MDPI

Article

Study on the Domain of Parameters in the Stochastic Reconstruction Method

Guangyao Zhao, Minglei Yang * and Feng Qian *

Key Laboratory of Smart Manufacturing in Energy Chemical Process, Ministry of Education, East China University of Science and Technology, 130 Meilong Rd, Shanghai 200237, China; y10160200@mail.ecust.edu.cn
* Correspondence: mlyang@ecust.edu.cn (M.Y.); fqian@ecust.edu.cn (F.Q.)

Abstract: In the stochastic reconstruction of petroleum fractions, the domain configuration for parameters is of great significance to the performance of the model in terms of accuracy and convergence. Based on the stochastic reconstruction model of gas oils, this work investigates the influence of different domains on the simulating accuracy, the convergence rate and the ability to predict detailed composition. In this paper, the parameters to be optimized in the model include histogram distributions and gamma distributions, which are used to represent the structural attributes. In each histogram distribution, the parameters are real numbers between zero and one that should increase progressively. The mean and shape parameters are used as the tuning parameters for gamma distribution. As for the mean, the lower and upper boundaries are configured as fixed values that depend on the predefined values of the structural attributes. For the shape parameter, multiple cases are considered for its upper boundary, which is set to 5, 10, 20, 50 and 100, respectively. The lower boundary is set to 0 or 1, depending on the structural attribute. The results indicate that the simulating accuracy of the stochastic reconstruction model improves with the increase of the upper boundary for the shape parameter. When the upper boundary of each shape parameter is 20, the simulating accuracy is better than that when the upper boundary is 5 and 10, and close to that when the upper boundary is 50 and 100. The detailed group-type analysis is involved in the model to predict the distributions of mass fraction in chemical families against the carbon number. The results show that the predicted distributions in normal paraffins, isoparaffins, naphthenes, and aromatics are closest to the experimental distributions when the upper boundary is 20 and 50. When the upper boundary is 5 and 10, the predicted distributions in normal paraffins and isoparaffins are lower and broader than the experimental distributions. On the contrary, the predicted distributions in normal paraffins and isoparaffins are higher when the upper boundary is 100. By studying the effects of different domains on the results of the stochastic reconstruction model, a better molecular characterization of petroleum fractions can be achieved.

Keywords: domain of parameters; stochastic reconstruction; gamma distribution; predicted distributions of mass fraction

---

1. Introduction

Petroleum fractions consist of tens of thousands of molecular species that contain carbon, hydrogen, sulfur, nitrogen, oxygen and metals. Due to the diverse origins of crude oils and refining technologies, the compositions of material streams in the refining processes vary widely. In recent years, with increasingly strict environmental regulations for fossil fuels as well as more and more heavy crude oils, it is urgent to improve refining technologies [1,2]. As the foundation for the study of refining technologies, the characterization of petroleum fractions and the construction of conversion mechanisms at the molecular level have become the most important issues in refineries [3,4].

Due to the limitation of analytical chemistry technologies, computer-aided molecular reconstruction is currently the most popular method to reflect the complicated compositions

of petroleum with limited properties [4]. Since the early 1990s when the first molecular reconstruction method was reported, a lot of molecular reconstruction methods have been proposed and applied in the molecular modeling of refining processes [5–8], including the stochastic reconstruction method (SR), structural-oriented lumping (SOL), molecular type homologous series (MTHS) method and SR-reconstruction by entropy maximization (SR-REM). Among them, the SR method is the most popular method in the characterization of heavy petroleum fractions.

Klein et al. [9,10] first introduced the SR method to characterize heavy residue feedstocks. In the SR method, molecules are treated as the collection of structural attributes (SA), and a probability density function (PDF) is imposed on each structural attribute. The type of PDF varies for different structural attributes, such as the histogram distribution for the determination of molecular families and the gamma distribution for ring numbers and the length of chains. Monte Carlo sampling with a quadrature method is applied to generate an equimolar set of molecules from PDFs. The parameters for each PDF were adjusted in an optimization loop for simulated annealing or genetic algorithm to make the bulk properties of generated mixtures close to those of the actual samples.

Many studies that use the SR method have been reported. Petti et al. examined [11] the usage of CPU resources in the SR method and suggested that a sample size of 10,000 molecules could balance the simulation accuracy and computation expense. Zhang et al. [12] extended a novel SR model to heavy vacuum residue fractions. The residue molecules were treated as a combination of approximately 600 building substructures. Deniz et al. [13] introduced a new structure parameter set for detailed ring and chain configurations into the SR method to improve the method performance in heavy petroleum fractions. Moreover, Deniz et al. investigated [14] the effects of methods for estimating the boiling point temperature and density of pure compounds on the simulating accuracy of the SR method. It is observed that the SR method has the highest accuracy with the group contribution method by Gani [15–18] and the Yen-woods equations [19]. Haktanlr et al. [20] proposed a novel SR method based on a custom predefined molecular library. This novel method focuses on characterizing the petroleum fractions with exhausted molecular species. Meanwhile, a sieving mechanism is introduced to make sure that the generated molecules are reasonable in structure. Glazov et al. studied [21] the relationships between different PDFs of structural attributes and bulk properties.

The general expression of the SR method is shown as below:

$$\begin{aligned}
&\text{E} = \min f[\boldsymbol{x}, \boldsymbol{\xi}] \\
&\boldsymbol{x}^{low} \leq \boldsymbol{x} \leq \boldsymbol{x}^{upper} \\
&\boldsymbol{x} \in \mathbb{R}
\end{aligned} \tag{1}$$

where $\boldsymbol{x}$ stands for the parameters to be optimized, $\boldsymbol{\xi}$ are series of uniform random numbers between 0 and 1. $\boldsymbol{x}^{low}$ and $\boldsymbol{x}^{upper}$ are the lower and upper boundaries of $\boldsymbol{x}$. The function f stands for the gap between the experimental values of bulk properties and the predicted values. The optimal parameters that make the predicted values of bulk properties close enough to the experimental values are obtained by effective algorithm optimization.

In the SR method, the input parameters $\boldsymbol{x}$ are the parameters of PDFs. The structural attributes generally contain the type of molecule, the number of naphthenic rings and aromatic rings, the length of paraffin chains and sidechains, the type of heteroatom-containing molecule, etc. The types of PDF used to represent them are histogram distribution and gamma distribution. The parameters in histogram distribution are real numbers between 0 and 1. If more than one parameter exists in a histogram distribution, they should increase progressively. Given the total number of parameters in the model and the simulating accuracy, the two-parameter gamma distribution is generally adopted in the SR method. The two parameters are the shape parameter (SP) and scale parameter, or the shape parameter and mean. To characterize the complicated and diverse compositions in petroleum fractions, it is crucial to reasonably configure the lower and upper boundaries of the parameters, which is the domain of parameters in the SR method. However, as far as the authors

know, no study has been reported on this issue. To this point, this paper aims to study the determination of domain and its effects on the performance of the SR method.

Section 2 illustrates the SR model based on gas oils. The bulk properties by experiments, the setting of structural attributes and the building diagram are provided in Section 2.1. Section 2.2 provides the configuration of the lower and upper boundaries for the parameters in the histogram distributions and gamma distributions. Section 3 provides the results and discussions. Finally, the conclusion is given in Section 4.

2. Methodology

Figure 1 shows the framework of the SR model adopted in this paper. The model starts with initializing the input parameters, which are the tuning parameters in PDFs. Input parameters are generated based on their lower and upper boundaries. For parameters in histogram distributions, the monotonicity constraint, that parameters in the same histogram distribution should increase progressively, should be satisfied. With the input parameters, Monte Carlo sampling is used to generate pseudo molecules, the properties of which are estimated by group contribution methods and empirical correlations. The bulk properties of pseudo mixtures are calculated by mixing rules. The objective function is configured to evaluate the input parameters by counting the gap between the experimental values and the estimated values of the bulk properties. If the objective function value meets the stop criterion, the optimal parameters and pseudo mixture are obtained. Otherwise, new input parameters are generated in the optimization algorithm and evaluated until all constraints are satisfied.

The setting of lower and upper boundaries for input parameters determines the values of the input parameters in the SR model. Multiple cases of boundaries for input parameters are tested and compared based on an SR model of gas oils.
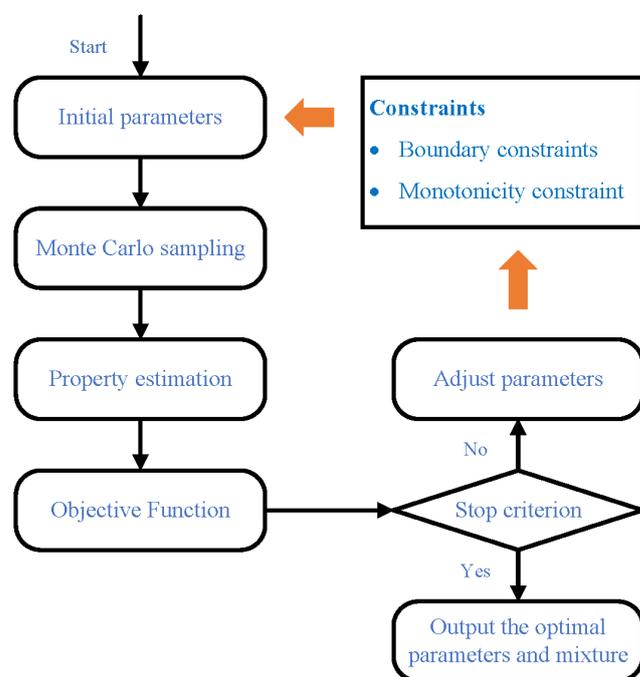

Figure 1. The framework of the SR model.

2.1. The SR Model of Gas Oils

The two gas oil samples used in this paper are from the work by Aleksandar [22], including a gas oil sample and a vacuum gas oil sample. As listed in Table 1, the bulk properties adopted contain the elemental analysis, density, H/C ratio, PINA analysis, aromatic sulfur content and simulated distillation.

Table 1. The bulk properties of gas oils.

Gas Oil A	Value	Gas Oil B	Value
Elemental analysis/wt%		Elemental analysis/wt%	
C	87.4	C	85.7
H	12.5	H	13.1
S	0.14	S	1.3
Density/kg·m^{-3}	0.862	Density/kg·m^{-3}	0.83
H/C ration/mol·mol^{-1}	1.77	Aromatic sulfur content/wt%	0.15
PINA analysis/wt%		PINA/wt%	
P	17.3	P	24.3
I	21.3	I	29.7
N	17.3	N	13.5
A	44.1	A	32.5
ASTM-D2887/K		ASTM-D2887/K	
0.5%	399	0.5%	409
10%	550	10%	535
30%	595	30%	576
50%	625	50%	602
70%	650	70%	622
90%	684	90%	655
99.5%	775	99.5%	705

Based on the molecular characteristics of two gas oil samples, the probability distributions containing the histogram distribution and gamma distribution are adopted to represent structural attributes. As shown in Table 2, the structural attributes represented by the histogram distribution are the type of molecule, the acceptance of a naphthenic ring on an aromatic ring, the acceptance of a branch on a paraffin chain, and the type of sulfur-containing compound. The gamma distribution is adopted for other structural attributes, such as the number of naphthenic rings, the number of aromatic rings, the length of paraffin chain, the length of sidechain, and the length of sulfur-containing chain.

Table 2. The structural attributes and PDFs.

Index	SA	The Type of PDFs	Parameters
1	The type of molecule	Histogram	x_1, x_2, x_3
2	If accept a naphthenic ring on the aromatic ring	Histogram	x_4
3	If the paraffin chain is branched	Histogram	x_5
4	The type of sulfur-containing molecule	Histogram	x_6, x_7
5	The number of naphthenic rings	Gamma	x_8, x_9
6	The number of aromatic rings	Gamma	x_{10}, x_{11}
7	The length of sidechain	Gamma	x_{12}, x_{13}
8	The length of paraffin chain	Gamma	x_{14}, x_{15}
9	The length of sulfur-containing chain	Gamma	x_{16}, x_{17}

Based on the structural attributes in Table 2, a building diagram for the construction of pseudo molecules is designed in Figure 2. In this building diagram, the first step is to determine the type of molecule. Paraffin, naphthene, aromatic and sulfur-containing compounds are considered in this work. If the type of molecule is paraffin, the next step is to determine whether the paraffin is branched. Then the length of paraffin is determined. If the type of molecule is naphthene, the number of naphthenic rings and the length of sidechain should be determined, in this order. When the type of molecule is aromatic, the number of aromatic rings should be determined first. If the number of aromatic rings is one or two, it should be determined whether a naphthenic ring is accepted. Lastly, the length of sidechain on aromatic rings is determined. There are three types of sulfur-containing compounds, including benzothiophene, dibenzothiophene, and sulfur-containing paraffin. Once the molecule is benzothiophene or dibenzothiophene, the length of sidechain should

be determined. Otherwise, the length of sulfur-containing chain is determined. In this work, the predefined range for the number of naphthenic rings in a core is one to three, and for the number of aromatic rings is one to four. The number of sidechains on a core is no more than one. The ranges for lengths of paraffin chains, sidechains and sulfur-containing chains are determined based on the distillation curve.
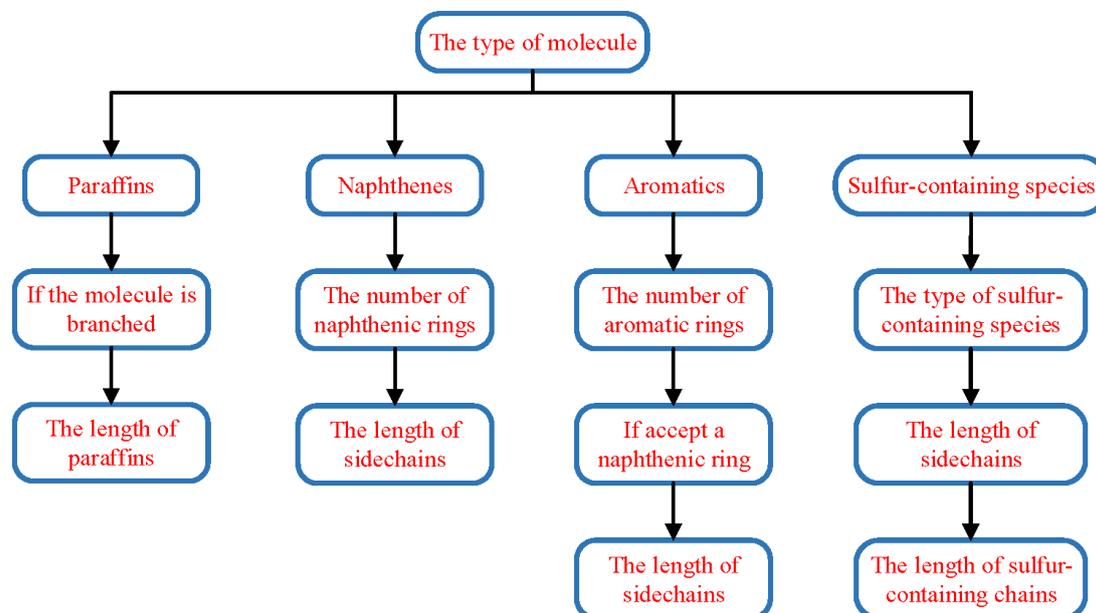

Figure 2. The building diagram in the SR model.

2.2. The Setting of Domains

For the parameters in the histogram distributions, their lower and upper boundaries are set to 0 and 1, respectively. According to Table 2, the parameters in the histogram distributions are x_1, x_2, x_3, x_4, x_5, x_6 and x_7. Because x_1, x_2, x_3 and x_6, x_7 are in a single histogram distribution, they should meet the constraint that parameters in the same histogram distribution should increase in order, as shown in Equation (2).

$$
\begin{aligned}
x_1 < x_2 < x_3 \\
x_6 < x_7 \\
x_1, x_2, \ldots, x_7 \in [0, 1]
\end{aligned}
\tag{2}
$$

The two-parameter gamma distribution is shown in Equations (3)–(5).

$$
p(x) = \frac{(x - r_{\min})^{\alpha - 1} e^{-\frac{(x - r_{\min})}{\beta}}}{\beta^{\alpha} \Gamma(\alpha)}
\tag{3}
$$

$$
\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt
\tag{4}
$$

$$
\beta = \frac{m - r_{\min}}{\alpha}
\tag{5}
$$

where α, β and $r_{\min}$ are the shape parameter, scale parameter and site parameter in the gamma distribution, respectively. In this work, the shape parameter α and mean m are adopted as the tuning parameters for each gamma distribution. The relationship between α, β, $r_{\min}$ and m is shown in Equation (5). $r_{\min}$ in each gamma distribution is a fixed value in this work, and depends on the predefined values of the corresponding structural attribute.

Before setting the boundaries of parameters in gamma distributions, the range of predefined values for structural attributes should be clarified. The range of predefined

values for the number of naphthenic rings and aromatic rings are 1–3 and 1–4 in both gas oils. According to the distillation curves in Table 1, the calculated ranges for the length of sidechains are 1–35 for gas oil A and 1–23 for gas oil B. The ranges for the length of paraffin chains are 8–41 and 8–30 for the two gas oil samples, respectively. Lastly, the ranges for the length of sulfur-containing chains are 6–40 for gas oil A and 6–29 for gas oil B. The r_{min} in each gamma distribution is equal to the minimum predefined value of the corresponding structural attribute minus 1.

In the gamma distributions, multiple cases for the boundaries of parameters are considered in Tables 3 and 4. There are five structural attributes that are represented by gamma distribution. The parameters in the model are from x_8 to x_{17}. x_8, x_{10}, x_{12}, x_{14} and x_{16} are the shape parameters of the gamma distributions for the number of naphthenic rings, the number of aromatic rings, the length of sidechains, the length of paraffin chains and the length of sulfur-containing chains, respectively. x_9, x_{11}, x_{13}, x_{15} and x_{17} are the mean of the gamma distributions. For both gas oils, the lower boundaries (LB) of x_8, x_{10} are 0, and the lower boundaries of x_{12}, x_{14} and x_{16} are 1. The upper boundaries (UB) of x_8, x_{10}, x_{12}, x_{14}, and x_{16} are given as 5, 10, 20, 50 and 100. As for the mean in each gamma distribution, the lower boundary is set to the minimum predefined value of the corresponding structural attribute minus 1, and the upper boundary is equal to the maximum predefined value of the corresponding structural attribute plus 1. For simplicity of presentation, the cases where the upper boundaries of the shape parameters are 5, 10, 20, 50 and 100 are marked below as case 1, case 2, case 3, case 4 and case 5, respectively.

Table 3. The boundaries of parameters in gamma distributions for gas oil A.

Index of SA	SP	LB	UB	Mean	LB	UB
5	x_8	0	5, 10, 20, 50, 100	x_9	0	4
6	x_{10}	0	5, 10, 20, 50, 100	x_{11}	0	5
7	x_{12}	1	5, 10, 20, 50, 100	x_{13}	0	36
8	x_{14}	1	5, 10, 20, 50, 100	x_{15}	7	42
9	x_{16}	1	5, 10, 20, 50, 100	x_{17}	5	41

Table 4. The boundaries of parameters in gamma distributions for gas oil B.

Index of SA	SP	LB	UB	Mean	LB	UB
5	x_8	0	5, 10, 20, 50, 100	x_9	0	4
6	x_{10}	0	5, 10, 20, 50, 100	x_{11}	0	5
7	x_{12}	1	5, 10, 20, 50, 100	x_{13}	0	24
8	x_{14}	1	5, 10, 20, 50, 100	x_{15}	7	31
9	x_{16}	1	5, 10, 20, 50, 100	x_{17}	5	30

3. Results and Discussions

3.1. Comparison of Simulating Performance in Different Cases

Tables 5 and 6 show the simulation results with the smallest relative error for two gas oils under different domains of the SR model. A fixed number of iterations that is specified as 2000 is adopted as the stop criterion of the SR model. It is observed that the relative errors are very small in all cases ranging from 0.03–0.06. This means that in the case of different domains, the SR model is able to generate pseudo mixtures, the bulk properties of which are very close to the experimental values (EV).

Table 5. Comparison between experimental values and estimated values of bulk properties in different cases for gas oil A.

Property	EV	UB of SP for Gas Oil A				
		5	10	20	50	100
Elemental analysis/wt%						
C	87.4	87.04	87.07	87.05	87.03	87.02
H	12.5	12.82	12.79	12.81	12.83	12.84
S	0.14	0.14	0.141	0.14	0.138	0.14
Density/kg·m^{-3}	0.862	0.868	0.864	0.863	0.861	0.862
H/C ratio/mol·mol^{-1}	1.77	1.77	1.76	1.77	1.77	1.77
PINA/wt%						
P	17.3	17.2	17.2	17.5	17.3	17.3
I	21.3	21.3	21.3	21.2	21.7	21.0
N	17.3	17.3	17.3	17.5	17.3	17.6
A	44.1	44.2	44.2	43.8	43.7	44.1
ASTM-D2887/K						
0.5%	399	392	406	406	406	406
10%	550	522	539	541	546	554
30%	595	588	600	602	605	607
50%	625	623	627	627	627	632
70%	650	661	655	651	650	651
90%	684	704	707	699	687	707
99.5%	775	754	773	767	754	773
The sum of relative errors		0.0441	0.0346	0.0332	0.0341	0.0335

Table 6. Comparison between experimental values and estimated values of bulk properties in different cases for gas oil B.

Property	EV	UB of SP for Gas Oil B				
		5	10	20	50	100
Elemental analysis/wt%						
C	85.7	85.2	85.2	85.6	85.0	85.0
H	13.1	13.5	13.5	13.2	13.7	13.7
S	1.3	1.34	1.31	1.28	1.28	1.3
Density/kg·m^{-3}	0.83	0.835	0.833	0.852	0.829	0.832
Aromatic sulfur content/wt%	0.15	0.15	0.15	0.15	0.15	0.15
PINA/wt%						
P	24.3	24.3	24.2	24.2	24.3	24.6
I	29.7	29.7	29.2	29.6	30.0	29.0
N	13.5	13.3	14.0	13.5	13.4	13.6
A	32.5	32.7	32.6	32.74	32.3	32.8
ASTM-D2887/K						
0.5%	409	403	403	403	403	403
10%	535	505	506	546	561	546
30%	576	554	557	582	582	568
50%	602	590	595	607	600	595
70%	622	627	629	627	631	633
90%	655	673	672	653	664	661
99.5%	705	710	709	694	705	701
The sum of relative errors		0.0591	0.0540	0.0523	0.0493	0.0511

　　Figures 3 and 4 compare the variations of the average objective function value from iterations 1–400 with different domains in the stochastic reconstruction of gas oil A. The average objective function value is calculated by 20 runs for each case. In Figure 3a, it is observed that at the early stage from iterations 1–200, the average objective function values in cases 1 and 2 are significantly lower than those in cases 3, 4 and 5. This means that in a

smaller domain the SR model is able to find parameters that can make the objective function value smaller at the beginning of the simulation. The variations in the average objective function value maintain the same trend for all cases between iterations 200–400. With the number of iterations exceeding 350, the average objective function value in case 1 starts to be larger than that in case 2. As the number of iterations increases from 400 to 1200, as shown in Figure 3c,d, the average objective function value in case 1 gradually exceeds the average objective function values in all other cases. Compared with the early stage of simulation, gaps of average objective function values among all cases are smallest when the number of iterations is around 1100. As the iteration continues, the average objective function value in case 2 gradually exceeds those in cases 3, 4 and 5. There are approximate rates in the decrease of average objective function values in case 3, 4 and 5. At the end of the iteration, it can be observed that average objective function values in case 3, 4 and 5 are very close, which are significantly smaller than those in case 1 and 2. This indicates that even though the average objective function values decrease faster in smaller domains at the beginning of the simulation, the SR model can find parameters that make the objective function value smaller with adequate iterations in larger domains. Meanwhile, for case 3, the decreasing rate of the average objective function value is larger than that in cases 4 and 5 at the early stage of simulation, and the final average objective function value is very close to that in cases 4 and 5, as shown in Figure 3f. This indicates that case 3, where the upper boundaries of shape parameters are 20, can not only make the SR model effectively find promising parameters, but also make sure the targeted objective function value is achieved.
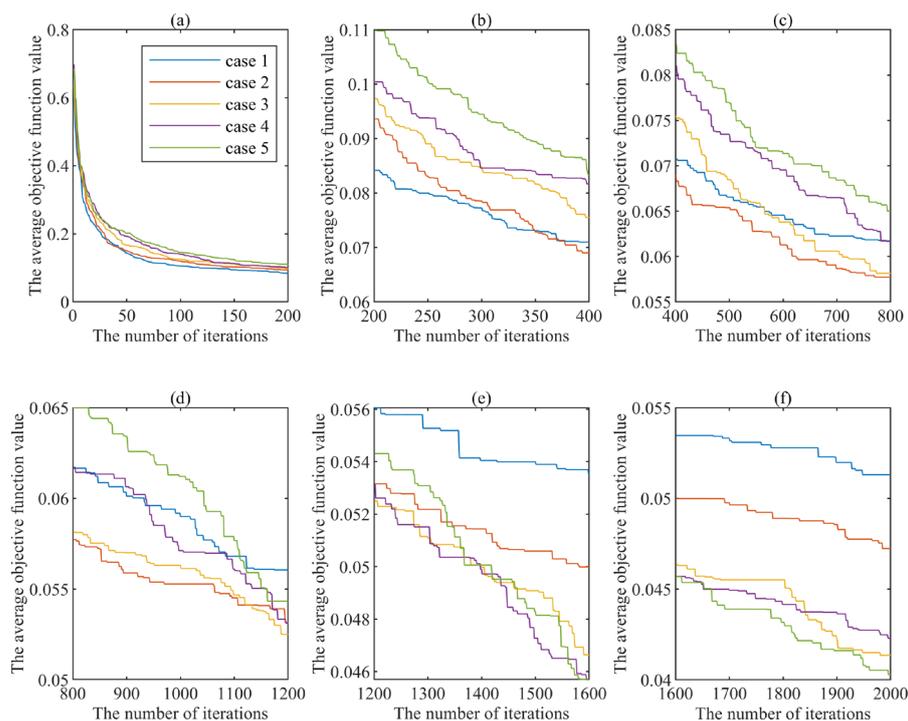

Figure 3. Variations in the average objective function value of 20 runs from iterations (**a**) 1–200, (**b**) 201–400, (**c**) 401–800, (**d**) 801–1200, (**e**) 1201–1600, (**f**) 1601–2000.

3.2. Prediction of Compositions by the SR Model in Different Domains

The predicted distributions of mass fraction against carbon number in normal paraffins, isoparaffins, naphthenes and aromatics are compared with experimental values in order to investigate the performance of the SR model on the prediction of detailed compositions in different domains. The representative predictions in each case are shown in Figure 4. It is observed that there are no cases where the predicted distributions in normal paraffins and isoparaffins fit well with the experimental distributions. The shapes of the predicted distributions in normal paraffins and isoparaffins are broader and lower than in

the experimental distributions. For naphthenes and aromatics, the predicted distributions are very close to the experimental distributions in cases 1, 2 and 3. As the upper boundaries of shape parameters increase, in cases 4 and 5, the shapes of the predicted distributions become higher. As a result, the upper boundaries of shape parameters in the gamma distributions should be less than 50 in order to fit well with the distributions of mass fraction in naphthenes and aromatics.

As shown in Tables 5 and 6 and in Figure 4, even though the estimated bulk properties in all cases are close to the experimental values, there are large deviations between the predicted distributions of mass fractions and the experimental values in all cases. In order to reduce the deviations between the predicted distributions and experimental distributions in all chemical families, the PINA values are replaced by the more the detailed experimental data from detailed group-type analysis.

The detailed group-type analysis data are mass fractions of homologous series in gas oil samples, which contain normal paraffins (P), isoparaffins (I), mononaphthenes (MN), dinaphthenes (DN), monoaromatics (MA), naphthenoaromatics (NA), diaromatics (DA), naphthenodiaromatics (NDA), triaromatics (TrA), naphthenotriaromatics (NTrA) and tetraaromatics (TeA). Based on detailed group-type analysis and other bulk properties, the stochastic reconstruction of gas oil A is implemented under different domains. As shown in Table 7, the predicted values of the bulk properties are also close to the experimental values in all cases, and the sums of relative errors are between 0.11 and 0.14.
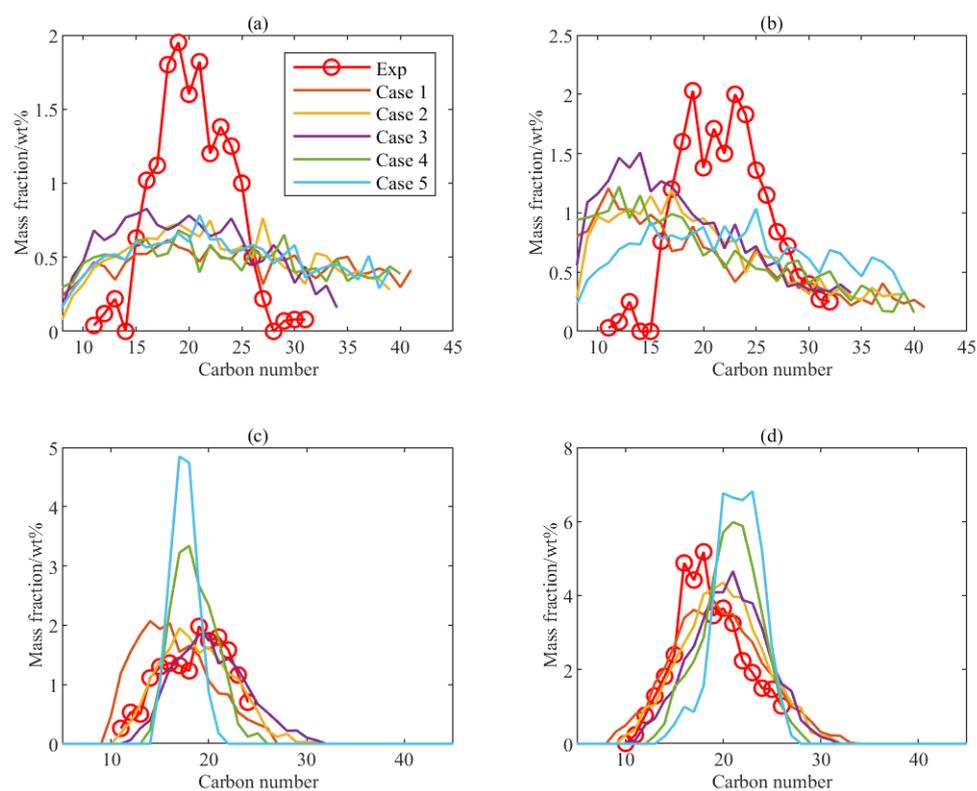

Figure 4. The distributions of mass fraction in (**a**) normal paraffins, (**b**) isoparaffins, (**c**) naphthenes, (**d**) aromatics for gas oil A by experiment and predictions.

Table 7. The comparison between experimental values and estimated values of bulk properties for gas oil A in different cases.

Property	EV	UB of SP for Gas Oil A				
		5	10	20	50	100
Elemental analysis/wt%						
C	87.4	87.1	87.1	87.1	87.0	87.06
H	12.5	12.8	12.8	12.8	12.8	12.8
S	0.14	0.138	0.141	0.140	0.141	0.138
Density/kg·m^{-3}	0.862	0.859	0.863	0.858	0.855	0.854
H/C ratio/mol·mol^{-1}	1.77	1.76	1.76	1.77	1.77	1.77
Detailed group-type analysis/wt%					-	-
P	17.00	17.45	17.28	17.76	17.51	17.34
I	21.20	21.21	21.12	21.71	21.06	21.37
MN	15.39	15.26	15.58	15.15	15.02	15.34
DN	2.20	2.07	2.19	2.25	2.24	2.20
MA	16.80	15.27	13.97	14.65	16.17	17.31
NA	7.40	8.96	8.56	9.14	9.17	8.07
DA	8.43	8.91	10.07	8.74	8.58	8.85
NDA	6.40	6.00	6.42	5.80	5.47	4.40
TrA	3.92	4.01	3.88	3.87	3.92	4.14
NTrA + TeA	0.74	0.86	0.93	0.93	0.86	0.98
ASTM-D2887/K						
0.5%	399	396	421	421	396	396
10%	550	503	522	521	507	503
30%	595	576	593	585	588	595
50%	625	629	635	624	629	629
70%	650	674	677	655	662	655
90%	684	729	727	692	699	687
99.5%	775	781	779	730	741	748
The sum of relative errors		0.1253	0.1340	0.1328	0.1170	0.1372

More accurate predictions on the distributions of mass fractions are observed with detailed group-type analysis. As shown in Figure 5, the domain range has a large influence on the predicted distributions. In normal paraffins and isoparaffins, the predicted distributions in cases 1 and 2, where the upper boundaries of the shape parameters are 5 and 10, are still lower and broader compared to the experimental values. In cases 3 and 4, the predicted distributions become higher than those in cases 1 and 2, and agree best with the experimental values. As the upper boundaries of the shape parameters increase, the predicted curves continue to be higher, and a significant deviation is observed in case 5. As for naphthenes and aromatics, similar predicted curves are shown in all cases. Small deviations between the prediction and experiment are observed. The predicted mass fractions of naphthenes are higher than the experimental values when the carbon number is small, and lower than the experimental values as the carbon number increases. In conclusion, the detailed group-type analysis can obviously improve the ability of the SR model to predict distributions of mass fraction in chemical families. Cases 3 and 4, where the upper boundaries of the shape parameters are 20 and 50, are the best cases for the prediction of detailed compositions of gas oil A.

The variation in predicted compositions in different domains is because of the mechanism of the SR method and the characteristic of gamma distribution. In the SR method, there are multiple combinations for the values of the parameters in the domain that make the estimated values of properties close to the experimental values. The gamma distribution curve tends to be lower and broader when the shape parameter is small, and tends to be higher and steeper as the shape parameter increases. As a result, with the experimental curves in this work, even though the target objective function value can be achieved in the small domain, there are large gaps between the predicted curves and the experimental

curves. When the value of the shape parameter is too large, for example, a value between 50 and 100 in this paper, an inappropriate curve can also be generated. Therefore, the appropriate prediction of detailed compositions depends on the reasonable configuration of domain.
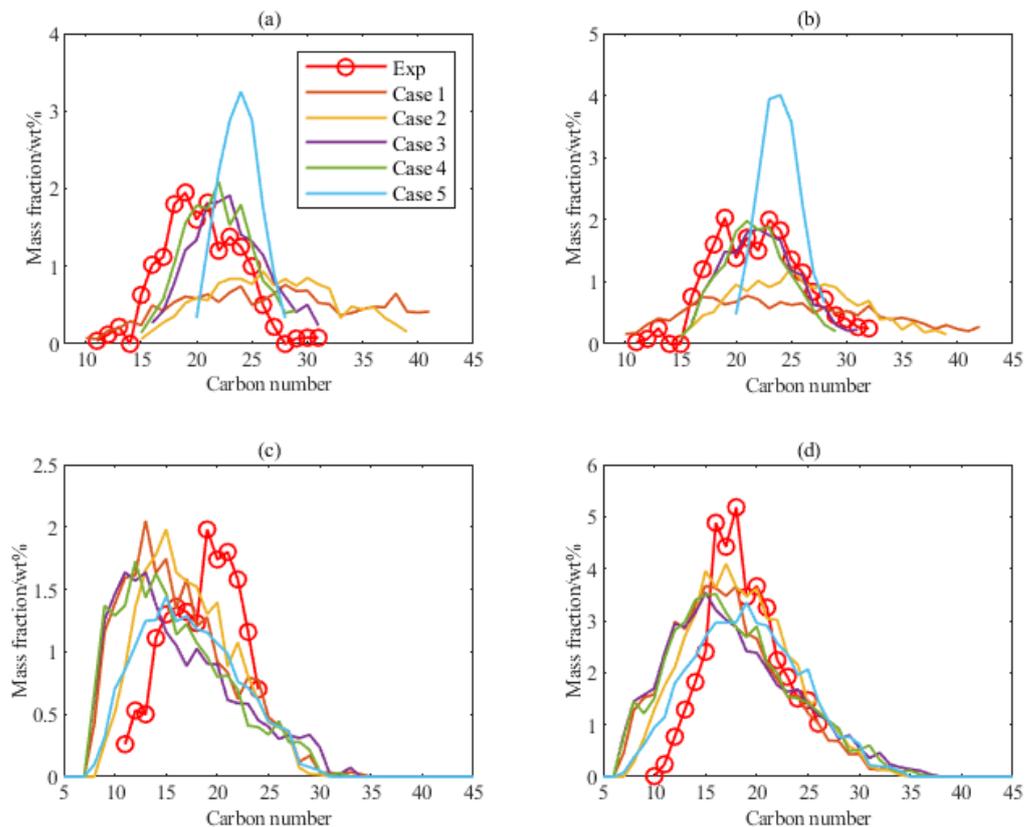

Figure 5. The distributions of mass fraction in (**a**) normal paraffins, (**b**) isoparaffins, (**c**) naphthenes, (**d**) aromatics for gas oil A by experiment and predictions.

4. Conclusions

Based on gas oil samples, this work studies the performance of the SR model in different domains of parameters. In the setting of domains, multiple cases for the upper boundaries of shape parameters in gamma distributions are considered. The results show that in each case, the SR model can generate pseudo mixture, the bulk properties of which agree well with the experimental values. The bulk properties adopted are elemental analysis, density, H/C ratio, aromatic sulfur content, PINA analysis and simulated distillation. By investigating the variations in average objective function value in each case, it is observed that even though the decrease of average objective function value in smaller domains is faster at the beginning of the simulation, the SR model can find parameters in wider domains that make the average objective function value smaller. Case 3 is the best case where the simulating accuracy and convergence performance are maintained compared with other cases. The prediction of the SR model on compositions of gas oil sample is not acceptable, despite the estimated values of bulk properties being very close to the experimental values. Large derivations are observed in all cases.

The detailed group-type analysis contributes to more accurate predictions on compositions. The predicted distributions in normal paraffins and isoparaffins become higher as the upper boundaries of shape parameters increase. In cases 3 and 4, the predicted distributions fit best with the experimental distributions, compared to the other cases. As for naphthenes and aromatics, the predicted distributions agree well with experimental values in all cases. Because case 3 performed best in maintaining simulating accuracy and

convergence rate, the best upper boundary of shape parameters is 20, compared to other values in this work.

The domain of parameters in probability functions has a great influence on the performance of the SR method. Due to the differences in bulk properties and compositions of petroleum fractions, reasonably setting the domain of parameters is indispensable for the accurate reconstruction of petroleum fractions.

Abbreviations

PDF	Probability density function
PINA	Normal paraffins, isoparaffins, naphthenes, and aromatics
SA	Structural attribute
SP	Shape parameter
UB	Upper boundary
LB	Lower boundary
MN	Mononaphthenes
DN	Dinaphthenes
MA	Monoaromatics
NA	Naphthenoaromatics
DA	Diaromatics
NDA	Naphthenodiaromatics
TrA	Triaromatics
NTrA	Naphthenotriaromatics
TeA	Tetraaromatics
EV	Experimental value

References

1. Qian, F.; Zhong, W.; Du, W. Fundamental Theories and Key Technologies for Smart and Optimal Manufacturing in the Process Industry. *Engineering* **2017**, *3*, 154–160. [CrossRef]
2. Shen, F.; Wang, M.; Huang, L.; Qian, F. Exergy analysis and multi-objective optimisation for energy system: A case study of a separation process in ethylene manufacturing. *J. Ind. Eng. Chem.* **2021**, *93*, 394–406. [CrossRef]
3. Stratiev, D.; Shishkova, I.; Tankov, I.; Pavlova, A. Challenges in characterization of residual oils. A Review. *J. Pet. Sci. Eng.* **2019**, *178*, 227–250. [CrossRef]
4. Ren, Y.; Liao, Z.; Sun, J.; Jiang, B.; Wang, J.; Yang, Y.; Wu, Q. Molecular reconstruction: Recent progress toward composition modeling of petroleum fractions. *Chem. Eng. J.* **2019**, *357*, 761–775. [CrossRef]
5. Hudebine, D.; Verstraete, J.J. Molecular reconstruction of LCO gasoils from overall petroleum analyses. *Chem. Eng. Sci.* **2004**, *59*, 4755–4763. [CrossRef]
6. 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*, 1020–1034. [CrossRef]

7. Quann, R.J.; Jaffe, S.B. Structure-oriented lumping: Describing the chemistry of complex hydrocarbon mixtures. *Ind. Eng. Chem. Res.* **1992**, *31*, 2483–2497. [CrossRef]
8. Wu, Y. Molecular Management for Refining Operations. Ph.D.Thesis, The University of Manchester, Manchester, UK, 2010.
9. Neurock, M.; Libanati, C.; Nigam, A.; Klein, M.T. Monte carlo simulation of complex reaction systems: Molecular structure and reactivity in modelling heavy oils. *Chem. Eng. Sci.* **1990**, *45*, 2083–2088. [CrossRef]
10. Neurock, M.; Nigam, A.; Trauth, D.; Klein, M.T. Molecular representation of complex hydrocarbon feedstocks through efficient characterization and stochastic algorithms. *Chem. Eng. Sci.* **1994**, *49*, 4153–4177. [CrossRef]
11. Petti, T.F.; Trauth, D.M.; Stark, S.M.; 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*, 570–575. [CrossRef]
12. Zhang, L.; Hou, Z.; Horton, S.R.; Klein, M.T.; Shi, Q.; Zhao, S.; Xu, C. Molecular representation of petroleum vacuum resid. *Energy Fuels* **2014**, *28*, 1736–1749. [CrossRef]
13. Deniz, C.U.; Yasar, M.; Klein, M.T. A New Extended Structural Parameter Set for Stochastic Molecular Reconstruction: Application to Asphaltenes. *Energy Fuels* **2017**, *31*, 7919–7931. [CrossRef]
14. Deniz, C.U.; Yasar, S.H.O.; Yasar, M.; Klein, M.T. Effect of Boiling Point and Density Prediction Methods on Stochastic Reconstruction. *Energy Fuels* **2018**, *32*, 3344–3355. [CrossRef]
15. Gani, R.; Nielsen, B.; Fredenslund, A. A group contribution approach to computer-aided molecular design. *AIChE J.* **1991**, *37*, 1318–1332. [CrossRef]
16. Constantinou, L.; Gani, R.; O'Connell, J.P. Estimation of the acentric factor and the liquid molar volume at 298 K using a new group contribution method. *Fluid Phase Equilibria* **1995**, *103*, 11–22. [CrossRef]
17. Marrero, J.; Gani, R. Group-contribution based estimation of pure component properties. *Fluid Phase Equilibria* **2001**, *183–184*, 183–208. [CrossRef]
18. Hukkerikar, A.S.; Sarup, B.; Ten Kate, A.; Abildskov, J.; Sin, G.; Gani, R. Group-contribution^{+} (GC^{+}) based estimation of properties of pure components: Improved property estimation and uncertainty analysis. *Fluid Phase Equilibria* **2012**, *321*, 25–43. [CrossRef]
19. Yen, L.C.; Woods, S.S. A generalized equation for computer calculation of liquid densities. *AIChE J.* **1966**, *12*, 95–99. [CrossRef]
20. Haktanır, M.; Karahan, S.; Yaşar, M. Structurally explicit composition model of petroleum vacuum residue. *Fuel* **2021**, *300*, 120977. [CrossRef]
21. Glazov, N.; Dik, P.; Zagoruiko, A. Effect of experimental data accuracy on stochastic reconstruction of complex hydrocarbon mixture. *Catal. Today* **2021**, *378*, 202–210. [CrossRef]
22. Bojkovic, A.; Dijkmans, T.; Dao Thi, H.; Djokic, M.; van Geem, K.M. Molecular Reconstruction of Hydrocarbons and Sulfur-Containing Compounds in Atmospheric and Vacuum Gas Oils. *Energy Fuels* **2021**, *35*, 5777–5788. [CrossRef]