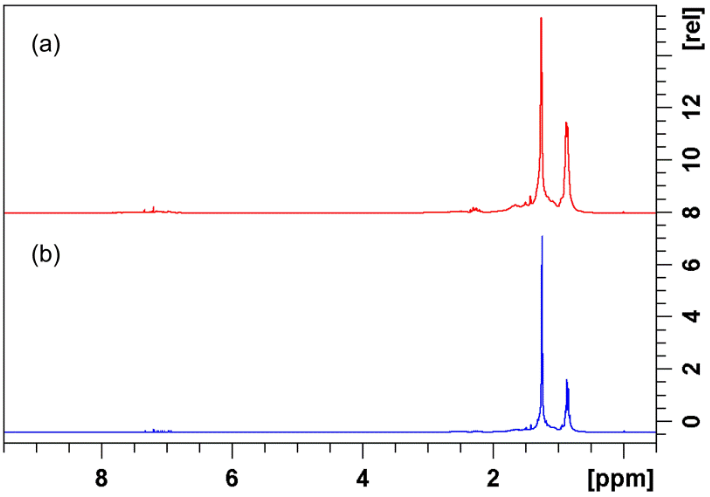

Fig. 2. ^{1}H NMR spectra of (a) Southwest Asia 1 and (b) North Africa 2 crude oil samples.

Table 2. Crude oil parameters calculated from ^{1}H and DOSY NMR spectra.

Sample No.	Designation	H_{ar} / % [a]	$I_{1,3ppm}$ / % [b]	$d_{rel,0.9\ ppm}$ [c]	$d_{rel,1.3\ ppm}$ [d]
1	North Africa 1	5.70	56.75	0.52	0.51
2	Southwest Asia 1	3.37	52.63	0.51	0.48
3	Southwest Asia 2	4.24	56.49	0.55	0.53
4	Southwest Asia 3	4.41	57.78	0.58	0.56
5	Central Europe 1	7.29	61.25	0.53	0.50
6	Eastern Europe 1	4.68	61.39	0.50	0.54
7	Eastern Europe 2	3.69	59.59	0.54	0.54
8	Central Europe 2	4.19	68.64	0.58	0.57
9	West Africa	4.54	44.83	0.50	0.52
10	Southwest Asia 4	5.53	61.40	0.53	0.52
11	Southwest Asia 5	5.66	53.16	0.55	0.54
12	Eastern Europe 3	3.96	60.14	0.51	0.51
13	NorthEast Asia	4.84	59.48	0.57	0.56
14	North Africa 2	2.10	58.92	0.80	0.74
15	North Asia	4.65	79.63	0.57	0.57
16	Central Europe 3	3.49	69.94	0.51	0.51
17	Central America	2.82	63.58	0.77	0.52

[a] aromaticity calculated as the difference between the sum of all signal integrals and those corresponding to aliphatic protons

[b] percentage of aliphatic chains obtained from the integral of the proton signal at 1.3 ppm

[c] relative diffusivities calculated from the DOSY signal at 0.9 ppm

[d] relative diffusivities calculated from the DOSY signal at 1.3 ppm

Further insight into the content and motional behavior of the crude oil components was obtained from DOSY NMR experiments. Representative DOSY NMR spectra of crude oil samples (Fig. 3) revealed differences in the shape and intensity of characteristic peaks. These signals belong to species with different diffusion properties and can be used to distinguish between the samples. Motional behavior of individual components is quantitatively described by their translational diffusion coefficients (*D*). However, the accuracy and reproducibility of the diffusion measurements is largely affected by experimental conditions. This impact can be minimized by introducing the relative diffusivity, $d_{\mathrm{rel}} = D_{\mathrm{sample}}/D_{\mathrm{toluene}}$. As shown for the signals at 0.9 and 1.3 ppm in Table 2, as well as for other resonances in Table S1, only the components of North Africa 2 and Central America samples exhibited considerably higher d_{rel} than average. On the other hand, diffusion properties of other crude oils were very similar to each other, despite their different origin. Hence, in order to separate and classify all crude oils additional information was extracted from DOSY NMR spectra by statistical analysis. For that purpose, an approach was employed that combines MWA and ML methods described in the following chapters.

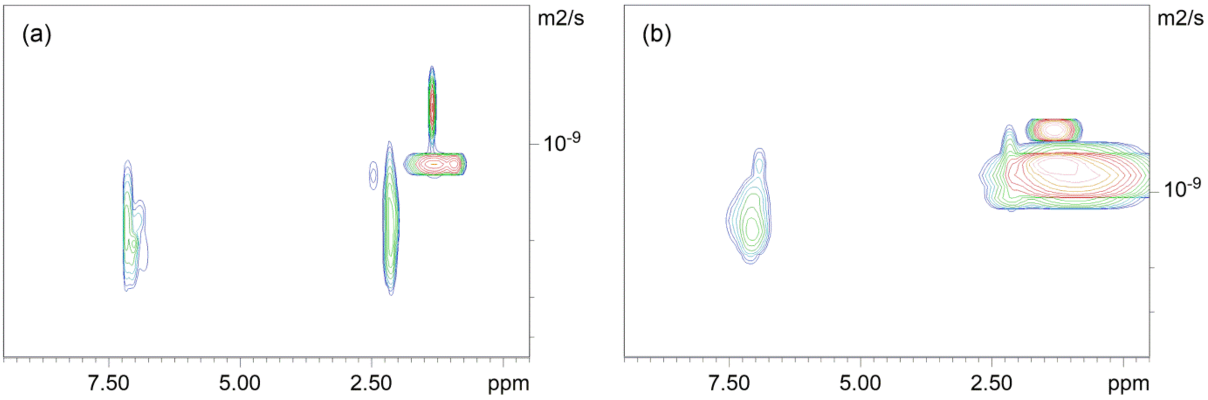

Fig. 3. DOSY NMR spectra of (a) Southwest Asia 1 and (b) North Africa 2 crude oil samples.

3.3. MWA

Using the TUCKER3 decomposition model for a set of DOSY NMR spectra, a *progressive* decomposition model search was performed starting from the model with dimensions (1,1,1). This was the simplest decomposition model which already explained 97.60 % of the total variance (Fig. 4). The search passed through all possible models up to the final tested decomposition model (5,5,5) that described 99.89 % of the total variance. Each dimension was gradually increased by 1 giving the total number of generated models 5×5×5=125. Explained variances for all investigated models are presented in Fig. 4.

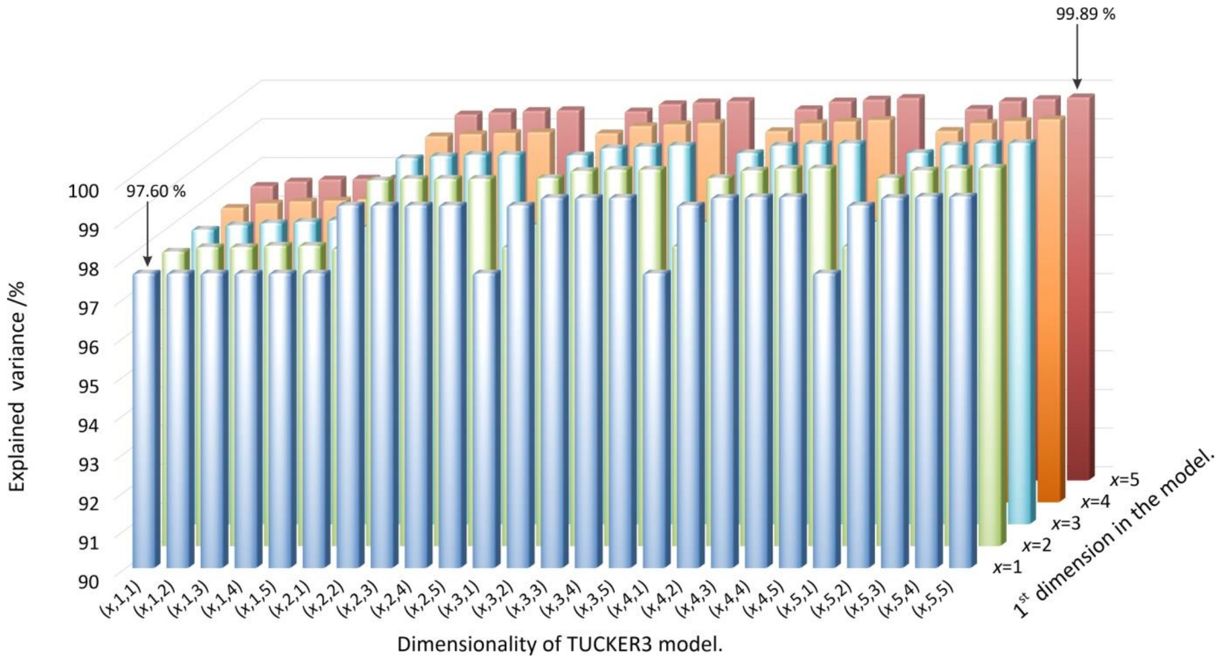

Fig. 4. Explained variance in TUCKER3 models in dependence of model dimensionality used in decomposition of 3rd-order data tensor (DOSY NMR spectra).

Model (5,5,5) was chosen for further analysis, while the first three components from this model were used for classification of samples, visualization and later regression. These three components described 99.72 % of the total variance. Their loadings plots are presented in Fig. 5. This percentage of the total described variance is high enough to ensure that the most important properties of the investigated systems relevant for the proper analysis were retained within the model.

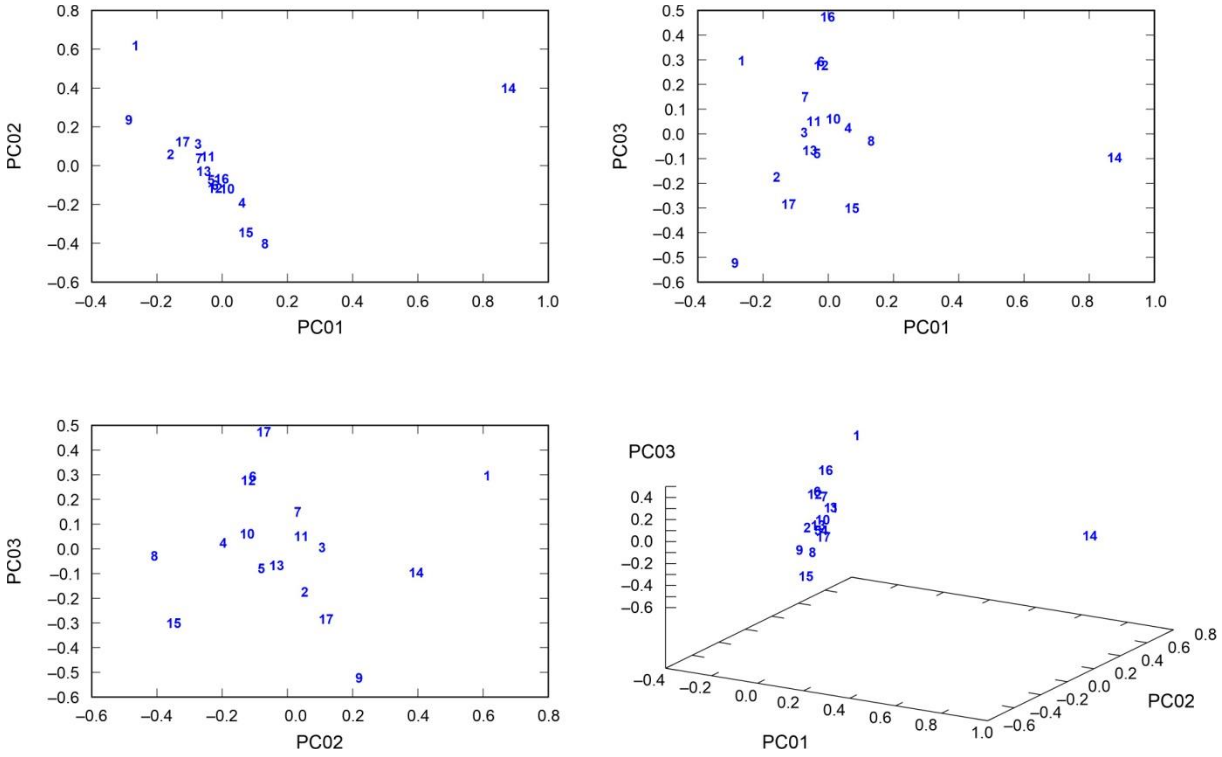

Fig. 5. Classification of the petroleum samples spanned in the space of the first three principal components for 3rd-way loadings calculated by TUCKER3 decomposition.

In the reduced space of 3rd-way loadings presented on Fig. 5, DOSY NMR spectra of the samples were represented as points (labeled as in Tables 1, 2 and S1). The distribution of all samples in this 3-dimensional space can be used for a classification of these samples based on their DOSY spectra. The sample **14** is highly distinguishable from the other samples, which is clear from the presented distribution. Moreover, from the variability among the samples one can see that the samples **6** and **12** are very similar. Investigation of 2-dimensional projections confirms that the same applies to the sample pairs (**10**,**11**) and (**5**,**13**).

3.4. Machine learning multivariate linear regression

3.4.1. Modeling stability with DOSY NMR spectra

To establish a connection between measured stability data and DOSY NMR spectra, extensive ML procedure was utilized. Stability data S_{asph}, S_{resin} and S_{total} from Table 1 were regressed to the first three principal components in the 3-dimensional reduced space of DOSY

NMR spectra. In this way, each crude oil sample was represented by the point in the reduced three-dimensional space and stability was modeled using these three predictors as independent variables. The total number of generated different models for each dependent variable was 1 717 869 184 (models were built as linear combination of original variables, as well as their higher-order polynomial terms, up to the 4th order). The 4th order was shown to be sufficient for building excellent regression models. Models with polynomial terms up to the 3rd order had predicted R^2 of 0.89 (for all three measured stability values). It was therefore justified to push it up to the order of 4 judging the model quality on the basis of adjusted R^2, predicted R^2 and LOO-CV mean-squared-error ensuring that there was no overfitting.

The best determined models among all different 1 717 869 184 tested models had predicted R^2 to be bigger than 0.98. For the measured stability of asphaltenes S_{asph}, predicted R^2 was 0.9808 whereas for S_{resin} predicted R^2 value was 0.9892 (Fig. 6). The quality of these models ensures that the crude oil stability in any similar crude oil sample can be predicted from the DOSY NMR spectra. This fact provides a broad range of possible applications using the DOSY NMR spectra for these or similar complex samples without the need for any additional chemical analyses. Properly predicting the stability of crude oils could *e.g.* directly reduce asphaltene remediation costs [29].

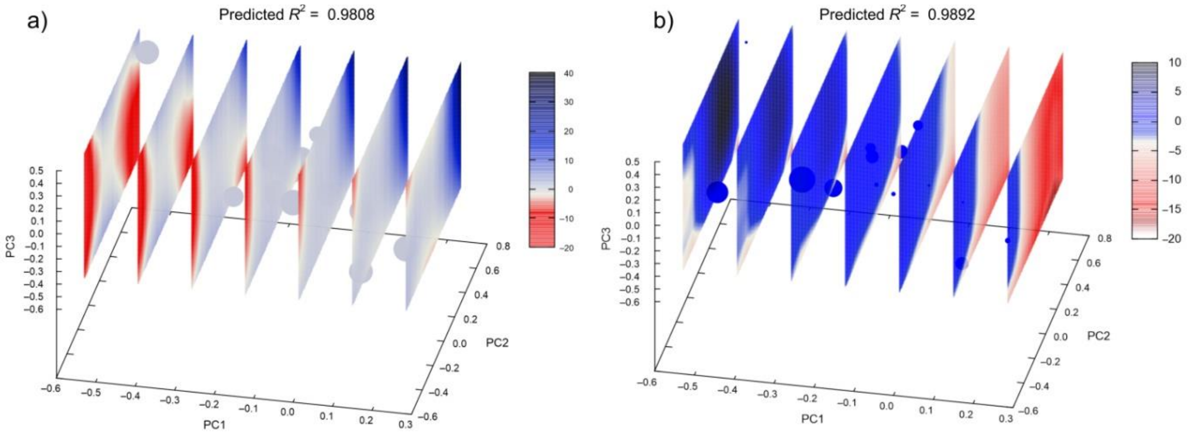

Fig. 6. The best multivariate regression model of the measured asphaltene and resin stability determined by machine learning: a) S_{asph}, and b) S_{resin} in dependence on the 1st, 2nd and 3rd

principal component of the DOSY NMR spectra of crude oil samples obtained by MWA. (Spheres represent points in 3D reduced space, and the planes are cuts of polynomial regression model, for easier interpretation 4th-dimension is represented redundantly with the color and with the size of the spheres.)

3.4.2. Modeling stability with other measured properties

Stability data was also regressed on 6 other measured properties: H_{ar}, $I_{1.3\ \mathrm{ppm}}$, w_{asph}, ρ_{API}, $d_{\mathrm{rel},0.9\ \mathrm{ppm}}$ and $d_{\mathrm{rel},1.3\ \mathrm{ppm}}$ (Tables 1 and 2). These measured properties were selected and their selection was further confirmed by investigation of linear correlation matrix with measured stability data where these properties showed some degree of linear correlation ($|R|>0.6$). In this case the number of possible models with linear combination of terms up to the polynomial order 2 was 134 217 728. Using parallelized ML code [23], it was possible to test all these models within one day and several excellent candidates were found. This search provided several regression models with values of predicted R^2 higher than 0.99 for all three measured stability parameters with the best ones having the following values of predicted R^2:

$R^2(S_{\mathrm{asph}})=0.9998$, $R^2(S_{\mathrm{resin}})=0.9997$ and $R^2(S_{\mathrm{total}})=0.9999$.

These are particularly good values for such complex mixtures, proving that this new model can accurately predict the crude oil stability and other important process parameters relevant for petroleum industry.

Best determined models:

$$S_{\mathrm{asph}} = 1.73\mathrm{E}{+}00 + 9.13\mathrm{E}{-}01\times \mathrm{H}_{\mathrm{ar}} + 1.99\mathrm{E}{-}02\times I_{1.3\ \mathrm{ppm}} - 5.84\mathrm{E}{-}02\times w_{\mathrm{asph}} - 3.39\mathrm{E}{-}02\times \rho_{\mathrm{API}} - 1.76\mathrm{E}{+}01\times d_{\mathrm{rel},0.9\ \mathrm{ppm}} - 3.99\mathrm{E}{-}02\times \mathrm{H}_{\mathrm{ar}}^2 - 1.21\mathrm{E}{-}02\times \mathrm{H}_{\mathrm{ar}} \times \rho_{\mathrm{API}} - 1.40\mathrm{E}{-}03\times I_{1.3\ \mathrm{ppm}}^2 - 1.11\mathrm{E}{-}02\times I_{1.3\ \mathrm{ppm}} \times w_{\mathrm{asph}} + 3.09\mathrm{E}{-}01\times I_{1.3\ \mathrm{ppm}} \times d_{\mathrm{rel},0.9\ \mathrm{ppm}} + 3.66\mathrm{E}{-}02\times w_{\mathrm{asph}}^2 + 1.06\mathrm{E}{+}00\times w_{\mathrm{asph}} \times d_{\mathrm{rel},0.9\ \mathrm{ppm}} + 2.02\mathrm{E}{-}01\times \rho_{\mathrm{API}} \times d_{\mathrm{rel},1.3\ \mathrm{ppm}} - 5.03\mathrm{E}{+}00\times d_{\mathrm{rel},0.9\ \mathrm{ppm}} \times d_{\mathrm{rel},1.3\ \mathrm{ppm}}$$

$$S_{\mathrm{resin}} = -2.77\mathrm{E}{+}01 + 1.52\mathrm{E}{+}01\times w_{\mathrm{asph}} + 4.25\mathrm{E}{+}00\times d_{\mathrm{rel},0.9\ \mathrm{ppm}} + 8.02\mathrm{E}{+}01\times d_{\mathrm{rel},1.3\ \mathrm{ppm}} + 7.83\mathrm{E}{-}02\times \mathrm{H}_{\mathrm{ar}}^2 - 7.88\mathrm{E}{-}01\times \mathrm{H}_{\mathrm{ar}} \times w_{\mathrm{asph}} + 3.75\mathrm{E}{-}03\times I_{1.3\ \mathrm{ppm}}^2 + 1.82\mathrm{E}{-}02\times I_{1.3\ \mathrm{ppm}} \times w_{\mathrm{asph}} - 6.05\mathrm{E}{-}03\times I_{1.3\ \mathrm{ppm}} \times \rho_{\mathrm{API}} - 5.48\mathrm{E}{-}01\times I_{1.3\ \mathrm{ppm}} \times d_{\mathrm{rel},1.3\ \mathrm{ppm}} - 1.40\mathrm{E}{-}01\times w_{\mathrm{asph}} \times \rho_{\mathrm{API}} - 1.49\mathrm{E}{+}01\times w_{\mathrm{asph}} \times d_{\mathrm{rel},0.9\ \mathrm{ppm}} + 1.85\mathrm{E}{-}02\times \rho_{\mathrm{API}}^2 - 1.27\mathrm{E}{+}00\times \rho_{\mathrm{API}} \times d_{\mathrm{rel},1.3\ \mathrm{ppm}}$$

$$S_{\mathrm{total}} = 1.78\mathrm{E}{+}02 - 9.21\mathrm{E}{+}00\times w_{\mathrm{asph}} - 7.15\mathrm{E}{+}00\times \rho_{\mathrm{API}} - 2.04\mathrm{E}{+}02\times d_{\mathrm{rel,1.3\ ppm}} + 2.52\mathrm{E}{-}02\times \mathrm{H_{ar}} \times I_{\mathrm{1.3\ ppm}} - 9.44\mathrm{E}{-}02\times \mathrm{H_{ar}} \times \rho_{\mathrm{API}} + 1.83\mathrm{E}{+}00\times \mathrm{H_{ar}} \times d_{\mathrm{rel,1.3\ ppm}} - 6.71\mathrm{E}{-}04\times I_{\mathrm{1.3\ ppm}}^{2} - 1.26\mathrm{E}{-}02\times w_{\mathrm{asph}} \times \rho_{\mathrm{API}} + 1.68\mathrm{E}{+}01\times w_{\mathrm{asph}} \times d_{\mathrm{rel,1.3\ ppm}} + 4.79\mathrm{E}{-}03\times \rho_{\mathrm{API}}^{2} + 1.32\mathrm{E}{+}01\times \rho_{\mathrm{API}} \times d_{\mathrm{rel,0.9\ ppm}} - 5.59\mathrm{E}{+}02\times d_{\mathrm{rel,0.9\ ppm}}^{2} + 3.39\mathrm{E}{+}02\times d_{\mathrm{rel,0.9\ ppm}} \times d_{\mathrm{rel,1.3\ ppm}}$$

4. Conclusion

With the application of *multi-way* analysis using the TUCKER3 decomposition model for a set of DOSY NMR spectra, principal components were determined for the model (5,5,5). This decomposition model described 99.89% of the total variance. A classification of crude oil samples using the reduced space of the first 3 principal components was performed. Similar samples were identified and reduced space was further utilized for the regression of measured stabilities. Extensive machine learning multivariate linear regression was proven useful for modeling crude oil stability based on DOSY NMR spectra and other measured properties. For both cases, very good models were established, up to the 4th polynomial order in the first case and up to the 2nd polynomial order in the second one. This approach can serve as an excellent tool for predicting stability of complex petroleum samples and can be applied for other similar systems.

Acknowledgment

This work was partially supported by the Croatian Science Foundation IP-2016-06-3775 (to T.H.).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Speight JG, El-Gendy NS. Chapter 1 - Petroleum composition and properties. In: Speight JG, El-Gendy NS, editors. Introduction to petroleum biotechnology, Oxford: Elsevier-Gulf Professional Publishing; 2018, p. 1–39. https://doi.org/10.1016/B978-0-12-805151-1.00001-1.

[2] Ashoori S, Sharifi M, Masoumi M, Salehi MM. The relationship between SARA fractions and crude oil stability. Egypt J Pet 2017;26:209–13. https://doi.org/10.1016/j.ejpe.2016.04.002.

[3] Durand E, Clemancey M, Lancelin J-M, Verstraete J, Espinat D, Quoineaud, A-A. Effect of chemical composition on asphaltenes aggregation. Energy Fuels 2010;24:1051–62. https://doi.org/10.1021/ef900599v.

[4] Gawrys KL, Spiecker PM, Kilpatrick PK. The role of asphaltene solubility and chemical composition on asphaltene aggregation. Petrol Sci Tech 2003;21:461–489. https://doi.org/10.1081/LFT-120018533.

[5] Evdokimov IN, Fesan AA, Losev AP. New answers to the optical interrogation of asphaltenes: monomers and primary aggregates from steady-state fluorescence studies. Energy Fuels 2016;30:4494–4503. https://doi.org/10.1021/acs.energyfuels.6b00027.

[6] Rogel E, León O, Contreras E, Carbognani L, Torres G, Espidel J, et al. Assessment of asphaltene stability in crude oils using conventional techniques**.** Energy Fuels 2003;17:1583–90. https://doi.org/10.1021/ef0301046.

[7] Wang J, Buckley JS. Asphaltene stability in crude oil and aromatic solvents - the influence of oil composition. Energy Fuels 2003;17:1445–51. https://doi.org/10.1021/ef030030y.

[8] Honse SO, Mansur CRE, Lucas EF. The influence of asphaltenes subfractions on the stability of crude oil model emulsions. J Braz Chem Soc 2012;23:2204–10. http://dx.doi.org/10.1590/S0103-50532013005000002.

[9] Schermer WEM, Melein PMJ, van den Berg FGA. Simple techniques for evaluation of crude oil compatibility. Pet Sci Technol 2004;22:1045–54. https://doi.org/10.1081/LFT-120038695.

[10] Vieira AP, Portela NA, Neto ÁC, Lacerda Jr. V, Romão W, Castro EVR, Filgueiras PR. Determination of physicochemical properties of petroleum using ^{1}H NMR spectroscopy combined with multivariate calibration. Fuel 2019;253:320–6. https://doi.org/10.1016/j.fuel.2019.05.028.

[11] Gao G, Cao J, Xu T, Zhang H, Zhang Y, Hu K. Nuclear magnetic resonance spectroscopy of crude oil as proxies for oil source and thermal maturity based on ^{1}H and ^{13}C spectra. Fuel 2020;271:117622. https://doi.org/10.1016/j.fuel.2020.117622.

[12] Rakhmatullin I, Efimov S, Tyurin V, Gafurov M, Al-Muntaser A, Varfolomeev M et al. Qualitative and quantitative analysis of heavy crude oil samples and their SARA fractions with ^{13}C nuclear magnetic resonance. Processes 2020;8:995. https://doi.org/10.3390/pr8080995**.**

[13] Parlov Vuković J, Novak P, Jednačak T. NMR spectroscopy as a tool for studying asphaltene composition. Croat Chem Acta 2019;92(3):323–29. https://doi.org/10.5562/cca3543.

[14] Durand E, Clemancey M, Lancelin J-M, Verstraete J, Espinat D, Quoineaud A-A. Aggregation states of asphaltenes: Evidence of two chemical behaviors by ^{1}H diffusion-ordered spectroscopy nuclear magnetic resonance. J Phys Chem C 2009;113:16266–16276. https://doi.org/10.1021/jp901954b.

[15] Lisitza NV, Freed DE, Sen PN, Song Y-Q. Study of asphaltene nanoaggregation by nuclear magnetic resonance (NMR). Energy Fuels 2009;23:1189–93. https://doi.org/10.1021/ef800631a.

[16] Parlov Vuković J, Novak P, Jednačak T, Kveštak M, Kovačević D, Smrečki V, et al. Magnetic field influence on asphaltene aggregation monitored by diffusion NMR spectroscopy: Is aggregation reversible at high magnetic fields? J Disper Sci Technol 2020;41:179–87. https://doi.org/10.1080/01932691.2018.1561302

[17] Parlov Vuković J, Hrenar T, Novak P, Friedrich M, Plavec J. New multiway model for identification of crude oil and asphaltene origin based on diffusion-ordered nuclear magnetic resonance spectroscopy. Energy Fuels 2017;31:8095–8101. https://doi.org/10.1021/acs.energyfuels.7b01358.

[18] Parlov Vuković J, Novak P, Plavec J, Friedrich M, Marinić Pajc Lj, Hrenar T. NMR and chemometric characterization of vacuum residues and vacuum gas oils from crude oils of different origin**.** Croat Chem Acta 2015;88:89–95. http://dx.doi.org/10.5562/cca2612.

[19] Standard Test Method for Determination of Asphaltenes (Heptane Insoluble) in Crude Petroleum and Petroleum Products, ASTM D 6560-17.

[20] Standard Test Method for Determination of Intrinsic Stability of Asphaltene-Containing Residues, Heavy Fuel Oils, and Crude Oils (*n*-Heptane Phase Separation; Optical Detection), ASTM D 7157-18.

[21] Hrenar T, Primožič I, Fijan D, Majerić Elenkov M. Conformational analysis of spiro-epoxides by principal component analysis of molecular dynamics trajectories. Phys Chem Chem Phys 2017;19:31706–13. https://doi.org/10.1039/C7CP05600A.

[22] Tucker L. Some mathematical notes on three-mode factor analysis. Psychometrika 1966;31:279–311. https://doi.org/10.1007/BF02289464.

[23] Hrenar T. *moonee,* Program for Manipulation and Analysis of Multi- and Univariate Data. Revision 0.6827, Zagreb, Croatia, 2021.

[24] Novak P, Kišić A, Hrenar T, Jednačak T, Miljanić S, Verbanec G. In-line reaction monitoring of entacapone synthesis by Raman spectroscopy and multivariate analysis. J Pharmaceut Biomed 2011;54:660–6. https://doi.org/10.1016/j.jpba.2010.10.012.

[25] Jović O, Smolić T, Primožič I, Hrenar T. Spectroscopic and chemometric analysis of binary and ternary edible oil mixtures: qualitative and quantitative study. Anal Chem 2016; 88:4516–4524. https://doi.org/10.1021/acs.analchem.6b00505.

[26] Jović O, Smolić T, Jurišić Z, Meić Z, Hrenar T. Chemometric analysis of Croatian extra virgin olive oils from central Dalmatia region. Croat Chem Acta 2013; 86:335–344. http://dx.doi.org/10.5562/cca2377.

[27] Parlov Vuković J, Telen S, Srića V, Novak P. The use of ^{13}C NMR spectroscopy and comprehensive two-dimensional gas chromatography, GC×GC, for identification of compounds involved in diesel fuel oxidative behavior. Croat Chem Acta 2011;84:537–41. http://dx.doi.org/10.5562/cca1874.

[28] IP 499/03 Standard Methods for Analysis and Testing of Petroleum and Related Products and British Standard 2000 Parts, Methods IP 361 to 501. The Institute of Petroleum London; 2003, 499.1.

[29] Kraiwattanawong K, Fogler HS, Gharfeh SG, Singh P, Thomason WH, Chavadej S. Thermodynamic solubility models to predict asphaltene instability in live crude oils. Energy Fuels 2007;21:1248–55. https://doi.org/10.1021/ef060386k.