energy&fuels
Article
Cite This: Energy Fuels 2018, 32, 1525−1533
pubs.acs.org/EF

Molecular Representation of the Petroleum Gasoline Fraction

Chen Cui,[†] Triveni Billa,[‡,§] Linzhou Zhang,[*,†] Quan Shi,[†] Suoqi Zhao,[†] Michael T. Klein,[*,‡,§] and Chunming Xu[†]

[†]State Key Laboratory of Heavy Oil Processing, China University of Petroleum, Beijing 102249, People's Republic of China
[‡]Department of Chemical and Biomolecular Engineering, University of Delaware, Newark, Delaware 19716, United States
[§]Center for Refining and Petrochemicals, King Fahd University of Petroleum and Minerals (KFUPM), Dhahran 31261, Saudi Arabia

ABSTRACT: The computer-aided reconstruction of gasoline composition is an active area of petroleum and petrochemical research as a result of the demand for molecular-level management of the petroleum feed streams. To that end, in this work, a molecular compositional model based on a predefined representative molecular set was built that allows for the conversion of conventional bulk property data to an approximate molecular composition. The selection of representative molecules was based on their presence in gasoline molecular compositional measurement and their potential contribution to the key physical properties. Around 170 hydrocarbons and heteroatom species were chosen as predefined identities of molecules that can exist in a gasoline sample. The physical property data of all of the representative molecules were collected, and suitable mixing rules for the gasoline range stream were applied for the accurate prediction of bulk properties. The approximate concentration of representative molecules was obtained through fitting the predicted bulk property to the measured data. The methodology was verified through intensive tests on various gasoline samples, including straight-run naphtha, catalytic cracking gasoline, coking gasoline, and reformates. The modeling was also accomplished in a sequential order using basic to advanced measurements to find the optimum number of measurements required for detailed composition evaluation on various feedstocks. The propagation of error in the experimental measurement and prediction method on composition has been evaluated.

1. INTRODUCTION

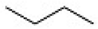
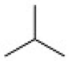
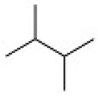
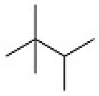
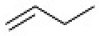
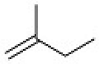
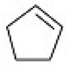
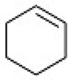
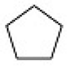
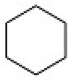
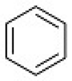
On a molecular basis, petroleum is a complex mixture of hydrocarbons with a small amount of organic compounds containing sulfur, oxygen, and nitrogen as well as metallic constituents. Characterization of petroleum molecular composition is an attractive research topic over a long time, and instrumental analysis is a direct method to access the molecular composition. A lot of progress has been achieved in the past few decades, and now there are a number of standard methods that have been proposed and used in various fields of the petroleum industry. However, compositional complexity of petroleum forms a challenge for modern analytical chemistry as a result of the increased number of isomers with an increase in the carbon number. Even gasoline, the lightest part of petroleum fractions, the state of art analytical techniques, such as gas chromatography−vacuum ultraviolet (GC−VUV) detector,[1,2] gas chromatography−mass spectrometry (GC−MS),[3] comprehensive two-dimensional gas chromatography (2D GC),[4] may not be able to give all of the molecular composition details. Moreover, in recent years, the gasoline quality standard upgraded rapidly for the reason of environmental issues. Motivated by the profit, refineries have to optimize the process of gasoline conversion and blending at a molecular level. Therefore, there is an urgent need to develop models to transform the available, usually indirect, analytical information into a representation of molecular composition of gasoline.

Computer-aided reconstruction of petroleum composition based on the limited experimental information became more and more popular as a result of the increased optimization level of the refinery process. The successful development of a full molecular-level petroleum compositional model started in the 1990s. Quann and Jaffe[5,6] proposed the structure-oriented lumping (SOL) framework, which allows for the simulation of petroleum composition and conversion at a molecular level, for example, hydrodesulfurization of fluid catalytic cracking (FCC) naphtha.[7] This SOL-based molecular composition model was generally obtained through fine tuning the compositional measurement data. Neurock et al.[8] developed an approach of sampling randomly different statistical distributions of structural attributes by the Monte Carlo method to reconstruct a large ensemble of molecules whose properties are close to experimental data. Hudebine et al.[9] extended the approach by introducing the entropy maximization method after stochastic reconstruction. Recently, Pan et al.[10] applied a similar strategy on gasoline, and the SOL vectors was used as structural attributes for random sampling. Because Monte Carlo sampling is time-consuming, some recent studies preferred to use a predefined molecular library, which consists of a series of representative molecules. Different strategies and frameworks have been developed to build a predefined molecular library. Peng[11] proposed a molecular-type homologous series (MTHS) matrix to represent the composition of the petroleum fraction, which has a similar form to the *n*-paraffin, isoparaffin, olefin, naphthene, and aromatics (PIONA) versus carbon number result of gas chromatography with a flame ionization detector (GC−FID). Zhang[12] predicted the composition matrix of gasoline using the interpolation method. Aye et al.[13] built a composition database of gasoline samples based on MTHS and then automatically

Received: November 16, 2017
Revised: January 12, 2018
Published: January 17, 2018

DOI: 10.1021/acs.energyfuels.7b03588

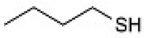
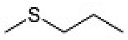
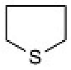
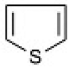
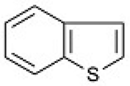
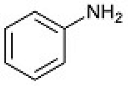
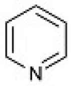
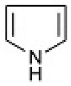
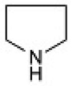
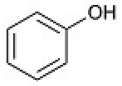
Table 1. Organization of the Hydrocarbon Homologous Series in the Predefined Gasoline Molecular Identity Library

Index	Chemical Family	Code Name	Representative Molecule (name)	Representative Molecule (structure)
1	Normal-paraffin	NP	*n*-Butane	
2	Methylparaffin	MP	Isobutane	
3	Dimethylparaffin	DP	2,3-dimethylbutane	
4	Trimethylparaffin	TP	2,2,3-trimethylbutane	
5	Normal-olefin	NO	1-butene	
6	Branched-olefin	BO	2-methyl-1-butene	
7	Five-membered Cyclic Olefin	CO_5	Cyclopentene	
8	Six-membered Cyclic Olefin	CO_6	Cyclohexene	
9	Five-membered Naphthene	N5	Cyclopentane	
10	Six-membered Naphthene	N6	Cyclohexane	
11	Aromatics	A	Benzene	

selected the sample matrix to generate a new matrix. Hu et al.[14] used the concept of MTHS to reconstruct the composition of reformat and extended its application to refinery optimization. Wu et al.[15] improved the transformation methodology of Aye and added several series in the matrix to describe gasoline composition more accurately. Albahri et al.[16] selected representative molecules directly from the gas chromatography (GC) analysis result. The molecular library proposed by Albahri contains 68 representative molecules for the light naphtha fraction. A similar approach was also used by Hudebine et al.[17] They established a FCC gasoline molecular library containing 230 different molecules.

The combination of GC analysis data and bulk properties was proposed by Ghosh et al.[18] In their model, the GC analysis data served as the initial set of gasoline composition, which was fine-tuned by defining an objective function in terms of bulk properties. This approach of modeling is widely applied in petroleum industries, which is not only governed by the computational technique but also by the analytical data. The precision of this modeling greatly depends upon the accuracy of the measured bulk properties and the exact number of measurements required for detailed composition evaluation. Various users of these models, such as researcher and process engineers, have a different expectation and tolerance for uncertainty, but they are all interested in the trade-off between information and measurement costs. In view of these aspects, a compositional model is introduced for approximate solution of gasoline samples and the impact of uncertainty associated with

Table 2. Organization of the Heteroatom Species Homologous Series in the Predefined Gasoline Molecular Identity Library

Index	Chemical Family	Code name	Representative Molecule (name)	Representative Molecule (structure)
1	Mercaptan	SI	N-butylmercaptan	
2	Sulfide	SII	Methyl-N-propylsulfide	
3	Tetrahydrothiophene	SIII	Tetrahydrothiophene	
4	Thiophene	SIV	Thiophene	
5	Benzothiophene	SV	Benzothiophene	
6	Aniline	NI	Aniline	
7	Pyridine	NII	Pyridine	
8	Pyrrole	NIII	Pyrrole	
9	Tetrahydropyrrole	NIV	Tetrahydropyrrole	
10	Phenol	O	Phenol	

the number of measurements required to find detailed feedstock composition has been evaluated.

2. METHODOLOGY

In most of the process simulators, the pseudo-component composition of the petroleum stream is computed using bulk properties as input, such as boiling point and density. In refineries, the physical and chemical properties of gasoline are available because they were measured frequently for the purpose of the product quality control. The present work follows a similar methodology that converts bulk properties into compositional information. This model computes compositional information at the molecular level rather than chemically undefined pseudo-components. It will potentially facilitate the development of more accurate blending and kinetic models.

The development of the present model can be categorized into two steps: the determination of molecular identities in a predefined universal gasoline molecule library and the calculation of the molecular fraction in a given stream.

2.1. Predefined Molecular Library. A typical gasoline stream mainly consists of hydrocarbons with a carbon number ranging from 4 to 12. Most hydrocarbons in gasoline are paraffins, olefins, naphthenes, and aromatics. Heteroatom (most of them are sulfur, nitrogen, and oxygen)-containing species occupy a relatively lower amount. On the other hand, the composition of gasoline streams varies according to their origin. For example, the straight-run gasoline does not contain olefins, and the FCC gasoline may have a high content of sulfur- and nitrogen-containing species. The predefined molecule in the library should have three features: First, it should be universal for all types of gasoline, and thus, the compositional difference between gasolines was expressed in terms of the variation of the molecular content in the library. Second, the molecule should have significant contribution to the bulk property. Third, the abundance of the molecules should be relatively high. Hence, under these three principles, different types of gasoline samples were collected and details of hydrocarbon-, sulfur-, and nitrogen-containing species compositional analysis were performed using GC coupled with a FID, sulfur chemiluminescence detector (SCD), and nitrogen chemiluminescence detector (NCD), respectively. The detectable

DOI: 10.1021/acs.energyfuels.7b03588

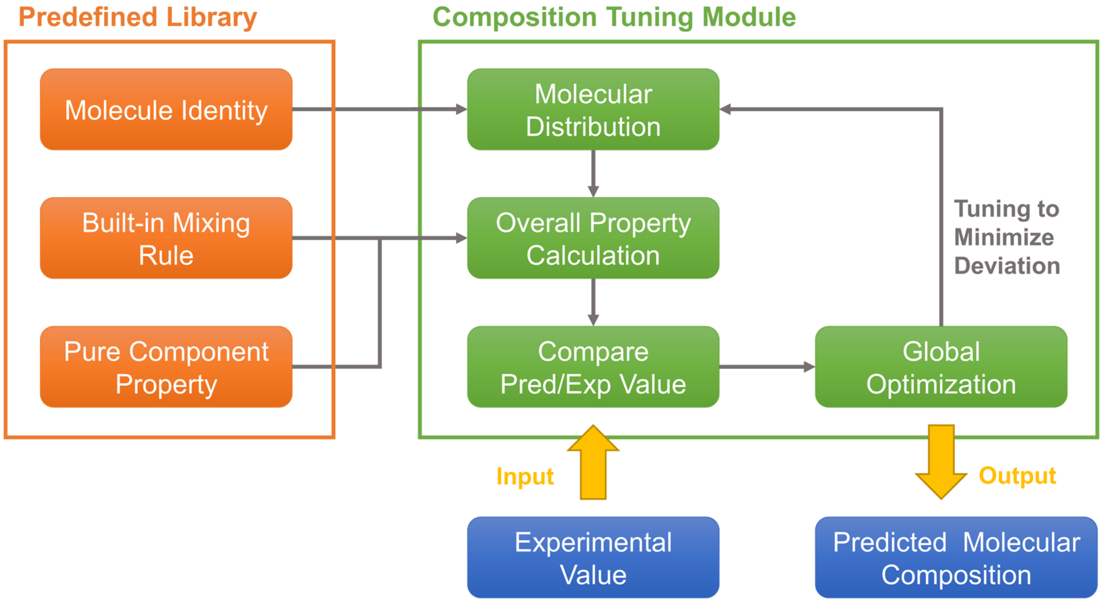


Figure 1. Flowchart of molecular composition model development for the gasoline stream.

species (summed up to ~400 molecules) was initially classified into different chemical families based on the structural characteristic. The contributions of the molecular structure on properties were manually analyzed, and an initial set of chemical families was selected. It should be noted that only the molecules with a carbon number of <8 can be clearly identified in detailed hydrocarbon analysis using GC−FID. For the C_{9+} components, the huge number of isomers is beyond the separation capacity of the GC column and a lot of compositional information is lost as a result of the co-elution problem. The C_{9+} components in the library were proposed through extension of each homologous series by adding a $-CH_2-$ functional group. The oxygenated compounds were given less attention in the present model, and only a phenol homologous series is added to provide the capability of oxygen elemental content calculation. The additives, such as ethanol and MTBE, have not yet been considered at this stage.

There are 21 representative chemical families in the predefined molecular library, which consists of 89 hydrocarbons and 81 heteroatom-containing species. As shown in Table 1, hydrocarbons were divided into 11 chemical families. The code name represents the chemical family name. The last two columns show the chemical name and structure of a representative molecule in the chemical family. In regular gasoline analysis, the gasoline molecules were divided into five groups named PIONA, which represents *n*-paraffin, isoparaffin, olefin, naphthene, and aromatics, respectively. In this work, the chemical family was defined with more structural details. The *n*-paraffin homologous series is kept as is because its structure is unique. Isoparaffins were extended into three types, which contain mono-, di-, and trimethyl functional groups, because the branching degree of paraffin strongly affects the octane number. There are four olefin homologous series in the library: *n*-olefin, branched olefin (represented by monobranched olefins), and five- and six-membered cyclic olefins. Naphthenes were further divided into five- and six-membered cyclic structures. The benzene homologous series was considered as a representation of aromatics. Similar to hydrocarbons, 10 chemical families of heteroatom species (S, N, and O) are shown in Table 2. Besides the structural identity, physical properties of pure components in the predefined molecular library were also collected. Most of the physical properties of the molecules in the library were directly obtained from thermodynamic databases and the literature. The group contribution method was employed to find the unknown properties of molecules.

2.2. Approximate Molecular Composition Calculation.

Once the molecular identities have been qualitatively specified, the other problem that is needed to be solved is the quantity of each molecule. Rather than direct analytical measurement, the present work uses bulk property as input to calculate approximate molecular composition. Once the experimental bulk property values were set, they served as constraints in an objective function to regress the molar fraction of the molecule in the gasoline stream. It should be noted that there are 170 molecules in the library. It means that the number of unknown variables (170) is much higher than the input data point number (depends upon the performed measurement; a regular value is ~20). Thus, more constraints are required to obtain a proper solution. Previous studies revealed that the molecular composition of petroleum is a continuum and follows a statistical distribution. The molar fraction of species in each chemical family can be organized by a probability density function with a limited number of parameters, which leads to a significant decrease in the number of variables.

The simulation framework of the present model is shown in Figure 1. The definition and pure component properties of representative molecules were described in the previous section. Mixing rules that are suitable for the gasoline range stream were collected, especially for the bulk property that shows nonlinear behavior during blending (e.g., ASTM D86 boiling point profile and octane number value). The predefined libraries of chemical identity, pure component property, and mixing rule allow for the calculation of the bulk property with a given molecular composition. The composition tuning module uses an optimization engine to search for the best molecular distribution function parameters that minimize the difference between the predicted and measured bulk property data. The Γ probability density function was used to represent the molecular distribution. The Γ distribution has been widely applied in the various petroleum molecular composition modeling studies. The mathematical expression of the Γ distribution is shown below

$$p(x) = \frac{(x-\eta)^{\alpha-1}\mathrm{e}^{-(x-\eta)/\beta}}{\beta^{\alpha}\Gamma(\alpha)} \tag{1}$$

where p is the probability of the Γ distribution. The three parameters of the Γ function, α, β, and η, determine the shape of the Γ function, and they were the tuning parameters for the optimization engine. Because the gasoline fractions were generally obtained through distillation, the boiling point of each molecule served as the parameter x in eq 1. Once α, β, and η were set, the content of molecules in the same chemical family could be calculated from the boiling point. Hence, eq 1 can be rewritten as

$$y_{i,j} = \frac{(T_{i,j}^{b}-\eta_j)^{\alpha_j-1}\mathrm{e}^{-(T_{i,j}^{b}-\eta)/\beta_j}}{\beta_j^{\alpha_j}\Gamma(\alpha_j)} \tag{2}$$

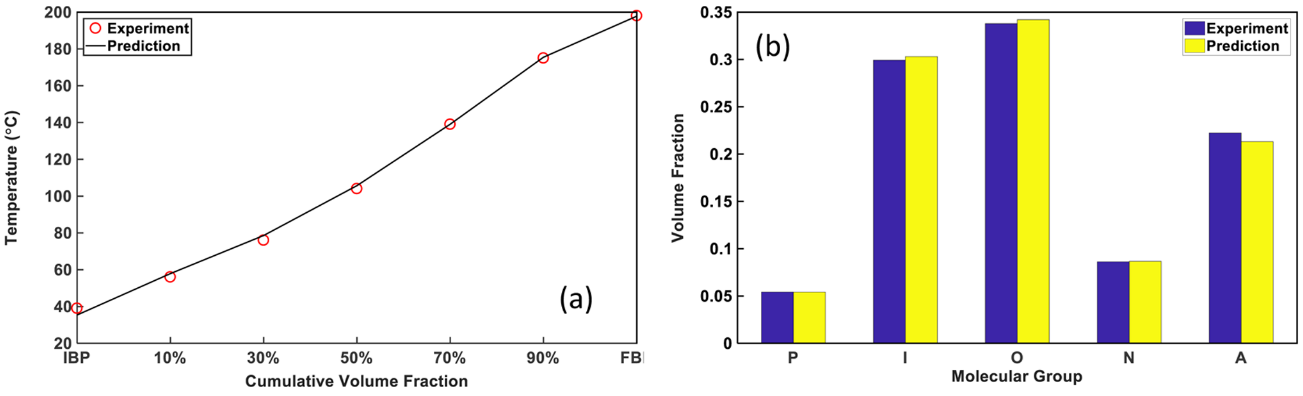


Figure 2. Comparison between predicted and experimental data: (a) boiling point cumulative distribution and (b) PIONA volume fractions.

Then

$$x_{i,j} = x_j \frac{y_{i,j}}{\sum_{ii} y_{ii,j}} \tag{3}$$

where $y_{i,j}$ in eq 2 stands for the fraction of the molecule (i, j) within the j series of the chemical family. Finally, the content of each molecule can be calculated by eq 3. The parameter x_j in eq 3, which stands for the content fraction of the j series within all of the series of the chemical family, will be optimized too. Thus, four parameters were used to describe each series. In the present work, there are 21 series, which give 84 fitting parameters in total.

A complete simulation process is initially experimental data of gasoline bulk properties input into the model. Then, the global optimization engine began to tune the Γ distribution parameters of each chemical family. Once the Γ function parameters were set, the molar fraction of each representative molecule was obtained through sampling. On the basis of the built-in mixing rules, the bulk property of obtained molecular composition was predicted. The objective function value was then calculated by comparing predictions and experimental data and delivered to the global optimization engine. The optimization will change the tuning parameter based on the feedback from the objective function. After the above steps were repeated, the objective function value was minimized and the best parameter values can be obtained. Finally, molecular composition generated at this situation was considered as the best predicted composition and was outputted. The objective function is

$$\text{obj} = \sum_i \left(\frac{P_i^{\text{msd}} - P_i^{\text{pred}}}{P_i^{\text{msd}}} w_i \right)^2 \tag{4}$$

where obj is the objective function, P^{msd} is the experimental data, P^{pred} is the predicted data, and w is the weight factor. The weight factor is manually set based on the accuracy of measurement and experimental error. For instance, if the final boiling point of a gasoline sample is 200 °C, given that the measurement error of the final boiling point is relatively large, the deviation of prediction within 10 °C was considered as acceptable. If the relative error was defined as

$$\text{relative error} = \frac{|P_i^{\text{msd}} - P_i^{\text{pred}}|}{P_i^{\text{msd}}} \times 100\% \tag{5}$$

that is, if the maximum relative error is 5% for a certain property, the weight of this property was set as 20 (reciprocal of 5%).

The run time depends upon the algorithm parameters set by users. In most cases, it takes around 1 min for a single run to obtain a fairly good result. This model has the advantages of good expansibility because there are only four variables that need to be optimized for each series, regardless of the number of molecules in a series. Thus, the model can be used to predict molecular details of heavier fractions, such as diesel or gas oil. Although the maximum entropy method can reduce the number of tuning variables significantly, properties, such as the octane number, cannot be added to the objective function as a result of their nonlinear nature. In contrast, a nonlinear constraint can be added to the objective function as easily as a linear constraint in the present model.

3. CASE STUDY

3.1. Applying the Model to a FCC Gasoline.

FCC gasoline was used as an example to illustrate the simulation process of the present model. The bulk property data for the FCC gasoline were measured and served as input. The bulk properties used in this case are density, vapor pressure, elemental composition, distillation profile, PIONA fraction, refractive index (RI), and octane numbers. The data were input into an in-house computer program with a preset weight fraction in the objective function. The optimization engine tuned the molecular distribution and minimized the difference in predicted and input property data. After optimization, the approximate molecular composition was obtained as well as properties that were not measured.

Table 3. Comparison of Experimental and Model-Predicted Properties for the FCC Gasoline

property	exp	pred	relative error (%)
SG	0.7312	0.7469	2.1
RVP (kPa)	56	55.2	1.4
RON	88.7	89.8	1.2
MON	77.4	79.3	2.4
C/H (w/w)	6.46	6.54	1.2
RI	1.4241	1.4126	0.81
total S (ppm)	100	101	1.0
total N (ppm)	77	80	3.9
MW		130.6	
P_c (kPa)		3168	
T_c (°C)		285.1	
V_c (m^3/kg)		0.0040	
K_w		11.85	
aniline point (°C)		58.8	
ω		0.3191	
Z_c		0.2708	
kinematic viscosity at 100 °F (cSt)		0.6491	

Figure 2 shows the comparison between predicted and experimental data for (a) boiling point distribution and (b) PIONA volume fractions. Distillate profile and PIONA fractions are the most fundamental properties of gasoline, and a good agreement between predicted and experimental was observed. It is important to note that the distillate profile was calculated in terms of the simulated distillate profile at the beginning of the

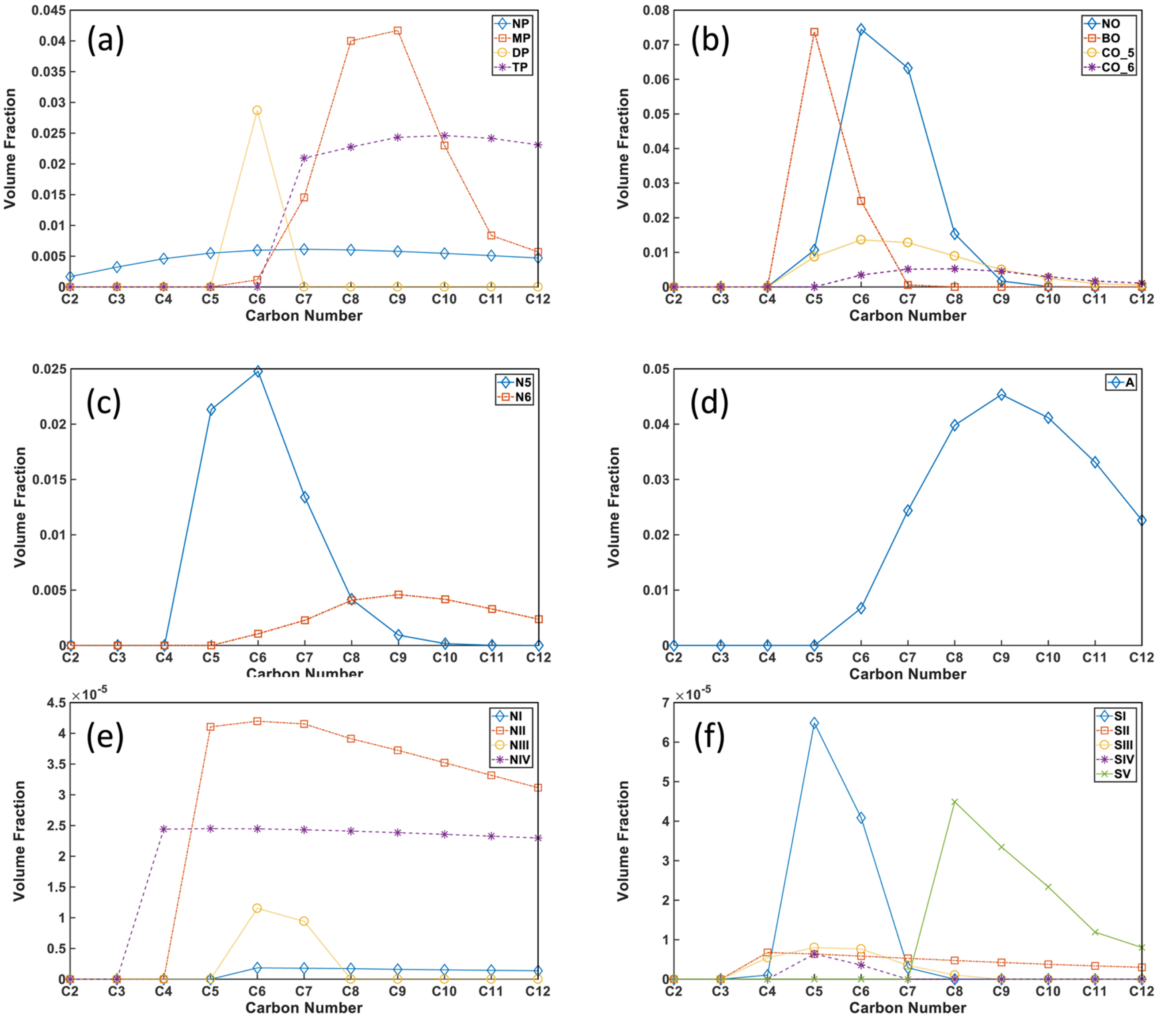


Figure 3. All types of hydrocarbons and heteroatom species distribution produced by simulation: (a) paraffins, (b) olefins, (c) naphathenes, (d) aromatics, (e) nitrogen-containing species, and (f) sulfur-containing species.

prediction. To match the experimental data, which was often given in terms of ASTM D86, and in the case of the situation of the predicted initial boiling point being lower than 0 °C, the simulated distillate profile had to be converted to the ASTM D86 profile at the phase of calculating the objective function. In addition to the boiling range and PIONA volume fraction, the other bulk properties were also compared in Table 3. The relative error of most properties is less than 5%. The developed method has the capacity of finding an approximate composition that has almost the same property of the targeted gasoline sample. On the basis of the obtained molecular composition, the properties that are difficult to measure could be predicted. For example, the critical properties (critical pressure, critical temperature, and critical volume) of the FCC gasoline were obtained. The parameters are very useful for process simulation, especially for phase equilibrium calculations.

Figure 3 shows predicted molecular composition of both hydrocarbons and heteroatom species. The detail molecular composition not only gives capacity of predicting various bulk properties but is also useful for molecular-level kinetic modeling. The molecular distribution of the chemical family was described by Γ functions. The Γ function parameters were tuned by the optimization engine. As seen in Figure 3, the molecular volume distribution of each chemical family varies significantly. It should be noted that the distribution of heteroatom species may have higher deviations. Because they only occupied a relatively small amount compared to hydrocarbons, their distribution was mainly constrained by overall elemental composition. If higher heteroatom species composition accuracy is required, more detailed heteroatom species experimental data (e.g., GC−SCD) should be added.

3.2. Model Validation for Different Types of Gasoline.

To prove the model validation, 51 gasoline samples were tested by the method, which included 2 FCC gasolines, 2 coker gasolines, 46 straight-run gasolines, and 1 reformate. The bulk property data of these samples were collected, such as the distillation profile, PIONA, specific gravity, octane number, and Reid vapor pressure (RVP). These measured properties were input into the model, and the predicted results were compared to the experimental data.

DOI: 10.1021/acs.energyfuels.7b03588
Energy Fuels 2018, 32, 1525−1533

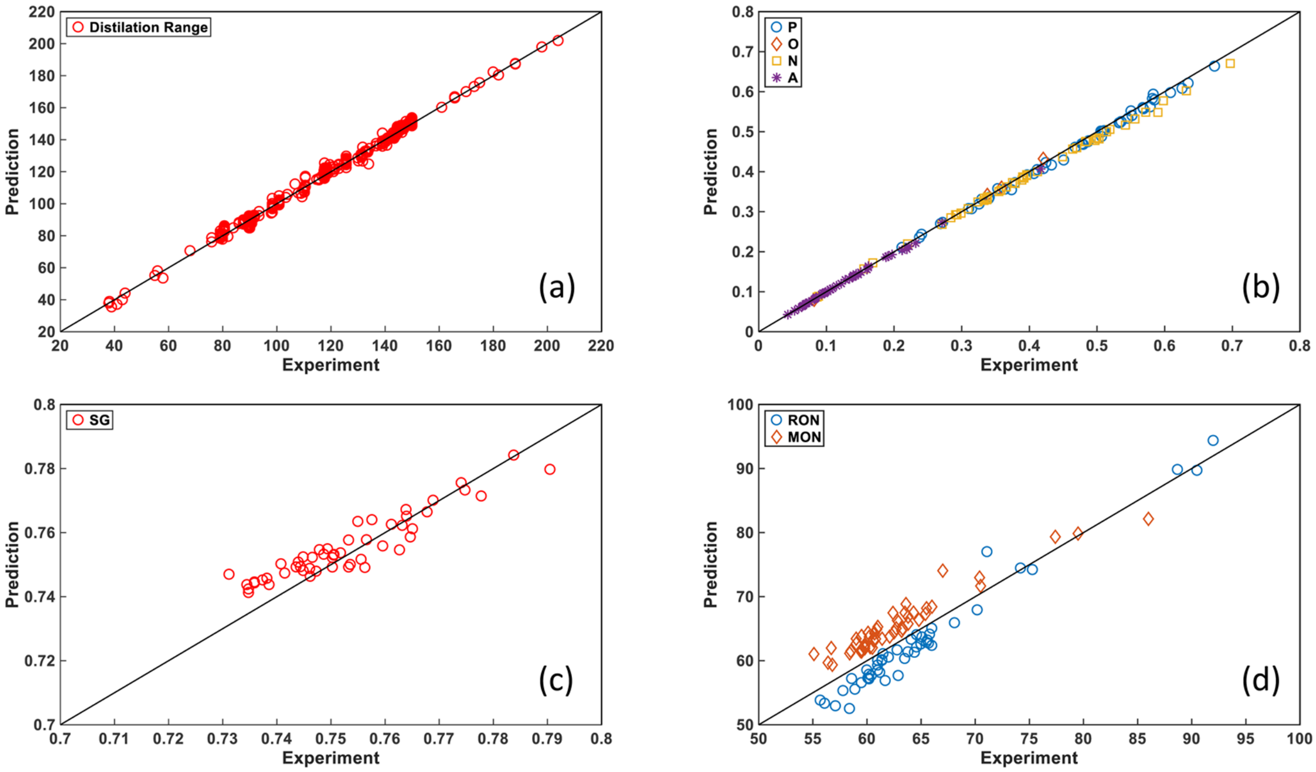


Figure 4. Comparison between predicted and experimental data for 51 gasolines: (a) distillation range, (b) PONA, (C) specific gravity, and (d) octane number.

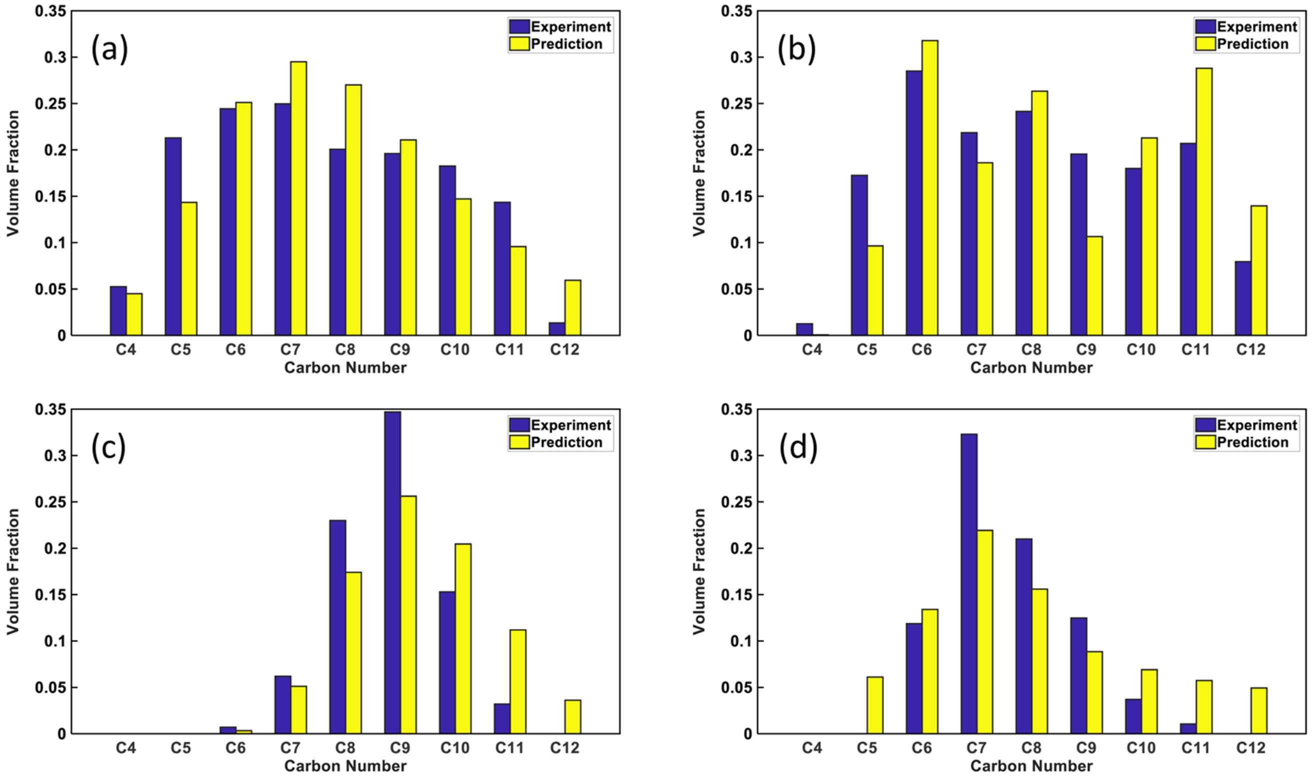


Figure 5. Comparison of the predicted and experimental data of a naphtha sample: (a) paraffin content comparison, (b) isoparaffin content comparison, (c) naphthene content comparison, and (d) aromatic content comparison.

A detailed comparison of the property can be seen in Figure 4. It is clear that this model has good performance on the distillation profile, PIONA content, and specific gravity. However, the deviation of the octane number is greater than other properties. The deviation is mainly for straight-run gasoline samples. As shown in this figure, the prediction of the motor octane number

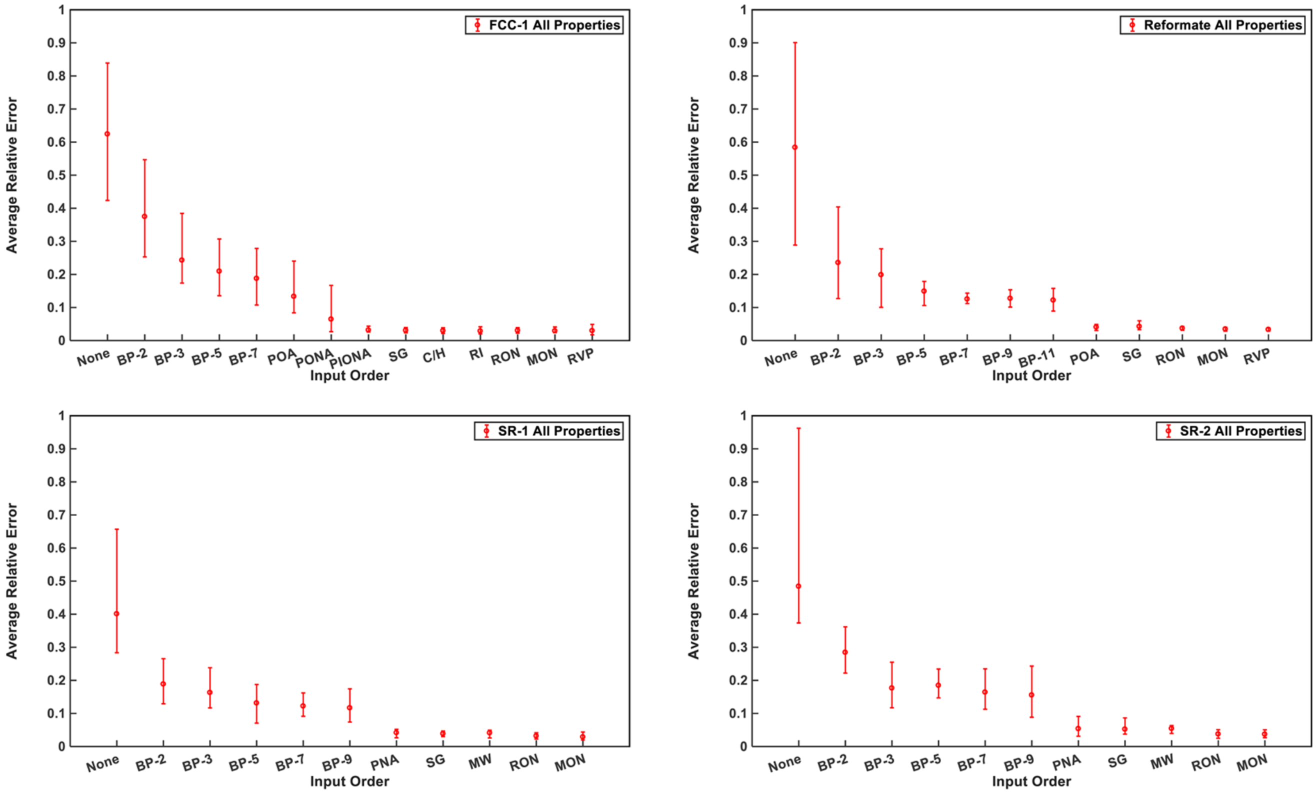


Figure 6. Uncertainty analysis of all predicted properties for four gasolines. The average relative error is calculated by averaging the error value of all measured properties of multiple simulations.

(MON) tends to be larger and that of the research octane number (RON) tends to be smaller. Thus, it seems that the deviation is caused by the systematic errors of the octane number prediction model. The error may be reduced by a correlative correction between predicted and experimental data.

3.3. Comparison of the Detailed Hydrocarbon Composition.

The comparison between the model-derived approximate composition and experimental group type and carbon number distributions was also made in this work. GC−FID is used to obtained the POINA versus carbon number distribution of the gasoline sample. A straight-run gasoline is used as the sample.

As shown in Figure 5, the prediction of the distribution trend of each series is basically the same as the experimental value. Especially for the isoparaffin series, the distribution calculated by this model matches the experimental distribution, which has three peaks in total. This is because isoparaffins are divided into three series in our molecular library, mono-, di-, and trimethyl series. Each series has its own distribution. Therefore, it is possible to result in the multi-peak situation when the three series are integrated into the new isoparaffins series. The results not only show the performance of our model on the prediction of the molecular composition but also the rationality of our molecular library.

4. EFFECT OF INPUT PROPERTY ON THE MODEL ACCURACY

Many bulk properties can be used as input. However, sometimes, it is not easy to access all of these properties because the cost associated in measuring these properties is very high. Therefore, it is important to find the key properties for the gasoline composition approximation. The influence of inputs was investigated by increasing the number of inputs from basic to detailed measurements. Given the random nature of the optimization engine, multiple runs were performed on each test that contain a certain number of inputs. The degree of influence was evaluated by the average deviation of each test. The fluctuation of average deviation was also a concern because it is another critical indicator for evaluation. To remove the influence of the weight factor, all of the weights were set equal. Here, the same test was performed on five gasoline samples to investigate whether the influence of the number of inputs was the same on different kinds of gasoline. The result shows that once the distillation profile and PIONA were input, the average deviation almost declined to the lowest point.

As shown in Figure 6, all experimental data would be input to the model according to the order shown on the *x* axis. When the input order of FCC-1 is taken as an example, "none" means that nothing was input. The boiling range is comprised of a series of boiling points against the cumulative volume, such as the initial boiling point (IBP), final boiling point (FBP), 10% distillate temperature, etc. Thus, "BP" stands for boiling point, and "-number" after "BP" stands for the number that we chose from the boiling range. For instance, "BP-3" means that the IBP, FBP, and 50% distillate temperature of these three points were selected. Similarly, the PIONA fraction could be subdivided into "POA", which stands for all saturated hydrocarbons, olefins, and aromatics, and "PONA", which stands for all paraffins, olefins, naphthenes, and aromatics. "SG", "C/H", "RI", "RON", "MON", and "RVP" stand for specific gravity, carbon−hydrogen mass ratio, refractive index, research octane number, motor octane number, and Reid vapor pressure, respectively. Each *x*-axis coordinate represents a new test, and the label is the new input, which was added for this test; that is, this certain label and the former would be input together for this test. When the number of input measurements is few, the possible solutions from the optimizer are more than one. To account for this, each simulation is performed multiple times (~20 times), and the results are shown as error bars. The *y* axis is the average relative error, which is calculated by averaging the relative error value of all properties in multiple simulations. It is clear that, from left to right of the *x* axis, as the details in input are added, the average relative error and its fluctuation decrease. It must be noticed that the average relative error and its fluctuation decrease sharply at two locations. One is when all of the boiling range details were included in the input, and the other is when all of the PIONA fraction was input. After that, the change of the average relative error and its fluctuation is not obvious. This phenomenon exists in all four subgraphs. Accordingly, the boiling range and PIONA fraction are considered as the most important bulk properties of the model. Therefore, these two properties should be given priority to ensure the completeness and accuracy of the model predictions.

This model is not sensitive to properties, such as specific gravity, because SG of each representative molecule in the library is not so far from that of gasoline. Thus, under the equal weight circumstances, the model will give priority to those properties that have bigger deviation. Therefore, this model is more sensitive to the distillation profile and PIONA.

5. CONCLUSION

(1) A predefined molecular library was built on the basis of feedstock source, physical and chemical properties, and GC−FID detection. (2) There were 170 molecules in total, which included both hydrocarbons and heteroatom species. This library could be used to represent different kinds of gasoline, such as naphtha, FCC gasoline, reformate, and coker gasoline. (3) The model was validated by 51 different types of gasoline samples. The predictions of properties were in good accordance with experimental data. Although the deviation of the predicted detailed hydrocarbon composition was larger than that of properties, the results were acceptable. (4) The relationship between the input and the model accuracy was also investigated in the present work. It could be found that once the boiling profile and PIONA were input, the average deviations almost declined to the lowest point. Thus, these two properties should be treated as the most critical properties in this model and should be given priority to ensuring the completeness and accuracy.

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: lzz@cup.edu.cn.

*E-mail: mtk@udel.edu.

ORCID

Linzhou Zhang: 0000-0002-8354-784X

Quan Shi: 0000-0002-1363-1237

Michael T. Klein: 0000-0001-5444-1512

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (21506254) and the Science Foundation of China University of Petroleum, Beijing (2462014YJRC020). Michael T. Klein acknowledges collaborations with and support of colleagues via the Saudi Aramco Chair Program at KFUPM and Saudi Aramco.

■ REFERENCES

(1) Walsh, P.; Garbalena, M.; Schug, K. A. Rapid analysis and time interval deconvolution for comprehensive fuel compound group classification and speciation using gas chromatography−vacuum ultraviolet spectroscopy. *Anal. Chem.* **2016**, *88* (22), 11130−11138.

(2) Weber, B. M.; Walsh, P.; Harynuk, J. J. Determination of hydrocarbon group-type of diesel fuels by gas chromatography with vacuum ultraviolet detection. *Anal. Chem.* **2016**, *88* (11), 5809−5817.

(3) Teng, S. T.; Williams, A. D.; Urdal, K. Detailed hydrocarbon analysis of gasoline by GC−MS (SI-PIONA). *J. High Resolut. Chromatogr.* **1994**, *17* (6), 469−475.

(4) Gröger, T.; Gruber, B.; Harrison, D.; Saraji-Bozorgzad, M.; Mthembu, M.; Sutherland, A. e.; Zimmermann, R. A vacuum ultraviolet absorption array spectrometer as a selective detector for comprehensive two-dimensional gas chromatography: Concept and first results. *Anal. Chem.* **2016**, *88* (6), 3031−3039.

(5) Quann, R. J.; Jaffe, S. B. Building useful models of complex reaction systems in petroleum refining. *Chem. Eng. Sci.* **1996**, *51* (10), 1615−1635.

(6) Quann, R. J.; Jaffe, S. B. Structure-oriented lumping: Describing the chemistry of complex hydrocarbon mixtures. *Ind. Eng. Chem. Res.* **1992**, *31* (11), 2483−2497.

(7) Ghosh, P.; Andrews, A. T.; Quann, R. J.; Halbert, T. R. Detailed kinetic model for the hydro-desulfurization of FCC Naphtha. *Energy Fuels* **2009**, *23* (12), 5743−5759.

(8) 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.

(9) Hudebine, D.; Verstraete, J. J. Molecular reconstruction of LCO gasoils from overall petroleum analyses. *Chem. Eng. Sci.* **2004**, *59* (22−23), 4755−4763.

(10) Pan, Y.; Yang, B.; Zhou, X. Feedstock molecular reconstruction for secondary reactions of fluid catalytic cracking gasoline by maximum information entropy method. *Chem. Eng. J.* **2015**, *281*, 945−952.

(11) Peng, B. Molecular modelling of petroleum process. Ph.D. Thesis, University of Manchester Institute of Science and Technology (UMIST), Manchester, U.K., 1999.

(12) Zhang, Y. A molecular approach for characterisation and property predictions of petroleum mixtures with applications to refinery modelling. Ph.D. Thesis, University of Manchester Institute of Science and Technology (UMIST), Manchester, U.K., 1999.

(13) Aye, M. M. S.; Zhang, N. A novel methodology in transforming bulk properties of refining streams into molecular information. *Chem. Eng. Sci.* **2005**, *60* (23), 6702−6717.

(14) Hu, S.; Towler, G.; Zhu, X. Combine molecular modeling with optimization to stretch refinery operation. *Ind. Eng. Chem. Res.* **2002**, *41* (4), 825−841.

(15) Wu, Y.; Zhang, N. Molecular characterization of gasoline and diesel streams. *Ind. Eng. Chem. Res.* **2010**, *49* (24), 12773−12782.

(16) Albahri, T. A. Molecularly Explicit Characterization Model (MECM) for Light Petroleum Fractions. *Ind. Eng. Chem. Res.* **2005**, *44* (24), 9286−9298.

(17) Hudebine, D.; Verstraete, J. J. Reconstruction of Petroleum Feedstocks by Entropy Maximization. Application to FCC Gasolines. *Oil Gas Sci. Technol.* **2011**, *66* (3), 437−460.

(18) Ghosh, P.; Hickey, K. J.; Jaffe, S. B. Development of a Detailed Gasoline Composition-Based Octane Model. *Ind. Eng. Chem. Res.* **2006**, *45* (1), 337−345.

DOI: 10.1021/acs.energyfuels.7b03588