

# Molecular Reconstruction of Petroleum Fractions: Application to Vacuum Residues from Different Origins

L. Pereira de Oliveira,<sup>†,‡</sup> A. Trujillo Vazquez,<sup>†</sup> J. J. Verstraete,<sup>†,\*</sup> and M. Kolb<sup>‡</sup>

<sup>†</sup>IFP Énergies Nouvelles, Rond-point de l'échangeur de Solaize - BP 3, 69360 Solaize, France

<sup>‡</sup>Laboratoire de Chimie, École Normale Supérieure, 69364 Lyon Cedex 07, France

**ABSTRACT:** A vacuum residue is a complex hydrocarbon mixture of several thousand different chemical species. Even today, no analytical technique is powerful enough to obtain the molecular detail that is required for the development of a detailed kinetic model. To overcome this drawback, a two-step reconstruction algorithm has been developed to build a representative set of molecules from partial analytical data. The first step, called Stochastic Reconstruction (SR), creates an initial mixture of molecules by a Monte Carlo sampling method. The second step, termed Reconstruction by Entropy Maximization (REM), modifies the molar fractions of the molecules in order to improve the representativeness of the generated mixture. The combined SR-REM algorithm creates a synthetic blend of molecules whose mixture properties are close to the analytical data of the petroleum fraction. The method has been applied to petroleum vacuum residue fractions from four different geographic locations with substantially different compositions. All cases are well represented, clearly illustrating the versatility of the SR-REM method. As an extension to this base algorithm, a novel indirect two-step reconstruction algorithm was developed, in which the SR step is used to build a single reference mixture. The set of molecules thus obtained is subsequently used in the second step to represent various petroleum fractions via the REM method. This allows to simultaneously reduce the computational burden and to represent the vacuum residue fractions with the same set of molecules. To validate this alternative approach, eight vacuum residues from different origins have been reconstructed with this technique. The results in terms of analysis prediction have shown a very good agreement.

## 1. INTRODUCTION

Over the past decade, the world demand for high-quality low-boiling products such as gasoline and diesel has been increasing continually, while at the same time the available crude oils are becoming heavier. In this context, the refining industry is focusing even more on recovering every valuable product from every single drop of petroleum crude oil.<sup>1</sup> Hence, refining processes that convert heavy petroleum cuts, such as vacuum residues, into lighter and more valuable products are vital to this industry. Petroleum residue conversion processes, such as residue hydrocracking, residue fluid catalytic cracking (RFCC), or delayed coking, are based on the degradation of the largest oil molecules by thermal and/or catalytic cracking reactions at high temperature.

To improve the performance of conversion processes, the development of reliable and accurate kinetic models is required. However, the development of kinetic models applied to complex petroleum feedstocks is not only intricate due to the large number of chemical species, reactions, and associated rate constants, but its complexity already starts with the difficulties arising in the characterization of this type of mixtures.

Classic kinetic models applied to complex hydrocarbon mixtures generally use a lumping strategy, where molecular components are grouped into several chemical families, according to their global properties, which are very often physical characteristics such as boiling point or solubility.<sup>2–5</sup> However, for heavy hydrocarbon mixtures, the number of families and reaction pathways turns out to be so vast that this lumped approach is no longer manageable.

The limitations of lumped models<sup>6,7</sup> motivated the development of molecular-based kinetic models, which require a molecular description of the feedstock. Unfortunately, no analytical technique is currently powerful enough to identify all the chemical species contained in heavy petroleum fractions. To overcome this limitation, a molecular representation of the feedstock can be built using a molecular reconstruction algorithm. This type of algorithm generates a set of molecules from overall petroleum analyses, model hypotheses and chemical knowledge.

In the literature, several approaches have been described to represent petroleum fractions.

The representation of the petroleum fractions, and especially of asphaltenes, by means of an average model molecule in 2D<sup>8–16</sup> or 3D<sup>17</sup> has been employed by several authors. Such molecules are generated in order to obtain an average structure that represents the various chemical functions of the mixture. This kind of approach may be inadequate for mixtures that contain a very large variety of chemical structures, since the information associated with the polydispersity of the mixture is entirely lost.

Allen and Liguas<sup>6</sup> proposed the first method that considers a set of molecules to represent a middle distillate. This method selects a set of predefined molecules and modifies their respective molar fractions in order to obtain a mixture whose properties are close to the desired analytical data. The analyses used for this

Received: May 6, 2012

Revised: April 7, 2013

Published: July 1, 2013

fitting include gas chromatography (GC),  $^{13}\text{C}$  NMR, and  $^1\text{H}$  NMR and provide up to 190 constraints on the molar fractions.

Khorasheh et al.<sup>18</sup> proposed a similar technique, termed structural group analysis (SGA), to obtain a set of predefined structural groups and their mole fractions. Elemental analysis,  $^{13}\text{C}$  NMR, and  $^1\text{H}$  NMR were employed by the authors. Using the information provided by SGA, Khorasheh et al.<sup>19</sup> developed an algorithm to generate a randomly chosen representative set of molecules for heavy hydrocarbon mixtures.

Quann and Jaffe<sup>2,3</sup> developed a method called structure oriented lumping (SOL) to represent vacuum gas oil and lighter fractions. Each molecule in the mixture is described by a vector of 22 elements, which represent the structural blocks of the molecule. The SOL approach is a very sophisticated lumping method, in which molecules with the same type and number of structural blocks, such as structural isomers, are described by the same vector. Jaffe et al.<sup>4</sup> have extended this method to heavy fractions and more specifically to the representation of multicore species.

Zhang<sup>20</sup> characterized petroleum cuts by a matrix of pseudo-compounds classified by chemical family and carbon number. In order to obtain this matrix, gas chromatography and mass spectrometry (GC/MS) had to be coupled. However, these analyses are long and complex, and difficult to apply on heavy petroleum residues. Several other approaches using variations of matrix representations with homologous series have also been proposed in the literature.<sup>21–25</sup>

Neurock et al.<sup>26,27</sup> developed a method called stochastic reconstruction (SR), where the petroleum cuts are characterized by a set of probability distribution functions for molecular structural attributes (number of cores, number of aromatic rings, the degree of substitution, etc.). These distributions are sampled by a Monte Carlo method in order to obtain a synthetic mixture of molecules. Coupled with an optimization loop, the method yields an accurate molecular representation of a heavy asphaltene feed-stock.<sup>28,29</sup> Several authors of this group<sup>30–35</sup> have continued to improve the method.

Hudebine et al.<sup>7,35,36</sup> developed two algorithms in order to generate a complex mixture of molecules from overall petroleum analyses. The first is based on stochastic reconstruction (SR), and the second is a novel approach called reconstruction by entropy maximization (REM). These algorithms were first applied to light cycle oil (LCO) gas oils. Hudebine and Verstraete<sup>36</sup> have coupled the SR and REM methods into a two-step molecular reconstruction (SR-REM) algorithm to overcome drawbacks of each of these methods: computational effort, noisiness of the objective function and sensitivity to the initial set of molecules. This approach was developed for and validated on LCO gas oils<sup>36,37</sup> and was later extended to vacuum gas oils<sup>38–40</sup> and vacuum residues.<sup>41</sup>

The present work applies the SR-REM method to four vacuum residues from different geographic locations and with different properties in order to assess the versatility of the method. The first part of this paper describes the vacuum residue characteristics. The second part reviews the SR-REM algorithm applied to vacuum residues. A brief description of the SR and REM steps will be given and improvements of the method will be described. The third part is devoted to the application of the SR-REM algorithm to four vacuum residues from different crude oils. In the last part, a novel indirect two-step reconstruction algorithm is presented, in which the SR step is used only once to build a reference mixture for vacuum residues. This reference mixture is then used in a second step to rebuild various vacuum residue

fractions via the REM method. This modified two-step reconstruction algorithm advantageously combines the power of the SR approach to generate appropriate molecules with the flexibility of the REM method, while significantly reducing the computational burden when various residues need to be reconstructed. To validate this alternative approach, eight vacuum residues have been reconstructed with this technique. Finally, Appendix A gives the description of the group contribution methods for boiling point and density that are used in the SR-REM algorithm.

## 2. VACUUM RESIDUE CHARACTERIZATION

A vacuum residue (VR) is a petroleum fraction that is drawn from the bottom of a vacuum distillation column. This petroleum cut is an extremely complex mixture of hydrocarbons that contains several thousands to one million different species.

The characterization of these mixtures is a formidable task given the number of different species. In many cases, vacuum residues are only roughly analyzed.<sup>42–44</sup> Many analyses focus on average physical characteristics, such as density, viscosity, pour point, softening point, sediments, and ash content. The most frequently utilized characterization technique is distillation or simulated distillation, in which the molecules are “sorted” according to their boiling point. A second characterization technique is elemental analysis, in which the abundance of each element in the petroleum fraction is determined. The molecules in vacuum residues are mainly composed of carbon and hydrogen, but they also contain heteroatoms such as sulfur, nitrogen, and oxygen. Metals, such as nickel and vanadium, are also present.<sup>42</sup> Sulfur is the most abundant heteroatom found in vacuum residues,<sup>42</sup> and its presence is undesirable due to its catalyst poisoning and pollution potential. Hence, there are strict environmental specifications that limit the sulfur content in fuels. Nitrogen and metals are also catalyst poisons and strong environmental pollutants. Therefore, these elements are also subject to tight regulations, even though their concentrations are lower than that of sulfur. A third technique that is frequently used to characterize residue fractions is SARA: a separation technique that classifies molecules into saturates, aromatics, resins, and asphaltenes. The saturate fraction is the lightest and least polar fraction of the vacuum residue. This fraction is mainly constituted by paraffinic and naphthenic molecules<sup>43</sup> with small amounts of heteroatoms.<sup>28</sup> Aromatics and resins are both intermediate fractions; however, the first fraction is less polar than the second one. In fact, aromatics are composed of molecules with a moderate concentration of heteroatoms (essentially sulfur), while the molecules with a higher concentration of heteroatoms (sulfur and nitrogen) can be found in the resin fraction. The asphaltenes are the heaviest and most polar fraction of vacuum residues, and they contain high concentrations of sulfur, nitrogen, oxygen, and metals organized in a large variety of chemical functions.<sup>45</sup>

In our work, the vacuum residues have been characterized by means of elemental analysis, average molecular weight, specific gravity, simulated distillation, SARA separation, and  $^{13}\text{C}$  NMR. All characterizations were carried out at IFP Energies nouvelles. The elemental analysis in terms of C–H–N was obtained by thermal conductivity measurements while O was obtained from infrared measurements. The sulfur, nickel, and vanadium contents were determined through wavelength-dispersive X-ray fluorescence. Simulated distillation was carried out by means of an ASTM D5307 derived gas chromatographic method. The gas chromatograph was equipped with a back-flush precolumn to avoid polluting the main column with the heaviest fraction.

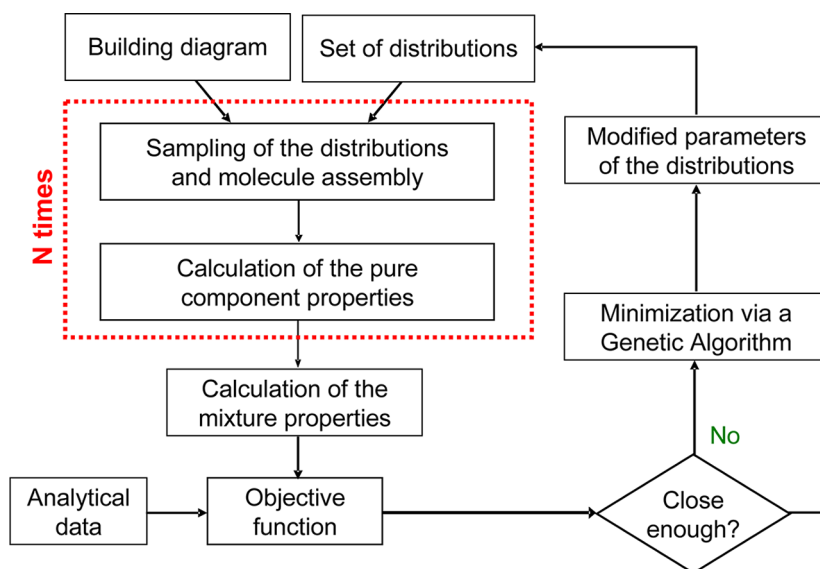


Figure 1. Flow diagram of the stochastic reconstruction step.<sup>36–41</sup>

To determine the amount of the sample entering the main column, the standard addition method was used by adding 4 wt % of a mixture of  $nC_{14}$ – $nC_{15}$ – $nC_{16}$ . For the heaviest VR samples, these three components are eluted before the actual sample, reducing the standard addition method to a calibration via an internal standard. Since this analysis is carried out up to 614 °C, only a partial distillation curve can be obtained for vacuum residues. The specific gravity was obtained through ASTM D4052. The  $^{13}\text{C}$  NMR analyses were performed with an Advance 300 MHz Bruker spectrometer using a 10 mm BBO 1H/X/D NMR probe. The chemical shifts were referenced using deuterated chloroform ( $\text{CDCl}_3$ ) as solvent. Samples were prepared by mixing 100 mg of product in 3 mL of  $\text{CDCl}_3$  so as to obtain a homogeneous solution. Direct  $^{13}\text{C}$  NMR acquisition spectra were obtained with a 60° flip angle at a radio frequency pulse of 20 kHz, which provided the quantity of saturated and unsaturated carbon atoms. In addition, two  $^{13}\text{C}$  NMR experiments based on the scalar coupling between protons and carbons were performed to obtain data about paraffinic, naphthenic, and aromatic carbon atoms. The spin–echo experiments allowed for the aromatic and aliphatic carbon species to be quantified separately, and the attached proton test (APT) series was applied to identify and quantify the proportion of carbon atoms as a function of the number of neighboring protons. The procedure is described in more detail elsewhere.<sup>46</sup> In the SARA separation, the deasphalting step is based on asphaltene flocculation by contacting the sample with *n*-heptane as paraffinic solvent at 80 °C with an oil/heptane ratio of 1:50 for 20 min. After filtering off the asphaltenes, the maltenes undergo a SAR fractionation by a preparative liquid chromatography (LC) procedure.<sup>41,47</sup> The LC system is equipped with a column filled with a bed of alumina followed by a bed of silica. After conditioning for 30 min under *n*-heptane flow, the sample diluted in *n*-heptane is injected onto the column. The group-type separation is performed at room temperature using a mobile-phase program consisting of *n*-heptane (to elute the saturate fraction), a mix of *n*-heptane and toluene (to elute the aromatic fraction), and a mix of tetrahydrofuran and methanol (to elute the resin fraction). The SARA analysis can also be used as a preparative technique in order to separate the four fractions. Then, each fraction can be characterized just as classic petroleum cuts. In the present work,

elemental analysis was carried out on the SARA fractions. Finally, the experimental average molecular weights were obtained from a correlation based on specific gravity and simulated distillation.<sup>48</sup>

Throughout this paper two groups of analyses will be distinguished. The first group, called constrained properties, is constituted by the elemental analysis, average molecular weight, partial simulated distillation, SARA separation, and global  $^{13}\text{C}$  NMR (total saturated carbon and total aromatic carbon). These analyses are used to reconstruct the vacuum residue; that is, they are included in the objective function. The second group, called nonconstrained properties, comprises the analyses that are not included in the objective function; they are to be predicted by the algorithm in order to validate the methodology. This nonconstrained group consists of the elemental analyses of each SARA fraction and the detailed  $^{13}\text{C}$  NMR analysis. This latter analysis measures the content of each type of saturated carbon atoms and aromatic carbon atoms.

### 3. DESCRIPTION OF THE TWO-STEP MOLECULAR RECONSTRUCTION ALGORITHM

In this section, a brief description of the two-step molecular reconstruction (SR-REM) algorithm, as well as the application of this method to the reconstruction of vacuum residues will be presented.

The SR-REM algorithm aims at generating a representative set of molecules whose mixture properties correspond to a given feedstock. This task is performed by using two methods in series. The first step, termed Stochastic Reconstruction (SR), allows to generate an initial set of molecules that are typical for a given type of petroleum fraction. The second step uses the Reconstruction by Entropy Maximization (REM) method in order to improve the properties of the previously generated set of molecules by modifying their molar fractions.

**3.1. Stochastic Reconstruction Method.** The Stochastic Reconstruction (SR) method is based on the characterization of petroleum cuts by means of a set of probability distribution functions (PDF) for the molecular structural attributes.<sup>7,26,27,35–40</sup> Any molecule in the petroleum feedstock can be considered to be an assembly of molecular attributes (e.g., type of molecule, number of cores, number of aromatic rings,



number of side chains, etc.). In practice, a petroleum molecule is constructed by randomly sampling the PDFs of the structural attributes. During the molecule construction, a building diagram and chemical rules must be applied to avoid the creation of impossible and improbable molecules. The building diagram therefore defines the hierarchic relationship between the distributions and the order of the sampling steps, while the chemical rules allow discarding molecules on thermodynamic or probabilistic grounds.

Once the molecule has been constructed, its properties are calculated, either directly, by inspection of the structure (e.g., chemical formula, molecular weight, NMR, mass spectra) or numerically, by group contribution methods (e.g., density, boiling point) or correlations.

To obtain a synthetic mixture, the molecule construction process is repeated  $N$  times by a Monte Carlo sampling technique. The average properties of the mixture are calculated and compared with available analyses through an objective function. This objective function is then minimized with an optimization method, by modifying the parameters of the PDFs of the molecular structural attributes. The various phases of the SR method are shown in Figure 1.

As described, any molecule can be viewed as an assembly of molecular attributes, each of which is represented by a PDF. Therefore, to apply the SR method to a vacuum residue, it is necessary to first choose the molecular attributes and the PDFs corresponding to these attributes. After that, a building diagram for vacuum residue molecules must be specified.

The choice of the molecular attributes needs to be based on expert knowledge of the chemical nature of the vacuum residue.<sup>9–13,15,49–53</sup> As already mentioned, a vacuum residue is a highly complex mixture of several thousands to millions of hydrocarbons species, in which some heteroatoms such as sulfur, nitrogen, and oxygen (in concentrations that vary typically from about 0.5 wt % to about 6 wt %) can be found besides carbon and hydrogen. Although metals are also found in VR fractions, their concentrations are at least 10 times lower on a weight basis. The number of these atoms in hydrocarbon structures is therefore extremely low. By way of example (see also Table 2), for every 1000 carbon atoms in the Buzurgan VR, there are approximately 1416 hydrogen atoms, 25 sulfur atoms, 4 nitrogen atoms, 7 oxygen atoms, 0.05 vanadium atoms, and 0.05 nickel atoms. Due to their low abundance, the metal atoms have been neglected during the residue reconstruction. In the present work, it is assumed that a vacuum residue can be described by 16 structural attributes shown in Table 1. In comparison with previous work,<sup>41</sup> the number of molecular attributes was reduced from 19 to 16. Previously, the probability of having heteroatoms in alkyl chains of monocore and multicore molecules was described by four structural attributes, two for each type of molecule. However, the information provided by current analytical techniques is not detailed enough to distinguish these structural attributes properly. Hence, it is now assumed that these structural attributes are identical for both types of molecules. A modification of how to add heteroatoms into alkyl chains also allowed to eliminate the last molecular attribute. A second major modification with respect to previous work<sup>41</sup> concerns the SARA classification, which separates the vacuum residue into four fractions: saturates, aromatics, resins, and asphaltenes. In the current work, the saturate fraction is considered to consist of  $n$ -paraffins and naphthenes, which contain saturated rings substituted by alkyl chains, while the other fractions are composed of three main structural attributes: aromatic rings, saturated rings, and alkyl chains. In previous work,<sup>41</sup> aromatic

**Table 1. Definition of the Structural Attributes Used in the Stochastic Reconstruction**

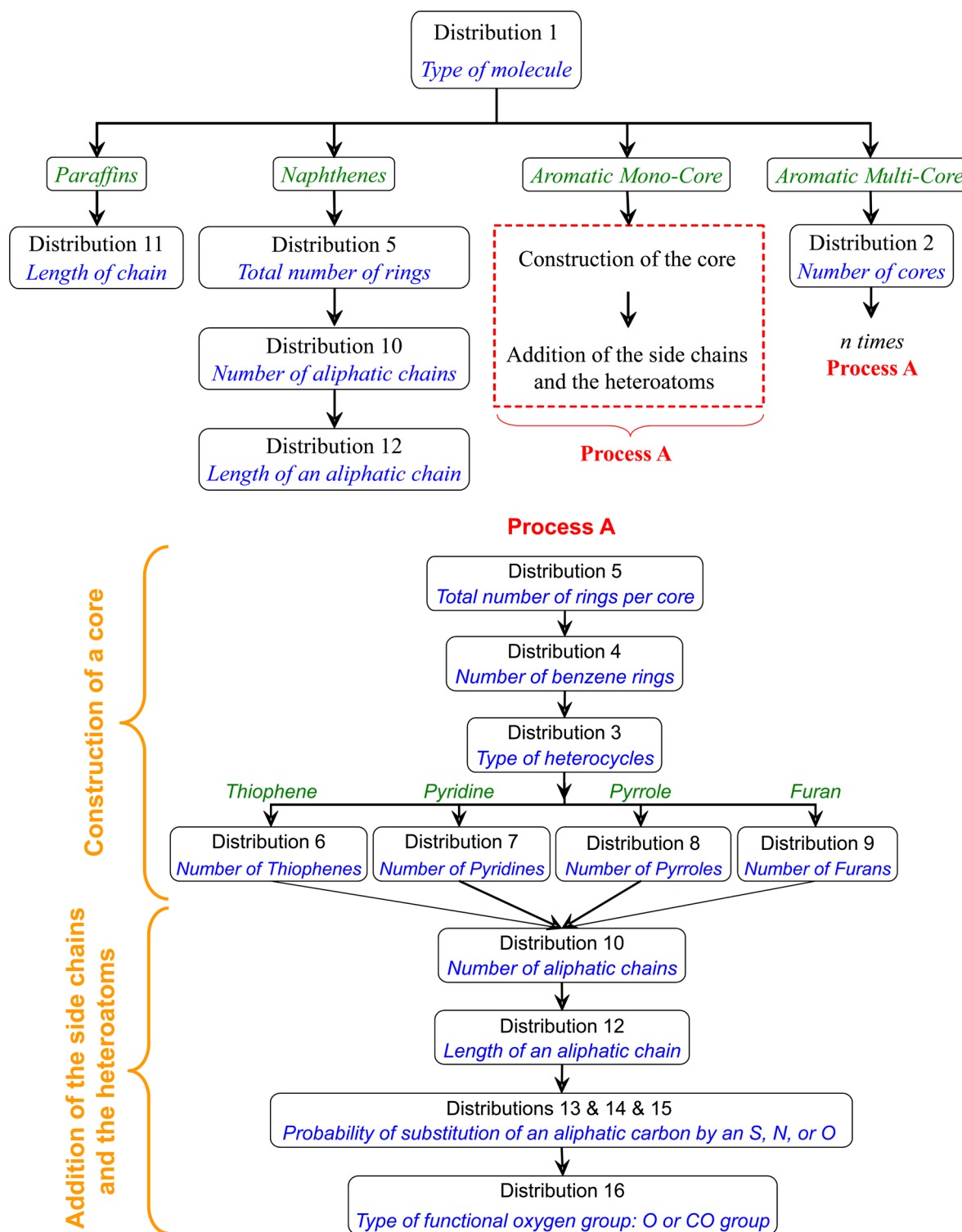
	structural attribute	values	distribution	param. no.
1	type of molecule <sup>a</sup>	0, 1, 2, or 3	histogram	0, 1, and 2
2	no. cores	>1	exponential	3
3	type of heterocycle <sup>a</sup>	0, 1, 2, or 3	histogram	4, 5, and 6
4	no. benzene rings per core	>0	exponential	7
5	total no. rings per core	>1	gamma	8
6	no. thiophenes per core	0, 1, or 2	histogram	9 and 10
7	no. pyridines per core	0, 1, or 2	histogram	11 and 12
8	no. pyrroles per core	0, 1, or 2	histogram	13 and 14
9	no. furans per core	0, 1, or 2	histogram	15 and 16
10	acceptance probability for a peripheral carbon	0 or 1	histogram	17
11	length of the paraffinic chains	>1	gamma	18
12	length of an alkyl chain (lateral and intercore)	>1	exponential	19
13	probability of sulfur substitution for aliphatic CH <sub>3</sub> or CH <sub>2</sub>	0 or 1	histogram	20
14	substitution probability of a carbon atom by a heteroatom	0 or 1	histogram	21
15	type of heteroatom substitution <sup>a</sup>	0 or 1	histogram	22
16	type of oxygen group <sup>a</sup>	0 or 1	histogram	23

<sup>a</sup>Type of molecule: 0, paraffin; 1, naphthene; 2, aromatic monocore; 3, aromatic multicore. Type of heterocycle: 0, thiophene; 1, pyridine; 2, pyrrole; 3, furan. Type of heteroatom: 0, nitrogen; 1, oxygen. Type of oxygen group: 0, ether function; 1, carbonyl function.

and resin molecules were considered to be composed of a single aromatic core, while the asphaltenes were represented by multiple aromatic cores linked together by alkyl chains. However, this approach did not classify some molecules, such as biphenyl, in the correct class. For this reason, a different approach was devised and will be described.

Each structural attribute is described by a PDF. Unfortunately, for vacuum residues, the shapes of the PDF are unknown. To overcome this difficulty, a highly flexible PDF is desirable, as well as a small number of parameters to characterize it. Three kinds of PDF's were used as optimization variables: histograms, exponential functions, and gamma functions. The histograms were used for structural attributes containing a narrow range of possible values (lower than four), while the exponential and gamma functions describe attributes with a wide range of possible values. The criterion to choose between the two latter PDFs was the shape of the distribution. When it was necessary to avoid large values for a structural attribute, the exponential distribution was chosen; otherwise a gamma distribution was used. The detailed list of PDFs of structural attributes is given in Table 1. In the case of histograms, the number of parameters is equal to the possible number of attribute values minus one. Usually, a gamma distribution has two free parameters: a shape parameter ( $\alpha$ ) and a scale parameter ( $\beta$ ). However, in order to reduce the number of parameters, it was assumed that the scale parameter is twice the shape parameter. Hence, to reconstruct a vacuum residue, 16 distributions with a total of 24 associated parameters are used, as shown in Table 1.

All PDFs must have a maximum and a minimum value. The minimum value was set to zero, except for the attributes which cannot be zero such as number of cores, length of paraffinic chains, etc. The maximum value is different for different PDFs. For histograms, the maximum value corresponds to the biggest value of the associated attribute. For instance, structural attribute



**Figure 2.** Building diagram for residue fractions. Process A constructs a single polycyclic core with side chains and heteroatoms.

1 (type of molecule) varies between 0 and 3 (0, paraffins; 1, naphthenes; 2, aromatic monocore; 3, aromatic multicore); thus, the maximum value of the corresponding histogram was set to three. In the case of exponential and gamma distributions, the maximum value of the PDF was fixed where the cumulative probability of the attribute value reaches 0.99. Once the distribution has been truncated, a renormalization is performed to ensure that the probabilities add up to 1.

The building diagram for the vacuum residue is illustrated in Figure 2. The molecule construction begins with choosing a molecule type (paraffin, naphthene, aromatic monocore, and aromatic multicore) by sampling distribution 1. Once the molecule type has been selected, the value of each structural attribute is determined.

According to the analytical data, saturates have a very low concentration of heteroatoms. Therefore, the heteroatom

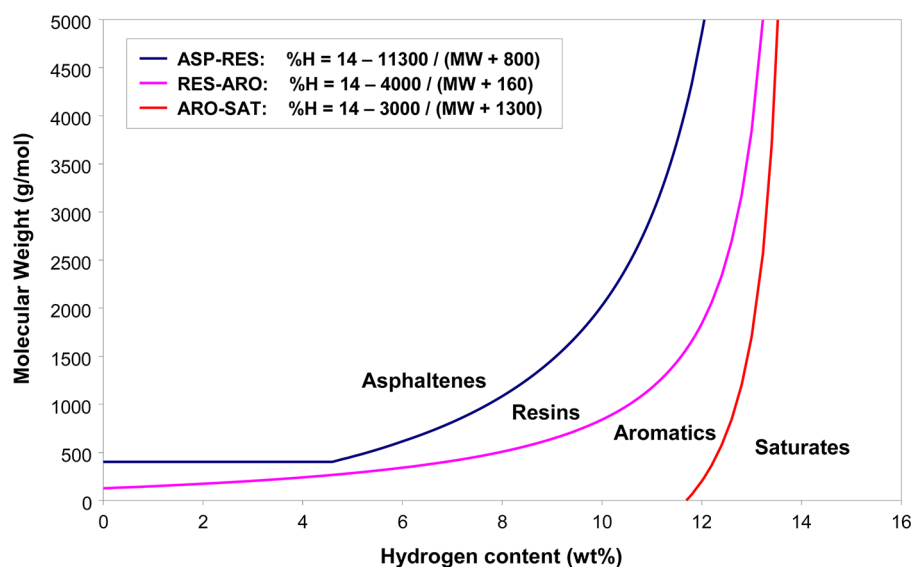


Figure 3. Diagram used to classify each molecule into one of the SARA fractions (based on Wiehe's solvent-resid phase diagram<sup>54,55</sup>).

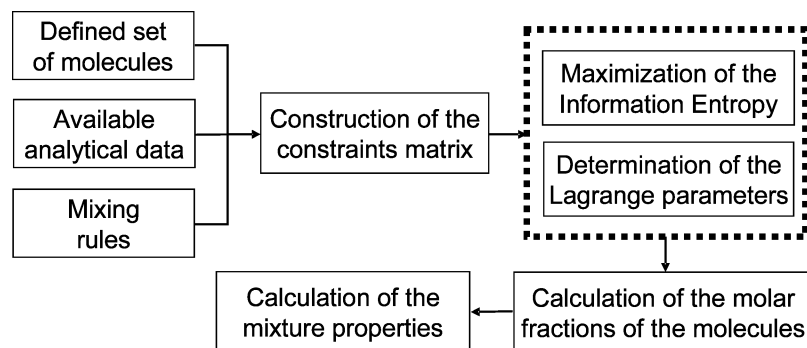


Figure 4. Flow diagram of the entropy maximization step.<sup>36,39,61</sup>

addition for this type of molecule was neglected. Thus, the length of the chain is enough to define the structure of paraffinic molecules. This attribute is determined by sampling distribution 11.

For naphthenic molecules, the present study considers that naphthenes have only one core, which is composed of cyclohexane rings and aliphatic side chains. Thus, this type of molecule is defined by three attributes: the number of rings, the number of side chains, and the length of these side chains. The naphthene molecules are therefore created by sampling distributions 5, 10, and 11.

The monocore (MONO-ARO) molecules and the aromatic multicore (MULTI-ARO) molecules only differ in the number of cores. Accordingly, the first type of molecule has only one core while the second type may have two or more cores. For the latter, the number of cores is determined by sampling distribution 2. Each core is constructed by determining the total number of rings in the core (distribution 5), the number of benzene rings (distribution 4), and the type and number of heterocycles (distributions 3, 6, 7, 8, and 9). The number of cyclohexane rings is obtained by difference. The number of aliphatic chains on a core is chosen by sampling distribution 10, while the length of each chain is determined by sampling distribution 12. Finally, the presence of heteroatoms (sulfur, nitrogen, and oxygen atoms) within a chain is determined according to distributions 13, 14, 15, and 16.

The molecular properties are calculated either by inspection of the structure of the molecule, or by group contribution methods. The chemical formula, molecular weight, and <sup>13</sup>C NMR spectrum are obtained by direct inspection of the structure, while the normal boiling point and the specific gravity are calculated through a novel group contribution method, which is described in Appendix A.

The classification of each molecule in the correct SARA fraction is carried out by means of the SARA diagram given in Figure 3. This SARA diagram is based on the solvent-resid phase diagram proposed by Wiehe.<sup>54,55</sup> In Wiehe's diagram, the domain of each SARA fraction is defined by two properties, molecular weight, and hydrogen content. Gray<sup>56</sup> demonstrated that the asphaltene domain, which is limited by the "heptane-solubility" line, is not clearly defined and proposed a modified "heptane-solubility" line. In the present methodology, Wiehe's solvent-resid phase diagram is used, but several modifications have been applied. The first modification concerns the "heptane-solubility" line, which was adjusted and has been placed between the line proposed by Wiehe<sup>54,55</sup> and the line proposed by Gray.<sup>56</sup> It was also assumed that asphaltenes can not have a molecular weight below 400 g/mol. For the domains of the other SARA fractions, the boundaries were directly drawn based on Wiehe's solvent-resid phase diagram. A second modification concerns the aromatic fraction for which it was assumed that aromatic molecules can have a molecular weight higher than 700 g/mol.

The equations of the proposed boundaries are given in Figure 3. For the aromatic fraction, an additional assumption was made by

**Table 2. Reconstruction of the Buzurgan Vacuum Residue: Comparison between Experimental and Calculated Constrained Properties after the Stochastic Reconstruction (SR) Step and the Entropy Maximization (EM) Step**

		exp.	sim. (after SR)	sim. (after SR and EM)
Elemental Analysis				
carbon	wt %	83.41	82.69	83.41
hydrogen	wt %	9.84	10.14	9.85
sulfur	wt %	5.56	5.76	5.56
nitrogen	wt %	0.39	0.47	0.39
oxygen	wt %	0.80	0.93	0.80
hydrogen/carbon	at/at	1.4157	1.4715	1.4171
sulfur/carbon	at/at	0.0250	0.0261	0.0250
nitrogen/carbon	at/at	0.0040	0.0049	0.0040
oxygen/carbon	at/at	0.0072	0.0084	0.0072
Molecular Weight				
	g/mol	762	738	759
SARA				
saturates	wt %	9.9	9.7	9.9
aromatics	wt %	37.1	36.9	37.1
resins	wt %	37.7	38.7	37.7
asphaltenes	wt %	15.3	14.7	15.3
<sup>13</sup> C NMR				
saturated C	mol %	66.9	68.6	66.9
aromatic C	mol %	33.1	31.4	33.1
Simulated Distillation				
initial BP	°C	379	354	354
5% BP	°C	468	436	468
10% BP	°C	501	500	501
20% BP	°C	539	607	540
30% BP	°C	573	702	573

considering that no nitrogen can be present in aromatic molecules. This assumption is based on the available experimental data.

Once all molecules have been constructed and their pure component properties determined, the mixture needs to be characterized. The average properties of this mixture of  $N$  molecules are calculated with linear mixing rules. The average specific gravity and the simulated distillation are estimated assuming an ideal mixture.<sup>7,35–41</sup>

In the present algorithm, the objective function  $F$  is defined as a weighted average of the relative deviations between calculated and experimental properties:

$$F = \frac{1}{N_A} \sum_{j=1}^{N_A} W_j \delta_j \quad (1)$$

where  $N_A$  is the number of analyses present in the objective function,  $W_j$  is the weighting factor of analysis  $j$ , and  $\delta_j$  is the relative deviation between the calculated and experimental values of analysis  $j$ . The relative deviation is calculated using the following expression:

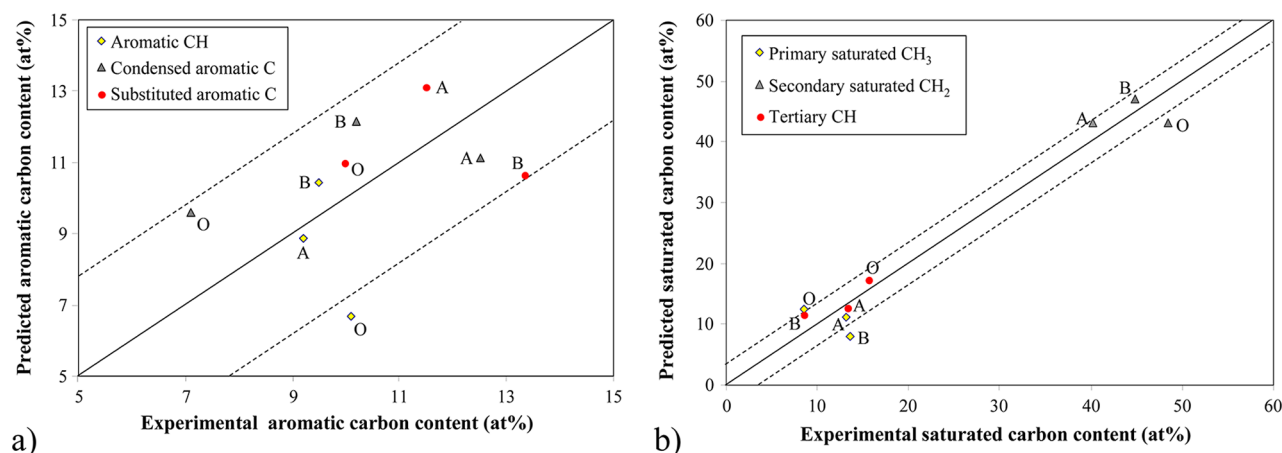
$$\delta_j = \frac{1}{N_{M,j}} \sum_{k=1}^{N_{M,j}} \frac{|X_{k,j}^{\text{exp}} - X_{k,j}^{\text{calc}}|}{X_{k,j}^{\text{exp}}} \quad (2)$$

where  $X_{k,j}^{\text{exp}}$  is the experimental value of the measure  $k$  of analysis  $j$ ;  $X_{k,j}^{\text{calc}}$  is the calculated value of the measure  $k$  of the analysis  $j$  and  $N_{M,j}$  is the number of measurements of analysis  $j$  (e.g., each fraction of the SARA analysis or each atom content in the elemental analysis).

The relative deviations between the calculated and experimental properties are not squared in order to avoid the amplification of the background noise which is present on an objective function obtained by a Monte Carlo based sampling method.

**Table 3. Comparison between Experimental and Calculated Constrained Properties after the Entropy Maximization (EM) Step of the Molecular Reconstruction of the Three Other Vacuum Residues**

		Ural		Maya		Athabasca	
		exp.	sim.	exp.	sim.	exp.	sim.
Elemental Analysis							
carbon	wt %	85.47	85.47	85.25	85.24	82.77	82.77
hydrogen	wt %	10.60	10.60	10.12	10.13	9.86	9.86
sulfur	wt %	2.72	2.72	3.50	3.50	5.74	5.74
nitrogen	wt %	0.58	0.58	0.58	0.58	0.63	0.63
oxygen	wt %	0.63	0.63	0.55	0.55	1.00	1.00
Molecular Weight							
	g/mol	727	724	764	760	800	796
SARA							
saturates	wt %	11.7	11.7	12.9	12.9	6.8	6.8
aromatics	wt %	46.1	46.1	38.7	38.7	31.2	31.2
resins	wt %	37.6	37.6	34.2	34.2	47.9	47.9
asphaltenes	wt %	4.6	4.6	14.2	14.2	14.1	14.1
<sup>13</sup> C NMR							
saturated C	mol %	72.8	72.79	69.5	69.5	66.8	66.8
aromatic C	mol %	27.2	27.21	30.5	30.5	33.2	33.2
Simulated Distillation							
initial BP	°C	384	359	302	277	358	334
5% BP	°C	496	496	484	484	474	474
10% BP	°C	520	520	520	520	505	505
20% BP	°C	550	550	558	558	540	541
30% BP	°C	574	574	585	585	577	578



**Figure 5.** Parity plots of detailed  $^{13}\text{C}$  NMR of Athabasca (A), Buzurgan (B), and Ural (O) vacuum residue: (a) aromatic carbon content and (b) saturated carbon content.

Finally, the objective function  $F$  is minimized by means of an elitist genetic algorithm<sup>7,57–59</sup> until the properties of the mixture are close to the analytical data.

**3.2. Reconstruction by Entropy Maximization.** The reconstruction by Entropy Maximization (REM) method aims at fine-tuning the properties of a predefined set of molecules by adjusting their molar fractions with respect to the analytical data.<sup>7,36,60,61</sup> The REM method is based on the principle of maximum Shannon entropy, according to which an information entropy criterion<sup>62</sup> must be maximized in order to obtain the optimal result. In the REM method, the optimal result corresponds to the set of molecules whose properties are identical to the desired analytical data. The entropic criterion is based on the Shannon entropy:<sup>62</sup>

$$E = - \sum_{i=1}^N x_i \ln(x_i) \quad (3)$$

with

$$\sum_{i=1}^N x_i = 1 \quad (4)$$

where  $E$  represents the Shannon entropy,  $x_i$  is the molar fraction of molecule  $i$ , and  $N$  is the number of molecules present in the predefined set of molecules.

This criterion ensures that no molecule is preferred over any other, when no information (i.e., analytical data of the vacuum residue) is available. The distribution of the predefined set of molecules will then be kept uniform.

Constraints can be added to this criterion by means of Lagrange parameters, as shown in the following equation:

$$H = \underbrace{- \sum_{i=1}^N x_i \ln x_i}_{\text{Information entropy}} + \underbrace{\mu \cdot (1 - \sum_{i=1}^N x_i)}_{\text{Mass balance}} + \underbrace{\sum_{j=1}^J \lambda_j \cdot (f_j - \sum_{i=1}^N x_i \cdot f_{i,j})}_{\text{Analytical constraints}} \quad (5)$$

where  $H$  represents the information entropy criterion,  $x_i$  is the molar fraction of molecule  $i$ ,  $f_j$  is the value of constraint  $j$ ,  $f_{ij}$  represents the property or coefficient of molecule  $i$  for constraint  $j$ ,  $\mu$  represents the Lagrange multiplier associated with the mass

balance constraint,  $\lambda_j$  represents the Lagrange multiplier associated with constraint  $j$ , and  $J$  represents the total number of equality constraints.<sup>7,60,61</sup>

When constraints (i.e., analytical data) are imposed, the uniform distribution is distorted until the correct properties are obtained with the constraints of the entropic criterion. The constraints used in this step for vacuum residues are the same as those used in the SR step.

Since the mixing rules are linear, the optimal result is obtained by maximizing a nonlinear equation with  $J$  parameters, where  $J$  is the number of constraints. A schematic diagram of the REM method is depicted in Figure 4.

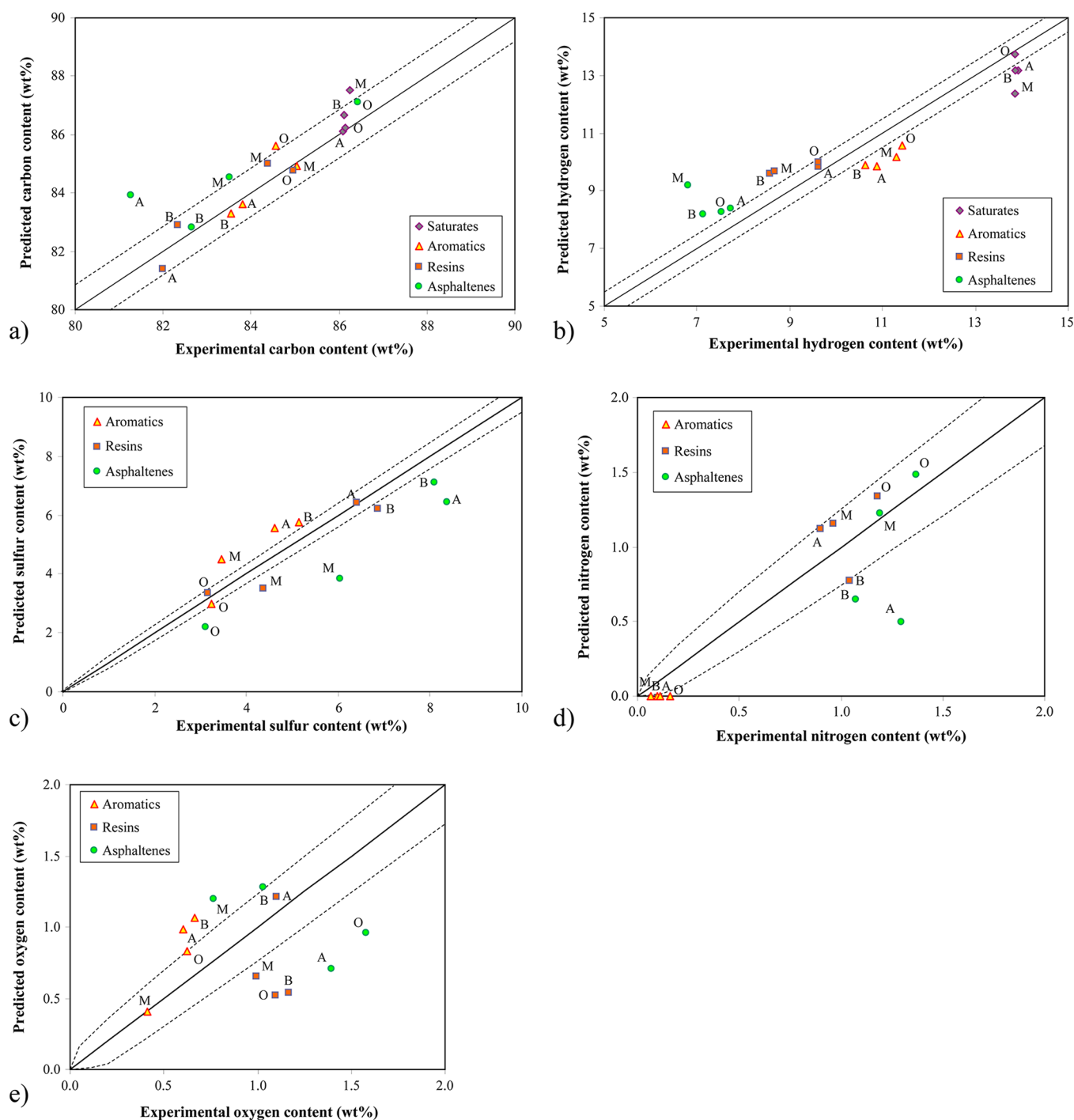
#### 4. APPLICATION OF THE TWO-STEP RECONSTRUCTION ALGORITHM

To test the versatility of the methodology, four vacuum residues with different chemical compositions have been reconstructed. The vacuum residues have been chosen to ensure a large diversity of their characteristics, as well as a good representation of the world's main petroleum reserves.

The methodology was applied to vacuum residues from Russian (Ural), Mexican (Maya), Middle Eastern (Buzurgan), and Canadian (Athabasca) crude oils. The characterization of the vacuum residues from different crude oils is presented in Tables 2 and 3. Ural (Table 3) is the lightest vacuum residue, with the lowest average boiling point, lowest molecular weight, highest hydrogen content, and lowest sulfur content. The Maya vacuum residue (Table 3) has a somewhat lower hydrogen content and a higher sulfur content but has the highest average boiling point of the four vacuum residues. Both the Buzurgan (Table 2) and Athabasca (Table 3) vacuum residue are very heavy fractions containing high amounts of sulfur and a very low hydrogen content. The Buzurgan vacuum residue contains the highest amount of asphaltenes, while the Athabasca vacuum residue has the highest average molecular weight and the highest amount of polar material (resins and asphaltenes).

For each vacuum residue, the molecular reconstruction was performed with sets of 5000 molecules in order to ensure an optimal balance between the required CPU time and the accuracy of the feedstock representation. In the objective function the same weighting factor was used for all analyses. In the genetic algorithm, an initial population of the 2048 individuals (sets of 5000 molecules) was used. To decrease the time required for parameter estimation, the population is progressively reduced to





**Figure 6.** Parity plots for the elemental analysis prediction of each SARA fraction of Athabasca (A), Buzurgan (B), Maya (M), and Ural (O) vacuum residues. The elemental analysis contains the measurement of carbon (a), hydrogen (b), sulfur (c), nitrogen (d), and oxygen content (e).

128 individuals. In each generation, the 50% best individuals are combined by one-point crossover in order to create their offspring. On the offspring, a single-point mutation was applied to each parameter, with a mutation probability of 0.25. Both the crossover and the mutation are carried out through floating point operators as described by Parker.<sup>41</sup> The parameter estimation procedure was performed during 100 generations (iterations).

To illustrate the importance of each step in the two-step SR-REM algorithm, the comparison between analytical and calculated properties of the Buzurgan vacuum residue is shown in Table 2 at the end of each step.

The properties of the synthetic mixture obtained after the SR step are already in good agreement with most of the analytical data. The elemental analysis, SARA fraction separation and <sup>13</sup>C NMR are well represented. Some differences are, however, observed in the partial simulated distillation curve and the average molecular weight. These differences may be caused by the group contribution method used to estimate the boiling point. This group contribution method has been developed from a large database of boiling point data, but most of the data available pertain to molecules with fewer than 42 carbon atoms. Therefore, extrapolation to very large compounds may increase the

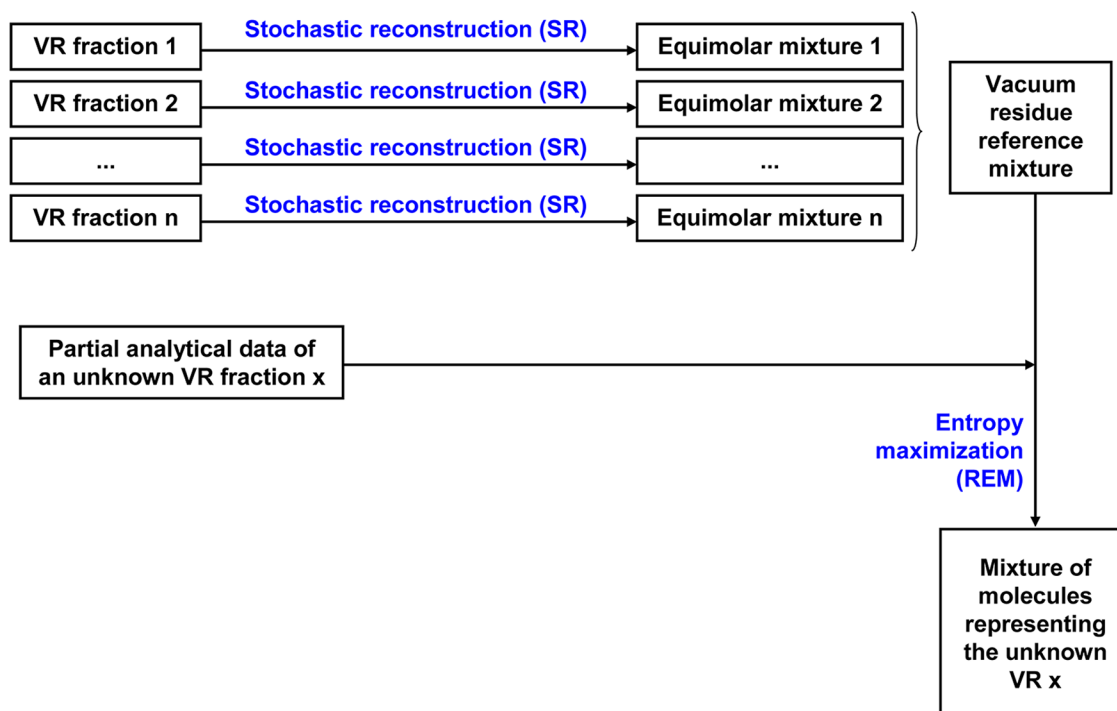


Figure 7. Modified two-step reconstruction algorithm.

uncertainty of the method. As the boiling point increases with the size of a hydrocarbon molecule and hence its molecular weight, including the average molecular weight in the objective function provides a second measure of the molecular size, together with the average boiling point.

However, after the REM step has been performed, a clear improvement is obtained in the properties of the mixture, as illustrated in Table 2. The elemental analysis, SARA fractions separation, and  $^{13}\text{C}$  NMR were perfectly adjusted. In the case of molecular weight and simulated distillation, the observed deviations were significantly reduced except for the initial boiling point. This is due to the fact that the initial boiling point in the algorithm corresponds to the boiling point of a single molecule, that is, the lightest one.

All four vacuum residues show a similar improvement after the REM step. Therefore, for the other three crude oils, only the results of the final mixture are shown. The experimental and the predicted properties of the Ural, Maya, and Athabasca vacuum residue are listed in Table 3.

Between the four vacuum residues, large differences are found in the experimental sulfur content, oxygen content, molecular weight, and SARA fractions. In spite of these differences, all properties are well predicted except for the initial boiling point, as explained.

The SR-REM methodology was validated by predicting properties that were not used during the molecular reconstruction, that is, the properties that are not included in the objective function. As indicated in the section Vacuum Residue Characterization, a detailed  $^{13}\text{C}$  NMR analysis of the vacuum residue and the elemental analyses of each SARA fraction were also performed and will now be used to validate the methodology through direct prediction of the characteristics.

The comparison between the predicted and experimental data is illustrated in the form of parity plots (Figures 5 and 6). The full line corresponds to a perfect fit (parity line), while the dotted lines delimit the uncertainty interval of the experimental data.

The detailed  $^{13}\text{C}$  NMR analyses are illustrated in Figure 5 in terms of aromatic carbon (a) and saturated carbon (b) content. The prediction of the Maya vacuum residue is not shown, since the experimental data was not available. The predicted values match the experimental values almost to within the range of the experimental error. As already mentioned and illustrated, the vacuum residues present quite some variation in their contents of the different types of carbon. This shows that in vacuum residues from different crude oils, the various types of molecules do not have the same abundance, even though the total carbon contents are relatively constant (see Tables 2 and 3). The SR-REM algorithm is able to select the appropriate molecules and their abundance for vacuum residues with different compositions.

Concerning the elemental analysis of the SARA fractions, the carbon contents (Figure 6a) are all well predicted, almost within the experimental analytical error, without providing any information.

For the hydrogen content of the SARA fractions (Figure 6b), its value is underestimated for saturates and aromatics fractions and overestimated for asphaltenes and resins fractions. In the case of saturates fractions, this bias can be explained by the fact that the algorithm generates mixtures that contain too many naphthenes and too few paraffins.<sup>41</sup> For the other fractions, the deviations may be caused by an imprecise definition of each SARA fraction range in the SARA diagram. Indeed, although the concept is relatively simple, the boundaries between aromatics and resins and between resins and asphaltenes solvent-resid phase diagram are not very clear and quite complicated to define.

The nitrogen contents of the SARA fractions (Figure 6d) are also well predicted, despite slight deviations in the asphaltene fraction. It should be stressed that this asphaltene fraction is very difficult to predict. Although the SARA diagram allows a reasonable identification of the SARA type of each molecule, the asphaltenes range in the solvent-resid phase diagram is the least clearly defined.<sup>56</sup> Moreover, since the asphaltenes are only a small fraction in the vacuum residue, the objective function is not very sensitive to variations in the characteristics of the small

Table 4. Characteristics of the Vacuum Residue Reference Database of 20 000 Molecules

		VR reference database			
		min.	max.	avg.	SD
Atom Distribution					
carbon no.		10	418	52.1	36.0
hydrogen no.		6	639	76.5	58.1
sulfur no.		0	16	1.05	1.40
nitrogen no.		0	7	0.31	0.69
oxygen no.		0	8	0.38	0.75
ring sulfur atoms		0	8	0.39	0.77
ring nitrogen atoms		0	7	0.26	0.64
ring oxygen atoms		0	7	0.21	0.58
saturated carbon atoms		0	309	36.4	30.6
aromatic carbon atoms		0	160	15.8	13.5
<sup>13</sup> C NMR Analysis					
saturated carbon	at %	0	100.00	63.68	25.22
saturated CH <sub>3</sub>	at %	0	36.36	9.76	4.45
saturated CH <sub>2</sub>	at %	0	97.98	44.29	20.45
saturated CH	at %	0	60.00	9.60	10.76
saturated C quaternary	at %	0	5.00	0.04	0.27
aromatic carbon	at %	0	100.00	36.32	25.22
aromatic CH	at %	0	71.43	12.29	14.20
aromatic C quat. cond.	at %	0	62.50	13.06	11.81
aromatic C quat. subst.	at %	0	36.36	10.96	6.09
Miscellaneous					
molecular weight	g/mol	156	6083	750	518
density	g/mL	0.771	1.861	1.085	0.135
boiling point	°C	277		972	742
Conradson carbon	wt %	0	100	16.68	22.11
SARA Analysis					
		no. molecules	mole fraction	wt fraction	avg MW
saturates		2225	0.11125	0.10894	732
aromatics		9379	0.46895	0.37047	590
resins		7433	0.37165	0.40371	812
asphaltenes		963	0.04815	0.11687	1813

number of asphaltene molecules. Another problem arises from the consistency of the analytical data. As illustrated for the Buzurgan vacuum residue, the prediction of nitrogen content was underestimated for all fractions, despite a well predicted overall content. This reveals the existence of inconsistencies in the analytical data. It should also be noted that the aromatic fractions contain small amounts of nitrogen. Given that the reconstruction scheme assumes that there are no nitrogen atoms in the aromatic fractions, this cannot be predicted.

As illustrated in Figure 6c and e, the sulfur and the oxygen are the elements that have the greatest variability between the different SARA fractions: 3.2–8.4% for sulfur and 0.4–1.6% for oxygen. The predicted sulfur content is reasonably close to the experimental values, except for the asphaltene fractions. The reason for this exception has been explained above. The prediction of the oxygen content is relatively inaccurate, especially in the case of the asphaltenes and the resins. However, it should be noted that the oxygen content is very low and the algorithm is less efficient in the prediction of low values.

Given the complexity of the chemical system as well as the experimental uncertainties of the analytical tests, the predicted properties, which have not been tuned to by the algorithm, are reasonably close to the experimental values. Hence, the SR-REM algorithm is able to reconstruct vacuum residues from very different crude oils with reasonable precision.

## 5. MODIFIED TWO-STEP RECONSTRUCTION ALGORITHM

As illustrated, the direct two-step reconstruction algorithm can be successfully applied to vacuum residues to generate, for each sample, a detailed synthetic mixture that adequately mimics the properties of the vacuum residues. However, as the first step of this reconstruction method uses a Monte Carlo approach to generate its initial set of molecules, it still remains a computationally demanding algorithm,<sup>63</sup> requiring typically several hours up to 1 day of CPU time. A modified indirect two-step reconstruction algorithm was therefore derived and may be applied when a large number of vacuum residues needs to be represented.

Instead of generating a set of molecules for each vacuum residue, the idea consists in generating a large reference database of representative molecules once and for all. Subsequently, only the REM step is employed, requiring only a few seconds of CPU time. The REM step starts from this reference mixture (instead of a Monte Carlo generated vacuum residue specific mixture) to obtain a new mixture whose properties are very close to those of the vacuum residue to be represented. This mixture will now contain the same molecules as the reference mixture but with modified molar fractions. Figure 7 illustrates this modified two-step reconstruction algorithm. To be representative of various types of vacuum residues, a reference mixture needs to be created that contains a sufficient number of molecules covering the various characteristics.

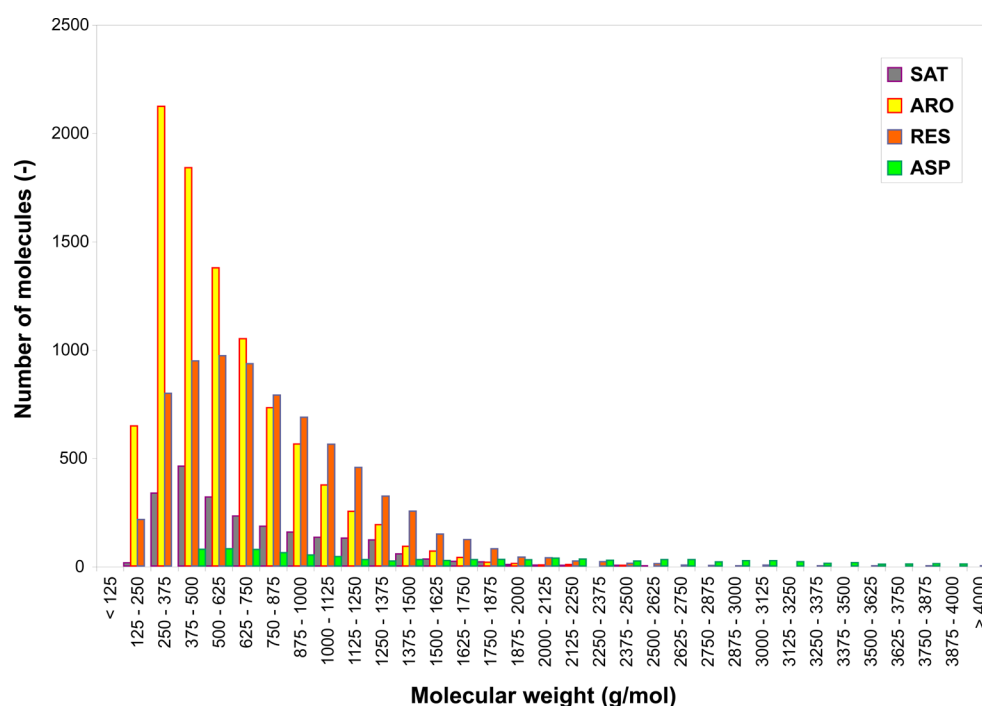


Figure 8. Distribution of the SARA classes in the vacuum residue reference database of 20 000 molecules.

Table 5. Comparison between Experimental and Calculated Constrained Properties after the Entropy Maximization (REM) Step of the Molecular Reconstruction Using the Modified Algorithm Based on a Single Vacuum Residue Reference Database

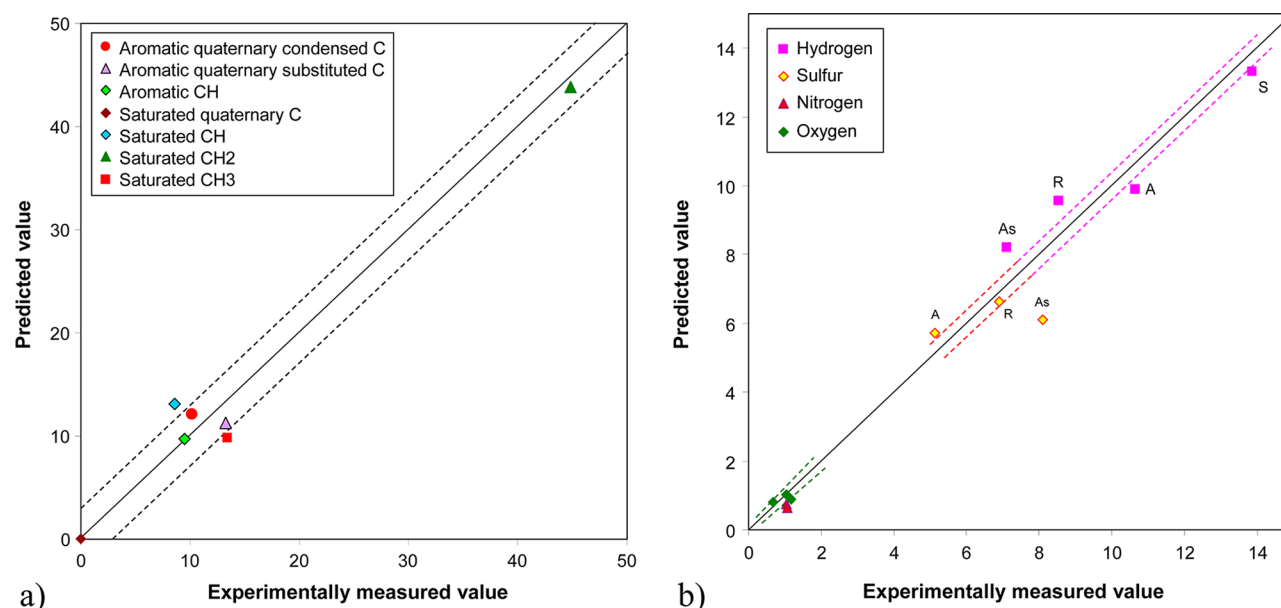
		Athabasca		Buzurgan		Maya		Ural	
		exp.	sim.	exp.	sim.	exp.	sim.	exp.	sim.
Elemental Analysis									
carbon	wt %	82.77	82.77	83.41	83.41	85.25	85.24	85.47	85.46
hydrogen	wt %	9.86	9.86	9.84	9.84	10.12	10.13	10.60	10.61
sulfur	wt %	5.74	5.74	5.56	5.56	3.50	3.50	2.72	2.72
nitrogen	wt %	0.63	0.63	0.39	0.39	0.58	0.58	0.58	0.58
oxygen	wt %	1.00	1.00	0.80	0.80	0.55	0.55	0.63	0.63
Molecular Weight									
	g/mol	800	797	762	760	764	761	727	725
SARA									
saturates	wt %	6.8	6.8	9.9	9.9	12.9	12.9	11.7	11.7
aromatics	wt %	31.2	31.2	37.1	37.1	38.7	38.7	46.1	46.1
resins	wt %	47.9	47.9	37.7	37.7	34.2	34.2	37.6	37.6
asphaltenes	wt %	14.1	14.1	15.3	15.3	14.2	14.2	4.6	4.6
<sup>13</sup> C NMR									
saturated C	mol %	66.8	66.8	66.9	66.9	69.5	69.5	72.8	72.8
aromatic C	mol %	33.2	33.2	33.1	33.1	30.5	30.5	27.2	27.2
Simulated Distillation									
initial BP	°C	358	334	379	354	302	277	384	359
5% BP	°C	474	474	468	468	484	484	496	496
10% BP	°C	505	505	501	501	520	520	520	520
20% BP	°C	540	541	539	540	558	558	550	550
30% BP	°C	577	577	573	573	585	585	574	574

To illustrate the approach, a vacuum residue reference mixture was created by collecting into a single database the sets of molecules obtained for each of the four vacuum residues described. In this way, the vacuum residue reference database contains 20 000 molecules in total. This new set of molecules constitutes an excellent starting point. Indeed, the constructed molecules are typical of vacuum residues given the building rules used by the stochastic reconstruction step. Moreover, the properties of the corresponding synthetic mixture are close to the average

properties of the four vacuum residues. The properties of the vacuum residue reference database are illustrated in Table 4 and in Figure 8. Consequently, this database of 20 000 molecules may now be used to rebuild vacuum residues from their analytical data by employing only the REM step.

This modified two-step reconstruction algorithm was validated in two different ways. First, the vacuum residue reference database was used to generate a mixture of molecules for each of the four vacuum residues used (Athabasca, Buzurgan,





**Figure 9.** Prediction of nonconstrained properties for the Buzurgan vacuum residue: (a) detailed  $^{13}\text{C}$  NMR analysis; (b) elemental composition of the SARA fractions.

**Table 6.** Comparison between Experimental and Calculated Constrained Properties after the Entropy Maximization (REM) Step of the Molecular Reconstruction Using the Modified Algorithm Based on a Single Vacuum Residue Reference Database

		Ardjuna		Djeno		Duri		Arab Medium	
		exp.	sim.	exp.	sim.	exp.	sim.	exp.	sim.
Elemental Analysis									
carbon	wt %	87.47	87.47	87.08	87.08	86.96	86.95	84.40	84.40
hydrogen	wt %	11.03	11.03	11.23	11.23	11.20	11.21	10.14	10.14
sulfur	wt %	0.27	0.27	0.41	0.41	0.56	0.56	4.72	4.72
nitrogen	wt %	0.31	0.31	0.66	0.66	0.56	0.56	0.31	0.31
oxygen	wt %	0.92	0.92	0.62	0.62	0.72	0.72	0.43	0.43
Molecular Weight									
	g/mol	811	809	853	850	847	844	784	781
SARA									
saturates	wt %	18.6	18.6	17.7	17.7	23.6	23.6	8.3	8.3
aromatics	wt %	36.1	36.1	30.1	30.1	30.8	30.8	43.6	43.6
resins	wt %	41.8	41.8	48.0	48.0	39.0	39.0	40.7	40.7
asphaltenes	wt %	3.5	3.5	4.2	4.2	6.6	6.6	7.4	7.4
$^{13}\text{C}$ NMR									
saturated C	mol %	76.1	76.1	78.1	78.1	79.7	79.7	68.5	68.5
aromatic C	mol %	23.9	23.9	21.9	21.9	20.3	20.3	31.5	31.5
Simulated Distillation									
initial BP	°C	450	425	407	382	463	438	420	395
5% BP	°C	541	540	530	530	541	541	532	532
10% BP	°C	557	557	553	553	562	562	552	553
20% BP	°C	578	578	585	585	592	592	581	581
30% BP	°C	599	599	611	611	613	613	603	604

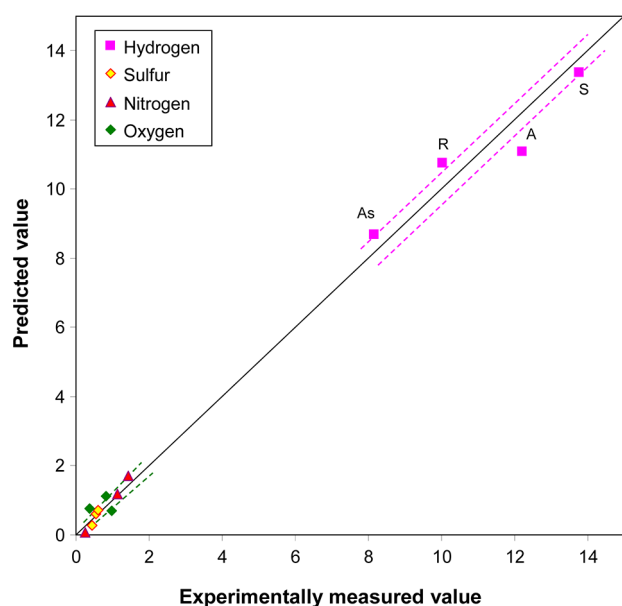
Maya, and Ural). In a second validation, the vacuum residue reference database was used to generate a mixture of molecules for four different vacuum residues: an Ardjuna VR, a Djeno VR, a Duri VR, and an Arab Medium VR.

Starting from this single vacuum residue reference database, the REM algorithm adjusts the mole fractions of the 20 000 components in the mixture to create a mixture that has exactly the same overall properties as provided by the user (Table 5). The mixtures obtained were further validated by predicting nonconstrained properties that were not used during the REM

step, that is, the properties that are not included during the objective function maximization. As shown for the Buzurgan vacuum residue, the various types of carbon are well-predicted (Figure 9a). The same is true for the elemental composition of the various SARA fractions (Figure 9b). For all vacuum residues studied, the results show a good agreement between the predicted and experimental analyses.

In a second validation, the vacuum residue reference database was used to generate a mixture of molecules for four different vacuum residues that were not used to generate this vacuum

residue reference database: an Ardjuna VR, a Djeno VR, a Duri VR, and an Arab Medium VR. The results are listed in Table 6. Although the three first vacuum residues contain more hydrogen, less sulfur, more saturates, and less aromatic carbon and have a higher molecular weight than the four vacuum residues that were used to create the vacuum residue reference database, the REM algorithm was able to modify the abundance of the 20 000 molecules in order to adjust to the user-provided analytical data. Concerning the prediction of the nonconstrained properties, the results are quite accurate as illustrated for the Djeno vacuum residue in Figure 10.



**Figure 10.** Prediction of nonconstrained properties for the Djeno vacuum residue: elemental composition of the SARA fractions.

## 6. CONCLUSIONS

A two-step molecular reconstruction (SR-REM) algorithm has been developed in order to obtain an accurate molecular representation of petroleum fractions. In this work, the direct SR-REM algorithm was applied to four vacuum residues from different crude oils, in order to illustrate the adaptability of this method.

As indicated by its name, the SR-REM algorithm is a method with two successive steps: a Stochastic Reconstruction (SR) step and a Reconstruction by Entropy Maximization (REM) step. The SR step provides an initial set of molecules whose properties are close to experimental data by sampling of attribute probability distribution functions (PDFs). Adding the REM method as a second step significantly improves the properties of the set of molecules by adjusting their molar fractions. The final set of molecules therefore has properties that are very close to the experimental data.

The application of the direct SR-REM algorithm to residues of the Middle East (Buzurgan), Russia (Ural), Mexico (Maya), and Canada (Athabasca) yields a synthetic mixture of molecules that mimics the properties of all vacuum residues very well, despite the important variability of the characteristics of the different crude oils.

Finally, to validate the methodology and to assess the predictive performance of the method, detailed  $^{13}\text{C}$  NMR and elemental analysis of the each SARA fraction of the set of molecules were calculated and these properties were then compared

with analytical data. The detailed  $^{13}\text{C}$  NMR are very well predicted. This fact reveals that the algorithm adapts well to mixtures that contain different types of molecules. The elemental analysis were reasonably predicted for saturates, aromatics, and resins. Concerning the asphaltenes fraction, there are notable deviations in the prediction of their elemental compositions. However, this fraction is too complicated to predict, since it is composed of the most complex molecules of the vacuum residue and represents the smallest fraction of this petroleum cut.

To reduce the computational burden, a modified indirect two-step reconstruction algorithm was developed. The idea consists of using the SR step to build a reference mixture. The set of molecules thus obtained is then used, in a second step, to build petroleum fractions via the REM method. This indirect two-step reconstruction algorithm advantageously combines the power of the SR approach to generate appropriate molecules with the flexibility of the REM method, while completely eliminating the computational burden when various residues need to be reconstructed. To validate this alternative approach, eight vacuum residues have been reconstructed with this technique. The results in terms of analytical predictions have shown a very good agreement.

In conclusion, the SR-REM algorithm is able to provide a molecular representation of a vacuum residue by combining its available analytical properties, and to distinguish the molecular composition of the different petroleum vacuum residues. The resulting synthetic mixtures then allow to access additional compositional information. The SR-REM method can therefore not only be used to obtain the molecular information that it is not provided by the analytical data but also to provide a molecular representation as input for molecule-based kinetic models.

## ■ APPENDIX A: GROUP CONTRIBUTION METHODS

Hudebine<sup>7,61</sup> developed a simple and accurate group contribution method to estimate normal boiling point and liquid density at 20 °C of organic compounds. In this method, the prediction of the properties is based exclusively on the structure of the molecule, which is performed by summing the number of occurrences of each group in the molecule times its corresponding contribution. In order to predict these properties 77 structural groups were defined. The definition of these groups along with sample assignments is given in Table A1.

The normal boiling point is estimated using the equation of the first-order groups proposed by Constantinou and Gani:<sup>64</sup>

$$\exp\left(\frac{T_b}{T_{b,0}}\right) = \sum_i (n_i \Delta T_{b,i}) + F_{T_b} \quad (\text{A.1})$$

where,  $\Delta T_{b,i}$  is the group contribution of group  $i$  (K);  $n_i$  is the number of groups of type  $i$ ,  $T_{b,0}$  is an adjustable parameter ( $T_{b,0} = 307.63$  K) and  $F_{T_b}$  is a boiling point correction term, which can be calculated by the following equation:

$$F_{T_b} = 0.2285N_R^2 + 0.4678N_R \quad (\text{A.2})$$

where  $N_R$  is the number of rings in the molecule.

The liquid density at 20 °C is estimated from the following equation:

$$d = \frac{MW}{\sum_i n_i \Delta V_{m,i} + F_{V_m}} \quad (\text{A.3})$$

where  $\Delta V_{m,i}$  is the group contribution of group  $i$  ( $\text{cm}^3/\text{mol}$ );  $n_i$  is the number of groups of type  $i$ , MW is the molecular weight of

Table A1. Description of the Structural Groups along with Sample Assignments

group name	description	examples (occurrences)
CH3_aliphatic	CH3 in a chain	isooctane (5)
CH2_aliphatic	CH2 in a chain	isooctane (1)
CH_aliphatic	CH in a chain	isooctane (1)
C_aliphatic	C in a chain	isooctane (1)
CH2_naphthenic	CH2 that is part of only one saturated ring	1,1,2-trimethylcyclohexane (3)
CH_naphthenic_substituted	CH that is part of only one saturated ring	1,1,2-trimethylcyclohexane (1)
CH_naphthenic_condensed	CH that is part of two saturated ring	4a-methyldecahydronaphthalene (1)
C_naphthenic_bisubstituted	C that is part of only one saturated ring	1,1,2-trimethylcyclohexane (1)
C_naphthenic_cond_subs	C that is part to least two saturated ring	4a-methyldecahydronaphthalene (1)
CH2_olefinic	CH <sub>2</sub> containing one double bond in a chain	ethylene (2)
CH_olefinic	CH containing one double bond in a chain	2-methyl-2,3-pentadiene (1)
C_olefinic	C containing one double bond in a chain	2-methyl-2,3-pentadiene (1)
C_diolefinic	C containing two double bond in a chain	2-methyl-2,3-pentadiene (1)
CH_ringolefinic	CH <sub>2</sub> containing one double bond in a nonaromatic ring	1,2,3,4,5,8-hexahydronaphthalene (2)
C_ringolefinic_condensed	CH <sub>2</sub> containing one double bond that is part to at least one nonaromatic ring	1,2,3,4,5,8-hexahydronaphthalene (2)
C_ringolefinic_substituted	CH <sub>2</sub> containing one double bond that is part of only one nonaromatic ring	1-methylcyclopentene (1)
CH_triplebonded	CH containing one triple bond in a chain	1-pentyne (1)
C_triplebonded	C containing one triple bond in a chain	1-pentyne (1)
CH_aromatic	CH that is part of only one aromatic ring	1-methylpyrene (9)
C_aromatic_substituted	C that is part of only one aromatic ring	1-methylpyrene (1)
C_aromatic_condensedperipheral	C that is part of two aromatic ring	1-methylpyrene (4)
C_aromatic_condensedinternal	C that is part of three aromatic ring	1-methylpyrene(2)
C_aromatic_condensedperipheral_naphtene	C that is part of one aromatic ring and one nonaromatic ring	tetraphthene (2)
C_aromatic_condensedinternal_naphtene	C that is part of two aromatic ring and one nonaromatic ring	phenalene (1)
C_aromatic_condensedinternal_binaphtene	C that is part of one aromatic ring and two nonaromatic ring	tetraphthene (1)
SH_mercaptan_aliphatic	SH without connection to a ring	1-hexanethiol (1)
SH_mercaptan_naphthenic	SH connected to a nonaromatic ring	cyclohexanethiol (1)
SH_mercaptan_thiophenolic	SH connected to an aromatic ring	benzenethiol (1)
S_sulfide_aliphatic	S without connection to a ring	dimethyl sulfide (2)
S_sulfide_naphthenic	S connected to at least one nonaromatic ring	methyl cyclohexyl sulfide (1)
S_sulfide_benzoic	S connected to at least one aromatic ring	cyclohexyl phenyl sulfide (1)
S_ring_sulfide	S in a nonaromatic ring	tetrahydrothiophene (1)
S_ring_sulfide_thiophenic	S in an aromatic ring	thiophene (1)
S_disulfide_aliphatic	S connected to another S and an aliphatic C	dimethyl disulfide (2)
S_disulfide_naphthenic	S connected to another S and a nonaromatic ring	methyl-cyclohexyl disulfide (1)
S_disulfide_benzoic	S connected to another S and an aromatic ring	ethyl phenyl disulfide (1)
NH2_aliphatic	NH <sub>2</sub> without connection to a ring	hexylamine (1)
NH2_naphthenic	NH <sub>2</sub> connected to a nonaromatic ring	hexahydroaniline (1)
NH2_benzoic	NH <sub>2</sub> connected to an aromatic ring	aniline (1)
NH_aliphatic	NH without connection to a ring	dimethylamine (2)
NH_naphthenic	NH connected to at least one nonaromatic ring	methyl-cyclohexylamine (1)
NH_benzoic	NH connected to at least one aromatic ring	cyclohexylaniline (1)
N_aliphatic	N without connection to a ring	trimethylamine (3)
N_naphthenic	N connected to at least one nonaromatic ring	dimethyl-cyclohexylamine (1)
N_benzoic	N connected to at least one aromatic ring	N-cyclohexyl-N-methylaniline (1)
NH_ring_amine	NH in a nonaromatic ring	pyrrolidine (1)
N_ring_amine	N in a nonaromatic ring	piperidine (1)
N_ring_pyridinic	N in an aromatic ring	pyrrole (1)
NH_ring_pyrrolic	NH in a five-membered aromatic ring	piridine (1)
N_ring_pyrrolic_subst	N in a five-membered aromatic ring	1-methyl-pyrrolidine (1)
OH_hydroxyl_aliphatic	OH without connection to a ring	1-hexanol (1)
OH_hydroxyl_naphthenic	OH connected to at least one nonaromatic ring	cyclohexanol (1)
OH_hydroxyl_phenolic	OH connected to at least one aromatic ring	phenol (1)
O_ether_aliphatic	O without connection to a ring	dimethyl ether (2)
O_ether_naphthenic	O connected to at least one nonaromatic ring	cyclohexyl methyl ether (1)
O_ether_benzoic	O connected to at least one aromatic ring	cyclohexyl phenyl ether (1)
O_ring_ether	O in a nonaromatic ring	tetrahydrofuran (1)
O_ring_ether_furanic	O in an aromatic ring	furan (1)
O_peroxy_aliphatic	O connected to another O and an aliphatic C	dimethyl peroxide (2)
O_peroxy_naphthenic	O connected to another O and a nonaromatic ring ring	cyclohexyl methyl peroxide (1)
O_peroxy_benzoic	O connected to another O and an aromatic ring	cyclohexyl phenyl peroxide (1)
O_ester_aliphatic	O connected to a CO (or COH) and an aliphatic C	methyl acetate (1)
O_ester_naphthenic	O connected to a CO (or COH) and a nonaromatic ring	cyclohexyl acetate (1)

Table A1. continued

group name	description	examples (occurrences)
O_ester_benzoic	O connected to a CO (or COH) and an aliphatic C	phenyl acetate (1)
CHO_aldehyde_aliphatic	CHO without connection to a ring	hexanal (1)
CHO_aldehyde_naphthenic	CHO connected to at least one nonaromatic ring	cyclohexanecarbaldehyde (1)
CHO_aldehyde_benzoic	CHO connected to at least one aromatic ring	Benzaldehyde (1)
CO_carbonyl_aliphatic	CO without connection to a ring	2-hexanone (1)
CO_carbonyl_naphthenic	CO connected to at least one nonaromatic ring	cyclohexylethanone (1)
CO_carbonyl_benzoic	CO connected to at least one aromatic ring	phenylethanone (1)
COOH_carboxyl_aliphatic	COOH without connection to a ring	hexanoic acid (1)
COOH_carboxyl_naphthenic	COOH connected to at least one nonaromatic ring	cyclohexanecarboxylic acid (1)
COOH_carboxyl_benzoic	COOH connected to at least one aromatic ring	benzoic acid (1)
CO_ester_aliphatic	CO connected to a O and an aliphatic C	methyl acetate (1)
CO_ester_naphthenic	CO connected to a O and a nonaromatic ring	methyl cyclohexanecarboxylate (1)
CO_ester_benzoic	CO connected to a O and an aliphatic C	methyl benzoate (1)
CHO_ester	CO connected to only one O	phenyl formate (1)

Table A2. Structural Groups and Their Contributions for Estimation of Normal Boiling Point and Liquid Density at 20 °C

group name	group contribution		group name	group contribution	
	NBP (K)	density (cm <sup>3</sup> /mol)		NBP (K)	density (cm <sup>3</sup> /mol)
CH3_aliphatic	0.8758	32.14	NH_aliphatic	0.5954	8.58
CH2_aliphatic	0.3101	16.38	NH_naphthenic	0.4497	5.99
CH_aliphatic	-0.3343	-0.93	NH_benzoic	1.1877	3.80
C_aliphatic	-0.9261	-19.45	N_aliphatic	-0.3200	-4.81
CH2_naphthenic	0.3852	13.93	N_naphthenic	-0.3042	-4.81
CH_naphthenic_substituted	-0.2519	-1.16	N_benzoic	0.0221	-12.75
CH_naphthenic_condensed	-0.1343	-3.98	NH_ring_amine	0.6033	2.23
C_naphthenic_bisubstituted	-0.9841	-15.89	N_ring_amine	-0.2623	-6.38
C_naphthenic_cond_subs	-0.8246	-18.28	N_ring_pyridinic	0.6455	-2.30
CH2_olefinic	0.8639	29.70	NH_ring_pyrrolic	1.1069	-1.12
CH_olefinic	0.3232	13.55	N_ring_pyrrolic_subst	0.1913	-13.21
C_olefinic	-0.2574	-4.12	OH_hydroxyl_aliphatic	1.6224	11.73
C_diolefinic	0.4095	8.79	OH_hydroxyl_naphthenic	1.2554	5.11
CH_ringolefinic	0.3702	10.97	OH_hydroxyl_phenolic	1.5741	16.29
C_ringolefinic_condensed	-0.0502	-6.85	O_ether_aliphatic	0.3232	5.61
C_ringolefinic_substituted	-0.2255	-4.03	O_ether_naphthenic	0.2409	5.61
CH_triplebonded	0.8041	25.79	O_ether_benzoic	0.3936	3.07
C_triplebonded	0.3850	8.85	O_ring_ether	0.4963	0.69
OCH_aromatic	0.3814	11.22	O_ring_ether_furanic	0.2540	14.63
C_aromatic_substituted	-0.1494	-6.15	O_peroxy_aliphatic	0.1952	12.26
C_aromatic_condensedperipheral	0.0067	-7.74	O_peroxy_naphthenic	0.1952	12.26
C_aromatic_condensedinternal	-0.3995	-10.97	O_peroxy_benzoic	0.1952	9.72
C_aromatic_condensedperipheral_naphtene	0.0067	-7.74	O_ester_aliphatic	0.5522	12.74
C_aromatic_condensedinternal_naphtene	-0.3995	-10.97	O_ester_naphthenic	0.3205	13.38
C_aromatic_condensedinternal_binaphtene	0.0067	-7.74	O_ester_benzoic	0.7520	11.71
SH_mercaptan_aliphatic	1.5323	28.67	CHO_aldehyde_aliphatic	1.6035	25.29
SH_mercaptan_naphthenic	1.3995	29.27	CHO_aldehyde_naphthenic	1.5538	26.69
SH_mercaptan_thiophenolic	1.5191	27.55	CHO_aldehyde_benzoic	1.6330	25.97
S_sulfide_aliphatic	0.9281	12.53	CO_carbonyl_aliphatic	1.0150	10.09
S_sulfide_naphthenic	1.0168	9.81	CO_carbonyl_naphthenic	0.8815	10.09
S_sulfide_benzoic	0.8741	12.34	CO_carbonyl_benzoic	0.9556	9.38
S_ring_sulfide	1.2553	11.99	COOH_carboxyl_aliphatic	2.7724	26.95
S_ring_sulfide_thiophenic	0.7137	12.46	COOH_carboxyl_naphthenic	3.0154	26.95
S_disulfide_aliphatic	0.7424	14.53	COOH_carboxyl_benzoic	2.7024	26.95
S_disulfide_naphthenic	0.7154	9.81	CO_ester_aliphatic	0.4196	7.35
S_disulfide_benzoic	0.6883	12.34	CO_ester_naphthenic	0.4302	7.35
NH2_aliphatic	1.3653	18.07	CO_ester_benzoic	0.2877	9.18
NH2_naphthenic	1.1798	19.48	CHO_ester	1.0910	21.30
NH2_benzoic	1.7276	17.05			



Table A3. Estimation of the Normal Boiling Point of Pyrene (Experimental Value: 668.2 K)

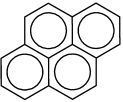
Component Name: Pyrene				
Group Name	Contribution	Occurrences	Value	Structure
CH_aromatic	0.3814	10	3.814	
C_aromatic_condensedperipheral	0.0067	4	0.0268	
C_aromatic_condensedinternal	-0.3995	2	-0.799	
$\Sigma (n_i \Delta T_{b,i}) =$			3.0418	
Correction Term				
4 Rings	5.5272	1	5.5272	
$T_b = 307.63 \ln(3.0418 + 5.5272) = 660.8 \text{ K}$				

Table A4. Estimation of the Normal Boiling Point of Cyclohexyl Phenyl Sulfide (Experimental Value: 561 K)

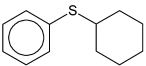
Component Name: Cyclohexyl phenyl sulfide				
Group Name	Contribution	Occurrences	Value	Structure
CH_aromatic	0.3814	5	1.9070	
C_aromatic_substituted	-0.1494	1	-0.1494	
CH2_naphthenic	0.3852	5	1.9260	
CH_naphthenic_substituted	-0.2519	1	-0.2519	
S_sulfide_benzoic	0.8740	1	0.8741	
$\Sigma (n_i \Delta T_{b,i}) =$			4.3058	
Correction Term				
2 Ring	1.8496	1	1.8496	
$T_b = 307.63 \ln(4.3058 + 1.8496) = 559.1 \text{ K}$				

Table A5. Estimation of the Normal Boiling point of 1,1-Dimethoxyethane (Experimental Value: 337.5 K)


Group Name	Contribution	Occurrences	Value	Structure
CH3_aliphatic	0.8758	3	2.6274	
CH_aliphatic	-0.3343	1	-0.3343	
O_ether_aliphatic	0.3232	2	0.6464	
$\Sigma (n_i \Delta T_{b,i}) =$			2.9395	
$T_b = 307.63 \ln(2.9394) = 331.7 \text{ K}$				

Table A6. Estimation of the normal boiling point of 1-ethylindene (Experimental value: 499.2 K)

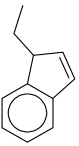
Component Name: 1-ethylindene				
Group Name	Contribution	Occurrences	Value	Structure
CH_aromatic	0.3814	4	1.5256	
C_aromatic_condensedperipheral_naphtene	0.0067	2	0.0134	
CH_ringolefinic	0.3702	2	0.7404	
CH_naphthenic_substituted	-0.2519	1	-0.2519	
CH3_aliphatic	0.8758	1	0.8758	
CH2_aliphatic	0.3101	1	0.3101	
$\Sigma (n_i \Delta T_{b,i}) =$			3.2134	
Correction Term				
2 Rings	1.8496	1	1.8496	
$T_b = 307.63 \ln(3.2134 + 1.8496) = 499.0 \text{ K}$				

Table A7. Estimation of the normal boiling point of N-Methyl aniline (Experimental value: 469.02 K)


Component Name: N-Methyl aniline				
Group Name	Contribution	Occurrences	Value	Structure
CH_aromatic	0.3814	5	1.9070	
C_aromatic_substituted	-0.1494	1	-0.1494	
NH_benzoic	1.18769	1	1.1877	
CH3_aliphatic	0.8758	1	0.8758	
$\Sigma (n_i \Delta T_{b,i}) =$			3.8211	
Correction Term				
1 Ring	0.6963	1	0.6963	
$T_b = 307.63 \ln(3.8211 + 0.6963) = 463.9 \text{ K}$				

Table A8. Estimation of the liquid density at 20°C of Pyrene (Experimental value: 1.270 cm<sup>3</sup>/g)


Component Name: Pyrene, MW = 202.26 g/mol				
Group Name	Contribution	Occurrences	Value	Structure
CH_aromatic	11.22	10	112.20	
C_aromatic_condensedperipheral	-7.74	4	-30.96	
C_aromatic_condensedinternal	-10.97	2	-21.94	
$\Sigma (n_i \Delta V_{m,i}) =$			59.30	
Correction Term				
4 Rings	100	1	100.00	
$d = \frac{202.26}{59.30 + 100} = 1.2697 \text{ cm}^3/\text{g}$				

Table A9. Estimation of the liquid density at 20°C of cyclohexyl phenyl sulfide (Experimental value: 1.036 cm<sup>3</sup>/g)

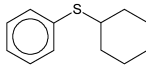

Component Name: Cyclohexyl phenyl sulfide MW = 192.32 g/mol				
Group Name	Contribution	Occurrences	Value	Structure
CH_aromatic	11.22	5	56.10	
C_aromatic_substituted	-6.15	1	-6.15	
CH2_naphthenic	13.93	5	69.65	
CH_naphthenic_substituted	-1.16	1	-1.16	
S_sulfide_benzoic	12.34	1	12.34	
$\Sigma (n_i \Delta T_{b,i}) =$			130.78	
Correction Term				
2 Ring	50	1	50	
$d = \frac{192.32}{130.78 + 50} = 1.0638 \text{ cm}^3/\text{g}$				

Table A10. Estimation of the liquid density at 20°C of 1,1-Dimethoxyethane (Experimental value: 0.8528 cm<sup>3</sup>/g)

Component Name: 1,1-Dimethoxyethane, MW = 90.12 g/mol				
Group Name	Contribution	Occurrences	Value	Structure
CH3_aliphatic	32.14	3	96.42	
CH_aliphatic	-0.93	1	-0.93	
O_ether_aliphatic	5.61	2	11.22	
$\Sigma (n_i \Delta V_{m,i}) =$			106.71	
$d = \frac{90.12}{106.71} = 0.8445 \text{ cm}^3/\text{g}$				

the molecule (g/mol) and  $F_{V_m}$  is a boiling point correction term that can be calculated by following equation:

$$F_{V_m} = 25N_R \quad (\text{A.4})$$

where  $N_R$  is the number of rings in the molecule.

The group contributions and the adjustable parameters have been determined by simultaneous linear regression. For this purpose, normal boiling point and liquid density at 20 °C for approximately 1850 components ranging from C5 to C42 were taken from the DIPPR data bank. The group contributions for these properties are listed in the Table A2.

In a number of cases, the choice of the correct group is not obvious using only its description. For example, N-methylcyclohexylamine (see Table A7) the  $-\text{NH}-$  could be described as  $\text{NH}_{\text{aliphatic}}$  or  $\text{NH}_{\text{naphthenic}}$ . In order to avoid ambiguous descriptions, a priority rule was defined between aliphatic, naphthenic and aromatic groups. An aromatic group has priority over the naphthenic and aliphatic groups and the naphthenic group has priority over the aliphatic group. Hence, in case of the N-methylcyclohexylamine,  $-\text{NH}-$  must be described as  $\text{NH}_{\text{naphthenic}}$ , given that the naphthenic group is more important than the aliphatic group.

To illustrate the application of the proposed method, a detailed procedure for the estimation of normal boiling point and liquid density at 20 °C is given in Tables A3–A7 and Tables A8–A12, respectively.

Table A11. Estimation of the liquid density at 20°C of 1-ethylindene (Experimental value: 0.9660 cm<sup>3</sup>/g)

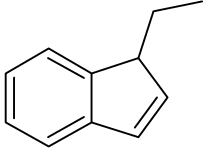

Component Name: 1-ethylindene, MW = 144.21 g/mol				
Group Name	Contribution	Occurrences	Value	Structure
CH_aromatic	11.22	4	44.88	
C_aromatic_condensedperipheral_naphtene	-7.74	2	-15.48	
CH_ringolefinic	10.97	2	21.94	
CH_naphthenic_substituted	-1.16	1	-1.16	
CH3_aliphatic	32.14	1	32.14	
CH2_aliphatic	16.38	1	16.38	
$\sum (n_i \Delta V_{m,i}) =$			98.7	
Correction Term				
2 Rings	50.00	1	50.00	
$d = \frac{144.21}{98.70 + 50} = 0.9698 \text{ cm}^3/\text{g}$				

Table A12. Estimation of the liquid density at 20°C of N-methyl aniline (Experimental value: 0.9861 cm<sup>3</sup>/g)

Component Name: N-Methyl aniline, MW = 107.15 g/mol				
Group Name	Contribution	Occurrences	Value	Structure
CH_aromatic	11.22	5	56.10	
C_aromatic_substituted	-6.15	1	-6.15	
NH_benzoic	3.80	1	3.80	
CH3_aliphatic	32.14	1	32.14	
$\sum (n_i \Delta V_{m,i}) =$			85.89	
Correction Term				
1 Ring	25.00	1	25.00	
$d = \frac{107.15}{85.89 + 25} = 0.9663 \text{ cm}^3/\text{g}$				

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail: jan.verstraete@ifpen.fr.

### Notes

The authors declare no competing financial interest.

## ABBREVIATIONS

ATP = attached proton test  
 ASTM = American Standard for Testing Materials  
 GC = gas chromatography  
 LC = liquid chromatography  
 LCO = light cycle oil  
 MS = mass spectrometry  
 NMR = nuclear magnetic resonance  
 PDF = probability distribution function  
 REM = reconstruction by entropy maximization  
 RFCC = resid fluid catalytic cracking  
 SARA = saturates, aromatics, resins, asphaltenes  
 SGA = structural group analysis  
 SOL = structure oriented lumping  
 SR = stochastic reconstruction  
 SR-REM = coupled sr-rem method  
 VR = vacuum residue

### Nomenclature

$E$  = Shannon entropy  
 $F$  = Objective function  
 $f_j$  = value of constraint  $j$  in the information entropy criterion  
 $f_{ij}$  = property or coefficient of molecule  $i$  for constraint  $j$  in the information entropy criterion

$H$  = information entropy criterion

$J$  = total number of equality constraints in the information entropy criterion

$N$  = number of molecules present in the predefined set of molecules

$N_A$  = number of analyses present in the objective function

$N_{M,j}$  = number of measurements of analysis  $j$

$X_{k,j}^{\text{exp}}$  = experimental value of the measure  $k$  of analysis  $j$

$X_{k,j}^{\text{calc}}$  = calculated value of the measure  $k$  of the analysis  $j$

$x_i$  = mole fraction of molecule  $i$

$W_j$  = weighting factor of analysis  $j$

$\delta_j$  = relative deviation between the calculated and experimental values of analysis  $j$

$\lambda_j$  = Lagrange multiplier associated with constraint  $j$  in the information entropy criterion

$\mu$  = Lagrange multiplier associated with the mass balance constraint

## REFERENCES

- (1) *Heavy Crude Oil—from Geology to Upgrading*; Huc, A.-Y., Ed.; Editions TECHNIP: Paris, 2011.
- (2) Quann, R. F.; Jaffe, S. B. Structure oriented lumping: Describing the chemistry of complex hydrocarbon mixtures. *Ind. Eng. Chem. Res.* **1992**, *31*, 2483.
- (3) Quann, R. F.; Jaffe, S. B. Building useful models of complex reaction systems in petroleum refining. *Chem. Eng. Sci.* **1996**, *51*, 1615.
- (4) Jaffe, S. B.; Freund, H.; Olmstead, W. N. Extension of structure-oriented lumping to vacuum residua. *Ind. Eng. Chem. Res.* **2005**, *44*, 9840.
- (5) Vynckier, E.; Froment, G. F. Modelling of the kinetics of complex processes based upon elementary steps. *Kinetic and Thermodynamic Lumping of Multicomponent Mixtures*; Astarita, G., Sandler, S. I., Eds.; Elsevier: Amsterdam, 1991; pp 131–161.
- (6) Allen, D. T.; Liguras, D. K. Structural models of catalytic cracking chemistry: A case study of a group contribution approach to lumped kinetic modeling. *Chemical Reactions in Complex Mixtures*; Sapre, A. V., Krambeck, F. J., Eds.; Van Nostrand Reinhold: New York, 1991; pp 102–125.
- (7) Hudebine, D. Reconstruction moléculaire de coupes pétrolières. PhD Thesis. École Normale Supérieure de Lyon, 2003.
- (8) Hirsch, E.; Altgelt, K. H. Integrated structural analysis—A method for the determination of average structural parameters of petroleum heavy ends. *Anal. Chem.* **1970**, *42*, 1330.

- (9) Speight, J. G. *The Chemistry and Technology of Petroleum*; Marcel Dekker, Inc.: New York, 1999.
- (10) Takegami, Y.; Watanabe, Y.; Suzuki, T.; Mitsudo, T.; Itoh, M. Structural investigation on column-chromatographed vacuum residues of various petroleum crudes by  $^{13}\text{C}$  nuclear magnetic resonance spectroscopy. *Fuel* **1980**, *59*, 253.
- (11) Suzuki, T.; Itoh, M.; Takegami, Y.; Watanabe, Y. Chemical structure of tar-sand bitumens by  $^{13}\text{C}$  and  $^1\text{H}$  NMR spectroscopic methods. *Fuel* **1982**, *61*, 402.
- (12) Ali, L. H.; Al-Ghannam, K. A.; Al-Rawi, J. M. Chemical structure of asphaltene in heavy crude oils investigated by NMR. *Fuel* **1990**, *69*, 519.
- (13) Ali, F. A.; Chaloum, N.; Hauser, A. Structure representation of asphaltene GPC fractions derived from Kuwaiti residual oils. *Energy Fuels* **2006**, *20*, 231.
- (14) Sato, S. The development of a support program for the analysis of average molecular structures by personal computer. *Sekiyu Gakkaishi* **1997**, *40*, 46.
- (15) Artok, L.; Su, Y.; Hirose, Y.; Hosokawa, M.; Murata, S.; Nomura, M. Structure and reactivity of petroleum-derived asphaltene. *Energy Fuels* **1999**, *13*, 287.
- (16) Gauthier, T.; Danial-Fortain, P.; Merdrignac, I.; Guibard, I.; Quoineaud, A.-A. Studies on the evolution of asphaltene structure during hydroconversion of petroleum residues. *Catal. Today* **2008**, *130*, 429.
- (17) Faulon, J. L.; Vandenbroucke, M.; Drappier, J. M.; Behar, F.; Romero, M. Modélisation des Structures Chimiques des Macromolécules Sédimentaires: le logiciel XMOL. *Rev. Inst. Fr. Pet.* **1990**, *45*, 161.
- (18) Khorasheh, F.; Gray, M. R.; Dalla Lana, I. G. Structural analysis of Alberta heavy gas oils. *Fuel* **1987**, *66*, 505.
- (19) Khorasheh, F.; Khaleli, R.; Gray, M. R. Computer generation of representative molecules for heavy hydrocarbon mixtures. *Fuel* **1998**, *77*, 241.
- (20) Zhang, Y. A molecular approach for characterization and property predictions of petroleum mixtures with application to refinery modelling. PhD Thesis. University of Manchester, 1999.
- (21) Gomez-Prado, J.; Zhang, N.; Theodoropoulos, C. Characterisation of heavy petroleum fractions using modified molecular-type homologous series (MTHS) representation. *Energy* **2008**, *33*, 974–987.
- (22) Ahmad, M. I.; Zhang, N.; Jobson, M. Molecular components-based representation of petroleum fractions. *Chem. Eng. Res. Des.* **2011**, *89* (4), 410–420.
- (23) López-García, C.; Hudebine, D.; Schweitzer, J.-M.; Verstraete, J. J.; Ferré, D. In-depth modeling of gas oil hydrotreating: From feedstock reconstruction to reactor stability analysis. *Catal. Today* **2010**, *150*, 279–299.
- (24) Hudebine, D.; Verstraete, J. J.; Chapus, T. Statistical reconstruction of gas oil cuts. *Oil Gas Sci. Technol.* **2011**, *66* (3), 461–477.
- (25) Pyl, S. P.; Hou, Z.; Van Geem, K. M.; Reyniers, M.-F.; Marin, G. B.; Klein, M. T. Modeling the composition of crude oil fractions using constrained homologous series. *Ind. Eng. Chem. Res.* **2011**, *50* (18), 10850–10858.
- (26) Neurock, M. A computational chemical reaction engineering analysis of complex heavy hydrocarbon reaction systems. PhD Thesis. University of Delaware, 1992.
- (27) Neurock, M.; Nigam, A.; Trauth, D. M.; Klein, M. T. Molecular representation of complex hydrocarbon feedstocks through efficient characterization and stochastic algorithms. *Chem. Eng. Sci.* **1994**, *49*, 4153–4177.
- (28) Trauth, D. M. Structure of complex mixtures through characterization, reaction, and modeling. PhD Thesis. University of Delaware, 1993.
- (29) Trauth, D. M.; Stark, S. M.; Petti, T. F.; Neurock, M.; Klein, M. T. Representation of the molecular structure of petroleum resid through characterization and Monte Carlo modeling. *Fuel* **1994**, *8*, 576–580.
- (30) Campbell, D. M. Modeling of structure and reaction in hydrocarbon conversion. PhD Thesis. University of Delaware, 1998.
- (31) Campbell, D. M.; Klein, M. T. Construction of a molecular representation of a complex feedstock by Monte Carlo and quadrature methods. *Appl. Catal., A* **1997**, *160*, 41.
- (32) Campbell, D. M.; Bennett, C.; Hou, Z.; Klein, M. T. Attribute-based modeling of resid structure and reaction. *Ind. Eng. Chem. Res.* **2009**, *48* (4), 1683–1693.
- (33) Sheremata, J. M.; Gray, M. R.; Dettman, H. D.; McCaffrey, W. C. Quantitative molecular representation and sequential optimization of athabasca asphaltene. *Energy Fuels* **2004**, *18*, 1377.
- (34) Boek, E. S.; Yakovlev, D. S.; Headon, F. H. Quantitative molecular representation of asphaltene and molecular dynamics simulation of their aggregation. *Energy Fuels* **2009**, *23*, 1209.
- (35) Hudebine, D.; Vera, C.; Wahl, F.; Verstraete, J. Molecular representation of hydrocarbon mixtures from overall petroleum analyses. *AIChE Spring Meeting* **2002**, 10–14.
- (36) Hudebine, D.; Verstraete, J. Molecular reconstruction of LCO gas oils from overall petroleum analyses. *Chem. Eng. Sci.* **2004**, *59*, 4755.
- (37) Pereira de Oliveira, L.; Verstraete, J. J.; Kolb, M.; Monte Carlo, A. Modeling methodology for the simulation of hydrotreating processes. *Chem. Eng. J.* **2012**, 207–208, 94–102.
- (38) Verstraete, J.; Revellin, N.; Dulot, H.; Hudebine, D. Molecular reconstruction of vacuum gas oils. *Prepr. Pap.—Am. Chem. Soc., Div. Fuel Chem.* **2004**, *49*, 20–21.
- (39) Revellin, N., Modélisation de l'hydrotraitement des distillats sous vide. PhD Thesis. École Normale Supérieure de Lyon, 2006.
- (40) Charon-Revellin, N.; Dulot, H.; López-García, C.; Jose, J. Kinetic modeling of vacuum gas oil hydrotreatment using a molecular reconstruction approach. *Oil Gas Sci. Technol.* **2011**, *66*, 479–490.
- (41) Verstraete, J.; Schnongs, P.; Dulot, H.; Hudebine, D. Molecular reconstruction of heavy petroleum residue fractions. *Chem. Eng. Sci.* **2010**, *65*, 304.
- (42) Boduszynski, M. M. Composition of petroleum residua. *Prepr. Pap.—Am. Chem. Soc., Div. Pet. Chem.* **2002**, *47* (4), 329–331 ACS.
- (43) Boduszynski, M. M. Composition of heavy petroleum 0.2. Molecular Characterization. *Energy Fuels* **1988**, *2*, 597.
- (44) Boduszynski, M. M. Composition of heavy petroleum 0.1. Molecular weight, hydrogen deficiency, and heteroatom concentration as a function of atmospheric equivalent boiling-point up to 1400 °F (760 °C). *Energy Fuels* **1987**, *1*, 2.
- (45) Sheu, E. Y. Petroleum asphaltene properties, characterization, and issues. *Energy Fuels* **2002**, *16*, 74.
- (46) Durand, E.; Clemancey, M.; Lancelin, J.-M.; Verstraete, J.; Spinat, D.; Quoineaud, A.-A. Effect of chemical composition on asphaltene aggregation. *Energy Fuels* **2010**, *24*, 1051.
- (47) Marques, J.; Maget, S.; Verstraete, J. J. Improvement of ebullated-bed effluent stability at high conversion operation. *Energy Fuels* **2011**, *25* (9), 3867–3874.
- (48) API. API procedure 2B2.1 for estimating the molecular weight of a petroleum fraction. *API Technical Handbook*; American Petroleum Institute: Washington, DC, 1987.
- (49) Murgich, J.; Abanero, J. A.; Strausz, O. P. Molecular recognition in aggregates formed by asphaltene and resin molecules from the Athabasca oil sand. *Energy Fuels* **1999**, *13*, 278.
- (50) Altgelt, K. H.; Boduszynski, M. M. Composition and analysis of heavy petroleum fractions; Marcel Dekker, Inc.: New York, 1994.
- (51) Miyabayashi, K.; Naito, Y.; Tsujimoto, K.; Miyake, M. Structure characterization of petroleum residuum components fractionated by high vacuum short-path distillation (DISTACT) and gel permeation chromatography (GPC). *Prepr. Pap.—Am. Chem. Soc., Div. Fuel Chem.* **1995**, *40* (3), 497–503.
- (52) Mullins, O. C.; Sheu, E. Y. *Structures and Dynamics of Asphaltenes*; Springer: New York, 1999.
- (53) Groenzin, H.; Mullins, O. C. Molecular size and structure of asphaltene from various sources. *Energy Fuels* **2000**, *14*, 677.
- (54) Wiehe, I. A. The pendant-core building block model of petroleum residua. *Energy Fuels* **1994**, *8*, 536.
- (55) Wiehe, I. A. *Process Chemistry of Petroleum Macromolecules*; CRC Press: Boca Raton, FL, 2008.

- (56) Gray, M. R. Consistency of asphaltene chemical structures with pyrolysis and coking behavior. *Energy Fuels* **2003**, *17*, 1566.
- (57) De Jong, K. An analysis of the behavior of a class of genetic adaptive systems. PhD Thesis. University of Michigan, Ann Arbor, MI, 1975.
- (58) Goldberg, D. E. *Genetic Algorithms in Search Optimization and Machine Learning*; Addison-Wesley: Boston, 1989.
- (59) Goldberg, D. E.; Deb, K.; Clark, J. H. Genetic algorithms, noise, and the sizing of populations. *Complex Syst.* **1992**, *6*, 333–362.
- (60) Van Geem, K. M.; Hudebine, D.; Reyniers, M. F.; Wahl, F.; Verstraete, J. J.; Marin, G. B. Molecular reconstruction of naphtha steam cracking feedstocks based on commercial indices. *Comput. Chem. Eng.* **2007**, *31*, 1020.
- (61) Hudebine, D.; Verstraete, J. J. Reconstruction of petroleum feedstocks by entropy maximization: Application to FCC gasolines. *Oil Gas Sci. Technol.* **2011**, *66*, 437–460.
- (62) Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*.
- (63) Petti, T. F.; Trauth, D. M.; Scott, M. S.; Neurock, M.; Yasar, M.; Klein, M. T. CPU issues in the representation of the molecular structure of petroleum resid through characterization, reaction, and Monte Carlo modeling. *Energy Fuels* **1994**, *8*, 570.
- (64) Constantinou, L.; Gani, R. New group contribution method for estimating properties of pure compound. *AIChE J.* **1994**, *40*, 1697.