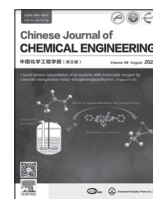




Contents lists available at ScienceDirect

Chinese Journal of Chemical Engineering

journal homepage: www.elsevier.com/locate/CJChE

Full Length Article

Molecular reconstruction of vacuum gas oils using a general molecule library through entropy maximization

Na Wang¹, Chong Peng^{2,*}, Zhenmin Cheng¹, Zhiming Zhou^{1,*}¹ School of Chemical Engineering, East China University of Science and Technology, Shanghai 200237, China² Dalian Research Institute of Petroleum and Petrochemicals, SINOPEC, Dalian 116000, China

ARTICLE INFO

Article history:

Received 31 March 2021

Received in revised form 16 June 2021

Accepted 21 June 2021

Available online 30 June 2021

Keywords:

Vacuum gas oil

Molecular reconstruction

Model

Algorithm

Optimization

ABSTRACT

Vacuum gas oil (VGO) is the most important feedstock for hydrocracking processes in refineries, but its molecular composition cannot be fully acquired by current analysis techniques owing to its complexity. In order to build an accurate and reliable molecular-level kinetic model for reactor design and process optimization, the molecular composition of VGO has to be reconstructed based on limited measurements. In this study, a modified stochastic reconstruction-entropy maximization (SR-REM) algorithm was applied to reconstruct VGOs, with generation of a general molecule library once and for all via the SR method at the first step and adjustment of the molecular abundance of various VGOs via the REM method at the second step. The universality of the molecule library and the effectiveness of the modified SR-REM method were validated by fifteen VGOs (three from the literature) from different geographic regions of the world and with different properties. The simulated properties (density, elemental composition, paraffin-naphthene-aromatics distribution, boiling point distribution, detailed composition of naphthenes and aromatics in terms of ring number as well as composition of S-heterocycles) are in good agreement with the measured counterparts, showing average absolute relative errors of below 10% for each property.

© 2021 The Chemical Industry and Engineering Society of China, and Chemical Industry Press Co., Ltd. All rights reserved.

1. Introduction

In recent decades the world has witnessed a growing demand for light fuel such as gasoline and diesel on the one hand, and a continuous increasing of heavier and inferior crude oil supply on the other hand. It is therefore significant for the refining industry to efficiently convert heavy petroleum fractions into light fractions by thermal and catalytic cracking processes such as fluid catalytic cracking, hydrocracking and delayed coking [1–3]. The conversion efficiency depends on the optimal operation of the reactor, which in turn relies on the reliable and accurate kinetic model. To circumvent the difficulty of quantifying individual species in the complex heavy fractions, a lumping technique is widely used in classical kinetic studies, in which hydrocarbon molecules are generally grouped into several lumps by physical properties such as boiling range [4–6]. Nevertheless, these lumps cannot yet satisfactorily characterize the feedstock and product compositions as each lump

contains a large number of molecules, and accordingly, the obtained kinetic model is restricted to specific feedstocks. Once some changes occur on the composition and properties of the feedstock, the rate parameters involved in the kinetic model cannot be feed-invariant. In this context, it is necessary to develop kinetic models at the molecular level [7].

The molecular-level kinetic models for chemical transformation of heavy petroleum fractions require the detailed molecular composition of feedstock as input. Unfortunately, current analysis tools cannot identify all species in the heavy fractions. In this case, the molecular reconstruction of feedstock is inevitably needed [8–10]. To date, many molecular-level reconstruction methods have been developed, e.g., stochastic reconstruction (SR) [11], structure-oriented lumping (SOL) [12], molecular type-homologous series (MTHS) [13], reconstruction by entropy maximization (REM) [14]. Among them, the SR method is particularly suitable for reconstructing heavy petroleum fractions such as asphaltene [15–17] and residue [18–20]. The MTHS and REM methods are suitable for light petroleum fractions as the former needs a detailed carbon number distribution [21] and the latter requires a complete molecule library [22], both of which can be accurately determined by modern analytic techniques for light

* Corresponding authors.

E-mail addresses: pengchong.fshy@sinopec.com (C. Peng), zmzhou@ecust.edu.cn (Z. Zhou).

fractions such as naphtha, but not for heavy fractions. Although the SOL method has been applied to heavy fractions [23,24], it suffers from too many variables, fewer constraints, and insufficient classification of types [10]. When applying the SR method, the Monte Carlo construction procedure is placed within the optimization loop together with a huge number of molecules to be constructed, thus creating a substantial computation burden to restrict its practical application.

Hudebine *et al.* [25] proposed a SR-REM two-step molecular reconstruction algorithm, which made use of the ability of SR to generate appropriate molecules at the first step and the flexibility of REM to adjust the molecular composition at the second step. The SR-REM method has advantages over the sole SR method in both computation time and accuracy, and has been applied to different fractions including light cycle oil (LCO) [25–27], vacuum gas oil (VGO) [28–30], and vacuum residue [31–33]. However, this method is still time-consuming if the oil samples that belong to the same fraction but with different properties are reconstructed individually (one-by-one) by the SR-REM method. Indeed, this strategy has been used in most previous studies. If a general molecule library specific to a given petroleum fraction is first built by the SR method, and then any sample from this fraction is reconstructed by the REM method based on this molecule library (a modified SR-REM method), the computational burden will be greatly reduced. To the best of our knowledge, to date only two studies have used this modified method to reconstruct LCO [25] and vacuum residue fractions [32]. de Oliveira *et al.* [32] reported that the computation time to reconstruct a VR fraction varied from several hours to one day by the traditional SR-REM method, which was greatly reduced to a few seconds using the modified SR-REM method.

The present work aims at reconstructing the molecular composition of VGOs via the modified SR-REM method. First, a general molecule library is built by the SR method using four VGO samples derived from different geographic regions of the world. In this step, two objective functions that contain different constrained properties are applied and compared: one with basic properties such as density, elemental composition and boiling point distribution, and one with not only the basic properties but also the detailed composition of naphthenic and aromatic rings as well as S-heterocycles. Next, the general molecule library is used to rebuild the above four VGOs by the REM method. The universality of the molecule library is examined by reconstructing eight other VGOs. Finally, this molecule library is applied to rebuild the VGOs reported in the literature, which further validates the effectiveness of the modified SR-REM method. Considering that the VGO fraction is the most widely used feedstock for hydrocrackers in refineries, this study lays a solid foundation for molecular-level kinetic modeling of VGO hydrocracking and subsequent reactor modeling and optimization.

2. Description of the Modified SR-REM Algorithm

The modified SR-REM two-step algorithm for molecular reconstruction of the VGO composition is shown in Fig. 1. At the first step, four VGO samples from different geographic regions with different properties are individually reconstructed by the SR method to generate a set of molecules (or molecule library), and then the generated four molecule libraries are united into a general molecule library. At the second step, the REM method is applied to adjust the molar fractions of molecules in the general molecule library, aiming at fine-tuning the properties of the generated mixture. Next, we will make a brief description of the SR-REM algorithm. More detailed information is available elsewhere [25,26,32,34].

2.1. Stochastic reconstruction (SR)

The SR method assumes that any molecule in the VGO feedstock is an assembly of different structural attributes according to certain probability distribution functions (PDFs). These structural attributes include type of molecules, number of side chains, number of benzene rings, number of naphthenic rings, type of heterocycles (thiophene, pyrrole and pyridine), *etc.*, as listed in Table 1. Although oxygen-containing species such as furans and phenols are also identified in VGO, they are less relevant for subsequent refinery modeling purpose and thus excluded from molecular reconstruction [34]. Each molecule is constructed by Monte Carlo sampling of the PDFs of the structural attributes based on a building diagram (Fig. 2), which is repeated N times to obtain a virtue mixture. N is set to 10000 in this work, as previous studies have shown that a sample size of 5000–10000 molecules can guarantee the prediction accuracy [32,35,36]. Worthy of mention is the use of structural attributes to differentiate aromatic cores with different configurations, *i.e.*, phenanthrene and anthracene (for triaromatics), pyrene and chrysene (for tetraaromatics), and perylene and benzopyrene (for pentaaromatics). The difference in the structural configuration of aromatic cores with multiple rings has recently been taken into account for heavy petroleum fractions such as asphaltene [17] and vacuum residue [37], but very few studies have been reported for VGO. Recently, Deniz *et al.* [38] reported that consideration of different configurations for aromatic cores was favorable for better fit of VGO through increased model flexibility.

The range of values and the number of parameters of PDFs of structural attributes are presented in Table 1, which are determined by referring to previous studies [34]. In addition, several chemical rules are used: (1) only single-core structures are allowed for naphthenes and aromatics; (2) only cyclohexane rings are considered for naphthenes; (3) aromatic species contains up to five benzene rings and two cyclohexane rings; (4) only one heterocycle such as thiophene, pyrrole or pyridine is allowed; (5) only one non-methyl branch is considered in the case of multiple alkyl branches; and (6) multiple branches are located as far as possible from each other on the parent structure. In total, there are 17 structural attributes with 27 associated parameters that need to be estimated from the experimental properties of VGO. It should be noted that the gamma distribution is normally associated with two parameters, *i.e.*, a shape parameter and a scale parameter. However, in the case of paraffin chain length and alkyl branch carbons, the scale parameter can be derived from the shape parameter and thus, the number of parameters is reduced to 1 [32,34].

During the molecular reconstruction process, the properties of each molecule are calculated either directly by its structure (chemical formula, molecular weight, *etc.*) or by group contribution methods (liquid density at 20 °C and normal boiling point [32]). The average density of the mixture is calculated using the linear mixing rule. After that, the calculated bulk properties of VGO are compared with the analytical data through an objective function. The parameters of the PDFs are optimized using a simulated annealing algorithm to minimize the objective function. The objective function is defined as the average of the relative residuals between the experimental and simulated properties:

$$F = \frac{1}{N_p} \sum_{i=1}^{N_p} \frac{|X_{i,\text{exp}} - X_{i,\text{sim}}|}{X_{i,\text{exp}}} \quad (1)$$

where $X_{i,\text{exp}}$ and $X_{i,\text{sim}}$ are the experimental and simulated values of the i -th constrained property, respectively, and N_p is the number of constrained properties in the objective function. Note that the relative residuals are not squared, with the aim of avoiding the amplification of the background noise available in the objective function

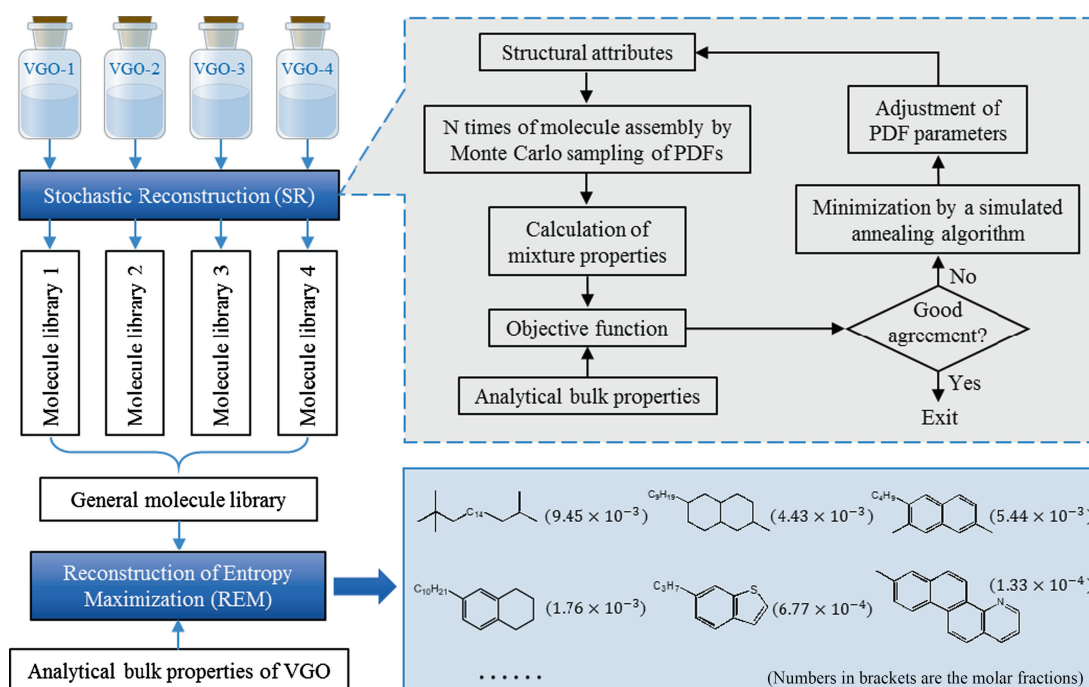


Fig. 1. Flow diagram of the molecular reconstruction of VGOs by a modified SR-REM algorithm.

Table 1

Mathematical description of structural attributes

Structural attributes	Values	PDF	Number of parameters
Type of molecules ^①	1, 2, 3	Histogram	2
Paraffins			
Paraffin chain length	15–40	Gamma	1
Acceptance of alkyl branches	0, 1	Histogram	1
Alkyl branch carbons	1–4	Exponential	1
Number of alkyl branches	1–3	Histogram	2
Naphthenes			
Number of naphthenic rings	1–5	Exponential	1
Alkyl branch carbons	≥0	Gamma	1
Number of alkyl branches	1–4	Histogram	3
Aromatics			
Number of benzene rings	1–5	Exponential	1
Benzene ring configuration ^②	1, 2	Histogram	3
Number of naphthenic rings	0–2	Histogram	2
Acceptance of heterocycles	0, 1	Histogram	1
Type of heterocycles ^③	1, 2, 3	Histogram	2
Alkyl branch carbons	≥0	Gamma	1
Number of alkyl branches	1–4	Histogram	3
Heteroatom substitution on alkyl chains and branches			
Acceptance of heteroatom substitution	0, 1	Histogram	1
Type of heteroatom substitution ^④	1, 2	Histogram	1

^① Type of molecule: 1-paraffins; 2-naphthenes; 3-aromatics.

^② Aromatic ring configuration: 1-phenanthrene or pyrene or perylene; 2-anthracene or chrysene or benzopyrene.

^③ Type of heterocycles: 1-thiophene; 2-pyridine; 3-pyrrole.

^④ Type of heteroatom substitution: 1-SH; 2-NH₂.

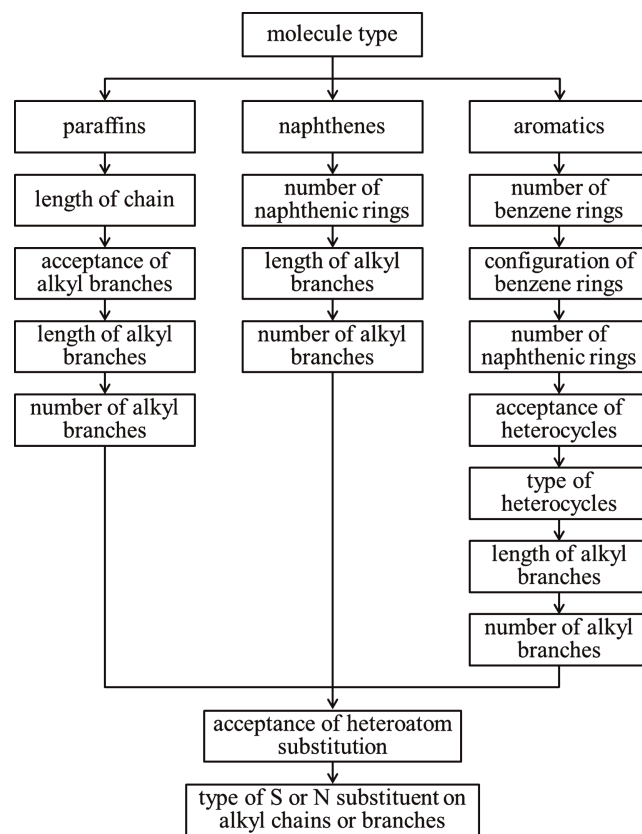


Fig. 2. Molecule building diagram for VGOs.

derived from the Monte Carlo-based sampling method [32]. In this study, two objective functions are considered in order to evaluate the effect of constrained properties on the simulation result. The constrained properties in the first objective function are some basic properties, including density, elemental analysis (C, H, S and N), paraffin-naphthene-aromatics (PNA) distribution and boiling point distribution. The second objective function is an extension of the

first objective function, which contains not only the basic properties, but also the detailed composition of naphthenic and aromatic rings (i.e., 1–5 rings) as well as S-heterocycles. A Fortran code is implemented on an Intel® Core(TM) i5-6500 based computer, and the CPU time required for each stochastic reconstruction is about 12 h.

2.2. Reconstruction by entropy maximum (REM)

The SR method generates a set of representative molecules, whose molar fractions are determined by the REM method that is based on the principle of maximum Shannon entropy [39]. The information entropy (H) to be maximized is expressed as:

$$H = - \sum_{j=1}^{N_m} y_j \ln(y_j) + \mu \left(1 - \sum_{j=1}^{N_m} y_j \right) + \sum_{i=1}^{N_p} \lambda_i \left(X_{i,\text{exp}} - \sum_{j=1}^{N_m} y_j X_{j,i} \right) \quad (2)$$

where y_j is the molar fraction of the j -th molecule, N_m is the number of molecules, μ is the Lagrange multiplier associated with the mass balance constraint, λ_i is the Lagrange multiplier associated with the i -th property, and $X_{j,i}$ is the coefficient of the j -th molecule for the i -th property. Further information on the REM method and its calculation procedure can be found elsewhere [40–43]. Compared to the SR method, the CPU time for REM is greatly reduced, being only around 2 min.

3. Characterization of Vacuum Gas Oil

Twelve VGO samples with different properties (Table 2) were characterized. The liquid density at 20 °C was measured by ASTM

Table 2

Physicochemical properties of twelve VGO samples

Samples	Description	Density at 20 °C/ g·cm ⁻³	S content/% (mass)	N content/ μg·g ⁻¹	Boiling point interval/ °C	Viscosity at 100 °C/ mm ² ·s ⁻¹	BMCI
VGO-1	Saudi Arabia medium VGO	0.9205	2.45	842	325–520	7.628	49.8
VGO-2	China VGO	0.9301	2.23	2978	232–527	4.135	57.6
VGO-3	Iran VGO	0.9095	1.59	1581	315–547	7.062	43.6
VGO-4	Venezuela VGO	0.9431	2.54	2000	349–504	8.958	59.9
VGO-5	Mixture of VGOs	0.9101	2.58	752	308–530	5.732	45.4
VGO-6	China coker gas oil	0.9473	3.00	3453	275–548	6.500	62.4
VGO-7	Iran VGO	0.9121	1.28	1325	335–532	6.333	45.0
VGO-8	Mixture of VGO-1 and VGO-2	0.9232	2.38	1269	296–524	6.378	51.2
VGO-9	Iran VGO	0.9065	1.71	1195	292–515	4.755	45.5
VGO-10	Iran VGO	0.9168	1.75	1681	336–542	6.901	43.1
VGO-11	Saudi Arabia light VGO	0.9036	2.98	692	339–451	3.822	45.6
VGO-12	Thailand VGO	0.9210	2.47	973	332–571	6.870	49.0

Table 3

Experimental and simulated properties of four VGO samples by the SR method

Properties	VGO-1			VGO-2			VGO-3			VGO-4		
	Exp	Sim-1 ^①	Sim-2 ^②	Exp	Sim-1 ^①	Sim-2 ^②	Exp	Sim-1 ^①	Sim-2 ^②	Exp	Sim-1 ^①	Sim-2 ^②
Density/g·cm ⁻³	0.9205	0.9136	0.9110	0.9301	0.9165	0.9156	0.9095	0.9088	0.8906	0.9431	0.9464	0.9318
Elemental analysis/(% (mass))												
Carbon	85.46	85.47	85.04	86.63	85.70	85.10	85.91	85.57	86.13	85.57	85.40	84.56
Hydrogen	12.01	12.13	12.11	10.84	11.8	11.90	12.34	12.61	12.95	11.69	11.93	12.12
Sulphur	2.45	2.34	2.73	2.23	2.20	2.70	1.59	1.63	0.73	2.54	2.50	3.21
Nitrogen/ppmw	842	564	1248	2978	2734	2640	1581	1925	1915	2000	1742	1147
PNA distribution/(% (mass))												
Paraffins	21.4	25.3	28.2	21.2	21.7	22.1	24.3	23.8	28.5	11.7	15.4	12.9
Naphthenes	25.6	23.7	22.4	27.3	27.2	23.9	34.8	36.9	33.9	47.3	46.1	46.0
Mononaphthenes	7.4	3.3	6.8	9.4	3.7	9.2	11.3	5.1	12.5	10.1	19.1	12.2
Dinaphthenes	7.1	3.9	5.6	6.8	4.8	6.1	9.4	6.4	8.6	11.8	11.7	10.5
Trinaphthenes	4.7	4.7	4.7	4.5	5.5	3.9	6.6	7.6	6.4	10.2	7.9	9.1
Tetranaphthenes	4.4	5.9	3.1	4.0	6.5	3.1	5.7	8.7	3.8	10.7	4.6	8.1
Pentanaphthenes	2.0	5.9	2.2	2.6	6.7	1.6	1.8	9.1	2.6	4.5	2.8	6.1
Aromatics	53.0	51.0	49.4	51.5	51.1	54.0	40.9	39.3	37.6	41.0	38.5	41.1
Monoaromatics	18.2	37.5	17.9	17.0	44.8	21.1	18.4	24.2	16.1	14.2	7.0	15.4
Diaromatics	10.2	8.3	9.4	11.7	4.3	10.9	10.4	6.2	8.8	11.3	6.7	9.2
Triaromatics	6.7	1.6	4.5	7.0	0.6	7.1	4.5	1.4	4.3	5.6	7.5	6.1
Tetraaromatics	4.4	0.3	2.3	4.8	0.1	3.7	1.7	0.3	2.0	3.7	6.2	3.3
Pentaaromatics	1.8	0.1	1.2	1.9	0.0	1.6	0.6	0.1	1.0	0.5	5.1	1.7
S-heterocycles	11.7	3.2	14.1	9.1	1.3	9.6	5.3	7.1	5.4	5.7	6.0	5.4
Boiling point distribution/°C												
IBP	325	319	306	232	230	228	315	312	308	349	333	325
10% BP	376	363	361	334	328	321	382	374	367	392	386	377
30% BP	411	399	385	377	370	362	417	411	403	413	403	387
50% BP	432	434	437	400	406	409	442	438	446	434	434	438
70% BP	458	460	464	426	432	440	469	466	472	456	465	471
90% BP	494	510	513	476	483	495	509	514	518	490	502	518
EBP	520	527	536	527	534	538	547	553	557	504	520	527

^① Properties in the 1st objection function include density, elemental composition, PNA distribution and boiling point data.

^② Properties in the 2nd objection function include density, elemental composition, PNA distribution, boiling point data, naphthenic and aromatic rings (i.e., 1–5 rings), as well as S-heterocycles.

D4052. The elemental composition (C, H, S and N) was determined by ASTM D5291/D1552. The boiling point distribution was acquired by simulated distillation according to ASTM D6352. The hydrocarbon-type distribution and S-heterocycles were analyzed by two-dimensional gas chromatography with a flame ionization detector and a sulfur chemiluminescence detector, respectively.

4. Results and Discussion

4.1. Construction of a general molecule library for VGOs

Four samples used for constructing a general molecule library for VGOs originate from Saudi Arabia (VGO-1), China (VGO-2), Iran (VGO-3) and Venezuela (VGO-4), which represent the main VGO fractions in the world. It can be seen from Table 2 that the four samples have different characteristics, e.g., the nitrogen content

of VGO-1 is the lowest among the four samples, the viscosity of VGO-2 is the lowest, and the density and sulfur content of VGO-3 are the lowest.

When the first objective function is applied, the analysis data (columns labeled “exp”, Table 3) of density, elemental composition, PNA distribution, and boiling point distribution are well simulated by the synthetic mixture (columns labeled “sim-1”, Table 3). Unfortunately, however, as far as the detailed composition of naphthenic and aromatic rings are concerned, remarkable deviations are observed between experimental and simulated data. Taking VGO-1 as an example, although the experimental and simulated contents of aromatics (53%(mass) vs. 51%(mass)) are close to each other, the simulated distribution of aromatics in terms of ring number (37.5%, 8.3%, 1.6%, 0.3% and 0.1%(mass) for mono-, di-, tri-, tetra- and penta-aromatics, respectively) is different from the experimental distribution (18.2%(mass), 10.2%(mass), 6.7% (mass), 4.4%(mass) and 1.8%(mass), respectively). A similar result is

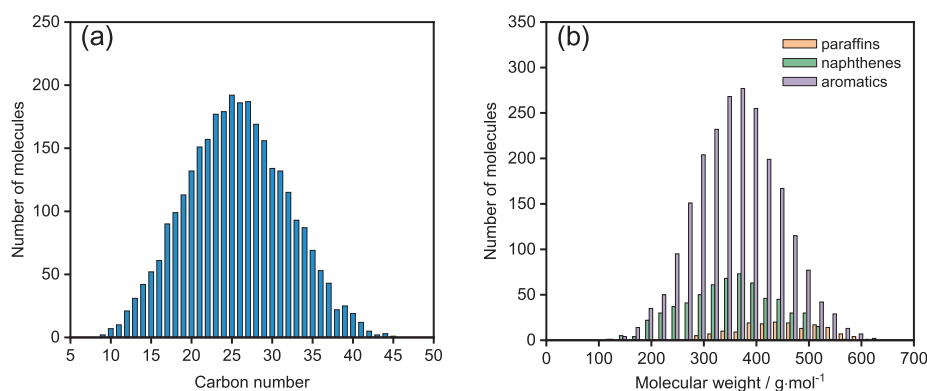


Fig. 3. (a) Carbon number distribution and (b) molecular weight distribution of the general molecule library.

Table 4

Experimental and simulated properties of four VGO samples by the REM method using a general molecule library

Properties	VGO-1		VGO-2		VGO-3		VGO-4	
	Exp	Sim	Exp	Sim	Exp	Sim	Exp	Sim
Density/g·cm ⁻³	0.9205	0.9275	0.9301	0.9303	0.9095	0.8961	0.9431	0.9380
Elemental analysis/(% (mass))								
Carbon	85.46	85.65	86.63	86.37	85.91	85.93	85.57	85.74
Hydrogen	12.01	11.64	10.84	11.58	12.34	12.35	11.69	11.85
Sulphur	2.45	2.61	2.23	1.75	1.59	1.56	2.54	2.21
Nitrogen/ppmw	842	907	2978	2893	1581	1648	2000	1933
PNA distribution/(% (mass))								
Paraffins	21.4	19.6	21.2	19.9	24.3	26.2	11.7	10.8
Naphthenes	25.6	26.6	27.3	26.8	34.8	31.6	47.3	48.7
Mononaphthenes	7.4	8.5	9.4	8.8	11.3	10.8	10.1	10.5
Dinaphthenes	7.1	7.4	6.8	7.5	9.4	8.1	11.8	11.5
Trinaphthenes	4.7	5.1	4.5	4.3	6.6	5.8	10.2	10.9
Tetranaphthenes	4.4	3.9	4.0	3.5	5.7	5.1	10.7	10.9
Pentanaphthenes	2.0	1.7	2.6	2.7	1.8	1.8	4.5	4.9
Aromatics	53.0	53.8	51.5	53.3	40.9	42.2	41.0	40.5
Monoaromatics	18.2	20.6	17.0	16.7	18.4	20.5	14.2	14.5
Diaromatics	10.2	10.6	11.7	12.9	10.4	10.7	11.3	10.1
Triaromatics	6.7	6.2	7.0	7.9	4.5	4.1	5.6	6.2
Tetraaromatics	4.4	3.9	4.8	5.3	1.7	1.4	3.7	3.2
Pentaaromatics	1.8	1.9	1.9	2.1	0.6	0.5	0.5	0.5
S-heterocycles	11.7	10.6	9.1	8.4	5.3	5.0	5.7	6.0
Boiling point distribution/°C								
IBP	325	326	232	232	315	315	349	349
10% BP	376	367	334	331	382	374	392	379
30% BP	411	399	377	382	417	406	413	412
50% BP	432	434	400	409	442	434	434	437
70% BP	458	464	426	429	469	472	456	463
90% BP	494	507	476	480	509	511	490	497
EBP	520	520	527	527	547	547	504	504

obtained for naphthenes. In addition, the simulated content of S-heterocycles (3.2% (mass)) is quite different from the experimental value (11.7% (mass)). Therefore, the first objective function does not yield a molecule library that can satisfactorily represent the VGO fractions.

Next, the second objective function is applied to construct a molecule library for VGO. The simulated properties are also listed in Table 3 (columns labeled “sim-2”). It seems that the second objective function gives a slightly poorer simulation of the basic properties of the four VGOs as compared to the first objective function, which is mainly due to more constrained properties in the second objective function for global optimization by simulated annealing. Nevertheless, the simulation results about the distribution of aromatics and naphthenes as well as the content of S-heterocycles are greatly improved when the second objective function is used. Taking the content of S-heterocycles as an example, the experimental data for four samples from VGO-1 to VGO-4 are 11.7% (mass), 9.1% (mass), 5.3% (mass) and 5.7% (mass), respectively, which deviate from the simulated data obtained with the first objective function (3.2% (mass), 1.3% (mass), 7.1% (mass) and

6.0% (mass), respectively) but close to the data obtained with the second objective function (14.1% (mass), 9.6% (mass), 5.4% (mass) and 5.4% (mass), respectively). It is therefore concluded from the comparison of the results of the two objective functions that, the second objective function that contains both the basic properties and the distribution of naphthenic and aromatic rings is appropriate to construct a molecule library for VGO fractions.

By applying the second objective function, four molecule libraries are constructed individually for the four VGOs. Note that although the number of Monte Carlo sampling is 10000 for each VGO, the number of molecules (each molecule is unique) in the final molecule library is less than 10000 due to repetition of some molecules during sampling. The number of molecules in each library (from VGO-1 to VGO-4) is 1571, 1644, 1507 and 1670, respectively. The four libraries are then united together by rejecting identical molecules, which finally yields a general molecule library for VGOs. The general library contains 3023 molecules with a gamma distribution within the carbon number range of 9–45 (Fig. 3(a)), whose molecular weight distribution by PNA family is shown in Fig. 3(b) and Table S1 (Supplementary Material). The

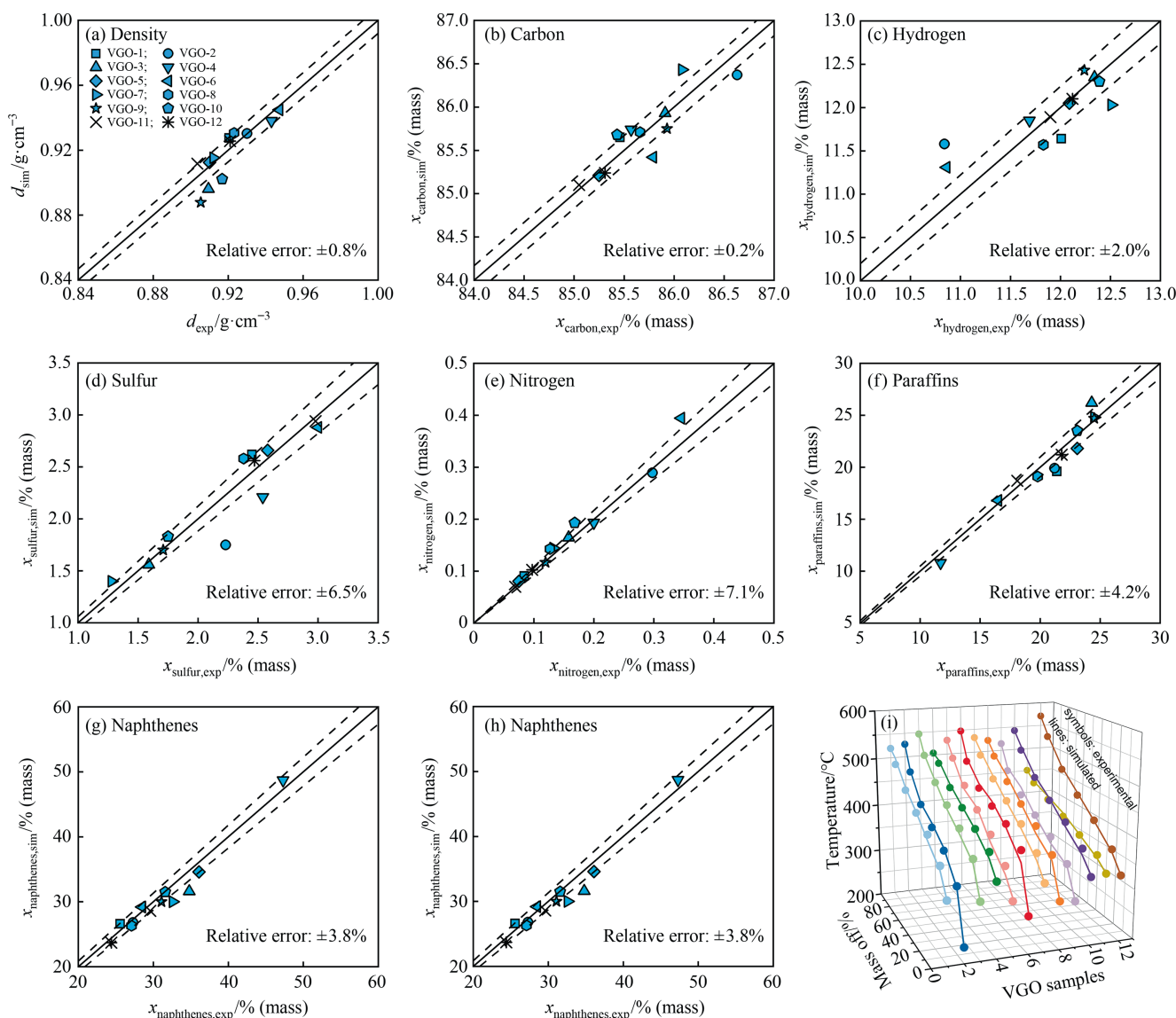


Fig. 4. (a–h) Parity plots for experimental and simulated properties of 12 VGOs and (i) experimental vs. simulated boiling point distribution curves (symbols: experimental; lines: simulated).

molecular weight range of the general molecule library is 105–630 g·mol⁻¹, which is close to those reported for VGOs in the literature, e.g., 160–580 g·mol⁻¹ [34] and 150–700 g·mol⁻¹ [44]. The aromatic species account for a major proportion of the molecule library (74.0%), followed by naphthenes (20.6%) and paraffins (5.4%). Considering that the aromatic hydrocarbons are more complex in structure than naphthenic and paraffinic hydrocarbons, more structural attributes are required to describe the aromatic hydrocarbons (Table 1), which results in the largest contribution of aromatics to the general molecule library. The most abundant species in the aromatic and naphthenic families have molecular weights of 355–380 g·mol⁻¹, which shift to 380–480 g·mol⁻¹ for paraffins. Next, the general molecule library consisting of an equimolar set of 3023 molecules is used to reconstruct VGOs by the REM method. The properties used in REM are the same as those in the second objective function of the SR method.

4.2. Reconstruction of VGOs by REM

Prior to REM of any VGO sample, the above general molecule library is reduced by eliminating the molecules with a normal boiling point lower than the initial boiling point or higher than the end boiling point of the VGO [22]. Table 4 compares the experimental and simulated properties of four samples from VGO-1 to VGO-4. Comparing with the simulated results obtained from the SR method (Table 3), it is apparent that, after the REM step using the general molecule library, a remarkable improvement is achieved in the properties of the four VGOs, i.e., density, elemental composition, boiling point distribution, and PNA distribution

including detailed composition of naphthenic and aromatic rings as well as S-heterocycles. In addition, eight other VGOs (from VGO-5 to VGO-12) with different properties can be well reconstructed by the REM method using the above generated molecule library, as summarized in Table S2. For all the twelve samples from VGO-1 to VGO-12, the average absolute relative error of all properties for each VGO $\left(\frac{1}{26} \sum_{i=1}^{26} \frac{|X_{j,i,exp} - X_{j,i,sim}|}{X_{j,i,exp}} \times 100\%\right)$ is calculated to be 5.7%, 5.4%, 5.4%, 4.0%, 4.4%, 4.6%, 5.4%, 4.2%, 2.9%, 3.3%, 7.7%, and 4.5%, respectively. The errors for the first four VGOs that are used for generating the molecule library are generally comparable to those for the other eight VGOs, demonstrating the reliability of the general molecule library.

Fig. 4(a–h) shows the parity plots of simulated versus experimental values for basic properties of the twelve VGOs. The average absolute relative errors for most basic properties $\left(\frac{1}{12} \sum_{j=1}^{12} \frac{|X_{j,i,exp} - X_{j,i,sim}|}{X_{j,i,exp}} \times 100\%\right)$ are smaller than 5% except for sulfur and nitrogen contents (6.5% and 7.1%, respectively). Moreover, the simulated boiling points are very close to the experimental values (Fig. 4(i)). Even regarding the detailed distributions of naphthenes and aromatics in terms of ring number as well as the content of S-heterocycles, the average absolute relative errors are smaller than 10% (Fig. S1). Therefore, the approach to reconstruct the molecular composition of VGOs from a general molecule library with the REM method is justified.

Finally, this molecule library is applied to reconstruct VGO samples in the literature. As listed in Table 5, sample A (VGO-A, from Canadian oil sands heavy crude) [34], sample B (VGO-B, from

Table 5
Experimental and simulated properties of literature reported VGOs

Properties	VGO-A ^①		VGO-B ^②		VGO-C ^③	
	Exp	Sim	Exp	Sim	Exp	Sim
Density/g·cm ⁻³	0.9760	0.9686	0.9115	0.9051	0.9127	0.9145
Elemental analysis/% (mass)						
Carbon	84.88	85.87	87.48	87.30	85.68	85.68
Hydrogen	11.23	10.89	12.60	12.60	12.38	12.14
Sulphur	3.50	3.11	0.09	0.09	2.24	2.12
Nitrogen/ppmw	1200	1289	129	100	456	590
PNA distribution/% (mass)						
Paraffins	0.0	0.0	3.2	2.9	22.1	20.0
Naphthenes	35.4	36.0	61.3	61.5	42.0	41.0
Aromatics	64.6	64.0	35.4	35.6	35.9	39.0
Monoaromatics	5.3	10.9	18.6	18.5	6.1	9.9
Diaromatics	28.1	26.6	9.1	9.1	12.5	11.9
Triaromatics	22.7	20.0	4.5	4.5	10.4	9.3
Tetraaromatics	8.2	6.3	1.9	2.0	6.2	7.1
Pentaaromatics	0.3	0.2	1.3	1.5	0.7	0.8
¹³ C spectrum/% (mass)						
Aliphatic CH ₃	18.4	10.4	20.2	17.2	14.5	11.0
Aliphatic CH ₂	37.2	52.3	45.4	56.2	52.5	64.0
Aliphatic CH	16.9	13.5	18.6	12.8	16.2	14.6
Aromatic CH	11.1	11.1	6.8	6.5	9.7	5.0
Substituted aromatic C	11.0	7.4	7.8	6.0	6.3	3.4
Bridge aromatic C	5.4	5.3	1.2	1.3	0.8	2.0
Boiling point distribution ^③ /°C						
IBP	304	304	168	178	209	209
10% BP	369	363	293	317	365	353
30% BP	410	426	352	347	418	406
50% BP	438	450	397	390	450	453
70% BP	465	478	441	453	479	472
90% BP	504	502	494	521	511	519
EBP	578	578	568	568	551	551

^① Experimental data are obtained from Ref. [34].

^② Experimental data are obtained from Ref. [29].

^③ Experimental boiling points are derived from the reported boiling point distribution curves [29,34].

Canadian oil sands bitumen) [29] and sample C (VGO-C, from Middle Eastern light crude) [34] have different properties: VGO-A and VGO-B are characteristic of no or only a very small amount of paraffins with the former rich in aromatics and the latter rich in naphthenes, while VGO-C has comparable amounts of paraffins, naphthenes and aromatics. Unfortunately, the detailed distribution of naphthenes in terms of ring number is not available in the literature for the three VGOS, but instead the ^{13}C NMR spectrum is provided. In this case, the constrained properties associated with the ring number distribution of naphthenes have to be removed from Eq. (2) when applying the REM method with the general molecule library. The simulated properties are presented in Table 5, which are close to the experimental data. In addition, the ^{13}C NMR spectrum data that are not used as constrained properties during the molecular reconstruction are generally predicted. For some ^{13}C NMR data, e.g., the aliphatic CH_2 group, there exists a relatively large difference between experimental and predicted values, which is probably caused by the lack of detailed distribution of naphthenes in the objective function. In summary, the molecular reconstruction of VGOS is fulfilled successfully by the REM method using the general molecule library. It should be noted that the molecule library generated in this study is restricted to VGOS with no or a very small amount of resins; otherwise, 6+-ring naphthenes and aromatics should be taken into account at the SR step.

5. Conclusions

The molecular composition of VGO was constructed by a modified SR-REM algorithm, where a general molecule library was first built by the SR method and then used to adjust the molecular composition of any VGO by the REM method. It was noteworthy that the general molecule library was built only once and for all, such that the computational burden was greatly reduced as compared to the conventional SR-REM algorithm. During construction of the molecule library from four VGOS differing in the geographic regions and in the physicochemical properties, it was found that the constrained properties in the objection function affected the simulation results. The commonly used properties such as density, elemental composition, PNA distribution and boiling point distribution were not adequate to generate an appropriate molecule library for VGO. Detailed compositions of naphthenes and aromatics in terms of ring number as well as S-heterocycles were also required. The general molecule library consisted of 3023 unique molecules with a gamma distribution within the carbon number range of 9–45, in which aromatics accounted for the largest proportion. With this general molecule library, eleven other VGOS including three from the literature were rebuilt by the REM method. A good agreement was achieved between the simulated and experimental properties, demonstrating the universality of the molecule library and the effectiveness of the modified SR-REM method for the molecular reconstruction of VGO.

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.

Acknowledgements

We thank Dr. Jinwen Chen for helpful discussion. This work was supported by the National Natural Science Foundation of China (21978093).

Supplementary Material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cjche.2021.06.007>.

References

- [1] C.I.C. Pinheiro, J.L. Fernandes, L. Domingues, A.J.S. Chambel, I. Graça, N.M.C. Oliveira, H.S. Cerqueira, F.R. Ribeiro, Fluid catalytic cracking (FCC) process modeling, simulation, and control, *Ind. Eng. Chem. Res.* 51 (2012) 1–29.
- [2] R. Sahu, B.J. Song, J.S. Im, Y.P. Jeon, C.W. Lee, A review of recent advances in catalytic hydrocracking of heavy residues, *J. Ind. Eng. Chem.* 27 (2015) 12–24.
- [3] R. Prajapati, K. Kohli, S.K. Maity, Slurry phase hydrocracking of heavy oil and residue to produce lighter fuels: An experimental review, *Fuel* 288 (2021) 119686.
- [4] C.S. Laxminarasimhan, R.P. Verma, P.A. Ramachandran, Continuous lumping model for simulation of hydrocracking, *AIChE J.* 42 (1996) 2645–2653.
- [5] T.C. Ho, Kinetic modeling of large-scale reaction systems, *Catal. Rev.: Sci. Eng.* 50 (2008) 287–378.
- [6] J. Ancheyta, Deactivation of Heavy Oil Hydroprocessing Catalysts, John Wiley & Sons, Hoboken, New Jersey, 2016.
- [7] M.T. Klein, G. Hou, R. Bertolacini, L.J. Broadbelt, A. Kumar, Molecular Modeling in Heavy Hydrocarbon Conversions, CRC Press, Boca Raton, Florida, 2005.
- [8] R. Van de Vijver, B.R. Devocht, K.M. Van Geem, J.W. Thybaut, G.B. Marin, Challenges and opportunities for molecule-based management of chemical processes, *Curr. Opin. Chem. Eng.* 13 (2016) 142–149.
- [9] L.P. de Oliveira, D. Hudebine, D. Guillaume, J.J. Verstraete, A review of kinetic modeling methodologies for complex processes, *Oil Gas Sci. Technol.* 71 (2016) 45.
- [10] Y. Ren, Z. Liao, J. Sun, B. Jiang, J. Wang, Y. Yang, Q. Wu, Molecular reconstruction: recent progress toward composition modeling of petroleum fractions, *Chem. Eng. J.* 357 (2019) 761–775.
- [11] M. Neurock, C. Libanati, A. Nigam, M.T. Klein, Monte Carlo simulation of complex reaction systems: molecular structure and reactivity in modelling heavy oils, *Chem. Eng. Sci.* 45 (1990) 2083–2088.
- [12] R.J. Quann, S.B. Jaffe, Structure-oriented lumping: describing the chemistry of complex hydrocarbon mixtures, *Ind. Eng. Chem. Res.* 31 (1992) 2483–2497.
- [13] B. Peng, Molecular modelling of petroleum processes Ph. D. Thesis, University of Manchester, Manchester, 1999.
- [14] D. Hudebine, C. Vera, F. Wahl, J. Verstraete, Molecular representation of hydrocarbon mixtures from overall petroleum analyses, AIChE Spring Meeting, New Orleans, 2002.
- [15] M. Neurock, A. Nigam, D. Trauth, M.T. Klein, Molecular representation of complex hydrocarbon feedstocks through efficient characterization and stochastic algorithms, *Chem. Eng. Sci.* 49 (1994) 4153–4177.
- [16] J.M. Sheremata, M.R. Gray, H.D. Dettman, W.C. McCaffrey, Quantitative molecular representation and sequential optimization of athabasca asphaltenes, *Energy Fuels* 18 (2004) 1377–1384.
- [17] C.U. Deniz, M. Yasar, M.T. Klein, A new extended structural parameter set for stochastic molecular reconstruction: application to asphaltenes, *Energy Fuels* 31 (2017) 7919–7931.
- [18] D.M. Trauth, S.M. Stark, T.F. Petti, M. Neurock, M.T. Klein, Representation of the molecular structure of petroleum resid through characterization and Monte Carlo modeling, *Energy Fuels* 8 (1994) 576–580.
- [19] D.M. Campbell, M.T. Klein, Construction of a molecular representation of a complex feedstock by Monte Carlo and quadrature methods, *Appl. Catal. A: Gen.* 160 (1997) 41–54.
- [20] D.M. Campbell, C. Bennett, Z. Hou, M.T. Klein, Attribute-based modeling of resid structure and reaction, *Ind. Eng. Chem. Res.* 48 (2009) 1683–1693.
- [21] K. Wang, S. Li, Modified molecular matrix model for predicting molecular composition of naphtha, *Chin. J. Chem. Eng.* 25 (2017) 1856–1862.
- [22] K.M. Van Geem, D. Hudebine, M.F. Reyniers, F. Wahl, J.J. Verstraete, G.B. Marin, Molecular reconstruction of naphtha steam cracking feedstocks based on commercial indices, *Comput. Chem. Eng.* 31 (2007) 1020–1034.
- [23] S.B. Jaffe, H. Freund, W.N. Olmstead, Extension of structure-oriented lumping to vacuum residua, *Ind. Eng. Chem. Res.* 44 (2005) 9840–9852.
- [24] L. Tian, B. Shen, J. Liu, Building and application of delayed coking structure-oriented lumping model, *Ind. Eng. Chem. Res.* 51 (2012) 3923–3931.
- [25] D. Hudebine, J.J. Verstraete, Molecular reconstruction of LCO gasoils from overall petroleum analyses, *Chem. Eng. Sci.* 59 (2004) 4755–4763.
- [26] A. Alvarez-Majmutov, J. Chen, R. Gieleciak, D. Hager, N. Heshka, S. Salmon, Deriving the molecular composition of middle distillates by integrating statistical modeling with advanced hydrocarbon characterization, *Energy Fuels* 28 (2014) 7385–7393.
- [27] M.L. Abelairas, L.P. de Oliveira, J.J. Verstraete, Application of Monte Carlo techniques to LCO gas oil hydrotreating: molecular reconstruction and kinetic modelling, *Catal. Today* 271 (2016) 188–198.
- [28] J.J. Verstraete, N. Revellin, H. Dulot, D. Hudebine, Molecular reconstruction of vacuum gasoils, *Prepr. Pap. Am. Chem. Soc., Div. Fuel Chem.* 49 (2004) 20–21.
- [29] A. Alvarez-Majmutov, J. Chen, R. Gieleciak, Molecular-level modeling and simulation of vacuum gas oil hydrocracking, *Energy Fuels* 30 (2016) 138–148.
- [30] A. Alvarez-Majmutov, J. Chen, Stochastic modeling and simulation approach for industrial fixed-bed hydrocrackers, *Ind. Eng. Chem. Res.* 56 (2017) 6926–6938.

- [31] J.J. Verstraete, P. Schnongs, H. Dulot, D. Hudebine, Molecular reconstruction of heavy petroleum residue fractions, *Chem. Eng. Sci.* 65 (2010) 304–312.
- [32] L.P. de Oliveira, A.T. Vazquez, J.J. Verstraete, M. Kolb, Molecular reconstruction of petroleum fractions: application to vacuum residues from different origins, *Energy Fuels* 27 (2013) 3622–3641.
- [33] A. Alvarez-Majmutov, R. Gieleciak, J. Chen, Modeling the molecular composition of vacuum residue from oil sand bitumen, *Fuel* 241 (2019) 744–752.
- [34] A. Alvarez-Majmutov, R. Gieleciak, J. Chen, Deriving the molecular composition of vacuum distillates by integrating statistical modeling and detailed hydrocarbon characterization, *Energy Fuels* 29 (2015) 7931–7940.
- [35] T.F. Petti, D.M. Trauth, S.M. Stark, M. Neurock, M. Yasar, M.T. Klein, CPU issues in the representation of the molecular structure of petroleum resid through characterization, reaction, and Monte Carlo modeling, *Energy Fuels* 8 (1994) 570–575.
- [36] L.P. de Oliveira, J.J. Verstraete, M. Kolb, Molecule-based kinetic modeling by Monte Carlo methods for heavy petroleum conversion, *Sci. China-Chem.* 56 (2013) 1608–1622.
- [37] C.U. Deniz, M. Yasar, M.T. Klein, Stochastic reconstruction of complex heavy oil molecules using an artificial neural network, *Energy Fuels* 31 (2017) 11932–11938.
- [38] C.U. Deniz, S.H.O. Yasar, M. Yasar, M.T. Klein, Effect of boiling point and density prediction methods on stochastic reconstruction, *Energy Fuels* 32 (2018) 3344–3355.
- [39] C.E. Shannon, A mathematical theory of communication, *Bell Syst. Tech. J.* 27 (1948) 623–656.
- [40] S. Guiasu, A. Shenitzer, The principle of maximum entropy, *Math. Intell.* 7 (1985) 42–48.
- [41] S.P. Pyl, K.M. Van Geem, M.F. Reyniers, G.B. Marin, Molecular reconstruction of complex hydrocarbon mixtures: An application of principal component analysis, *AIChE J.* 56 (2010) 3174–3188.
- [42] D. Hudebine, J.J. Verstraete, Reconstruction of petroleum feedstocks by entropy maximization. Application to FCC gasolines, *Oil Gas Sci. Technol.* 66 (2011) 437–460.
- [43] Y. Pan, B. Yang, X. Zhou, Feedstock molecular reconstruction for secondary reactions of fluid catalytic cracking gasoline by maximum information entropy method, *Chem. Eng. J.* 281 (2015) 945–952.
- [44] S. Feng, C. Cui, K. Li, L. Zhang, Q. Shi, S. Zhao, C. Xu, Molecular composition modelling of petroleum fractions based on a hybrid structural unit and bond-electron matrix (SU-BEM) framework, *Chem. Eng. Sci.* 201 (2019) 145–156.