

## Novel Parameter Estimation Method for Molecular Reconstruction of Naphtha by Gamma Distribution

Yu Ren<sup>a</sup>, Zuwei Liao<sup>b,\*</sup>, Jingyuan Sun<sup>a</sup>, Binbo Jiang<sup>a</sup>, Jingdai Wang<sup>b</sup>, Yongrong Yang<sup>b</sup>, Qing Wu<sup>c</sup>

<sup>a</sup>Zhejiang Provincial Key Laboratory of Advanced Chemical Engineering Manufacture Technology, College of Chemical and Biological Engineering, Zhejiang University, Hangzhou 310027, P. R. China

<sup>b</sup>State Key Laboratory of Chemical Engineering, College of Chemical and Biological Engineering, Zhejiang University, Hangzhou 310027, P. R. China

<sup>c</sup>China National Offshore Oil Corporation, Beijing 100010, P. R. China  
[liaoZW@zju.edu.cn](mailto:liaoZW@zju.edu.cn)

Molecular reconstruction is an effective tool to predict the detailed composition of complex mixtures under obtainable bulk properties and chemical details. With the introduction of statistical distribution, although the number of variables in the model is reduced, the ranges of statistical distribution parameters have a great influence on the prediction accuracy. A novel parameter estimation method using the analysis of the probability density plot of gamma distribution based on the basic available information is proposed to obtain parameter ranges with wide universality and non-specificity. The obtained ranges are used in a molecular reconstruction model to predict the bulk properties and molecular composition of naphtha. The model assumes molecular composition within each homologous series follows a gamma distribution against carbon number. The results of a naphtha sample show that the calculated bulk properties and molecular composition are in good agreement with the experimental values.

### 1. Introduction

Facing the increasingly serious energy crisis and environmental pressure, energy conservation and consumption reduction are important issues for the chemical industry. Other than process intensification tools (Yang et al., 2018), energy integration (Hong et al., 2018) and recovery (Li et al., 2019), system optimization (Hong et al., 2019) and integration (Lou et al., 2019), molecular management is an effective alternative tool to improve refinery efficiency. The basic idea of molecular reconstruction is to determine the detailed molecular composition of petroleum fractions through obtainable bulk properties and chemical details (Ren et al., 2019). Due to the complexity of petroleum composition, there are numerous variables in the model while available constraints are insufficient. To reduce the degree of freedom, the introduction of statistical distribution, which is an important feature of petroleum, has been widely adopted. Based on the Molecular type-homologous series (MTHS) matrix model (Peng, 1999), as shown in Figure 1, Wu and Zhang (2010) assumed that the compositions of the entries within each homologous series follow a gamma distribution with boiling point. The numbers of variables to represent molecular composition are reduced from the unknown fractions of each matrix entry to three parameters for the gamma distribution and one fraction of each homologous series. Therefore, the degrees of freedom are reduced dramatically. Liu (2015) followed the above method of using gamma distribution but improved the matrix form of the MTHS model by combining carbon numbers and pseudo-components in the matrix framework. Wang and Li (2017) used normal distribution instead of the gamma distribution to describe the statistical distribution within each homologous series in the proposed novel MTHS matrix. Recently, Cui et al. (2018) also used gamma probability density function to describe the fractions of species in each chemical family with a library containing 89 hydrocarbons and 81 heteroatom-containing species in their molecular representation of the petroleum gasoline fraction.

The introduction of statistical distribution results in reducing model variables. However, it brings a problem that the boundary of variables is difficult to determine during the optimization of the model. Noted that the tuning parameters for the optimization engine at this time are no longer the fractions of each molecule whose boundary fall within [0, 1], but the parameters of the statistical distribution whose boundary are wide enough to reach infinity (Cui et al., 2018). Therefore, reasonable ranges of parameters are critical to the accuracy of composition prediction. Wu and Zhang (2010) stated that to avoid the distribution between molecules within each homologous series being too narrow or too wide, the scale parameter  $\alpha$  and the shape parameter  $\beta$  of gamma distribution are limited in the range of 1~20 and 19~100, respectively. However, the authors did not indicate how these ranges were determined. Wang and Li (2017) summarized the ranges of normal distribution parameters by analyzing the parameters of 16 naphtha and gasoline samples. These ranges were used to generate the initial value of distribution parameters, and the prediction calculation was optimized by the multi-population genetic optimization algorithm in this range too. By this way, the authors guaranteed the initial value and optimal value of distribution parameters obtained in the correct range, which could decrease the predicting error and calculating time largely.

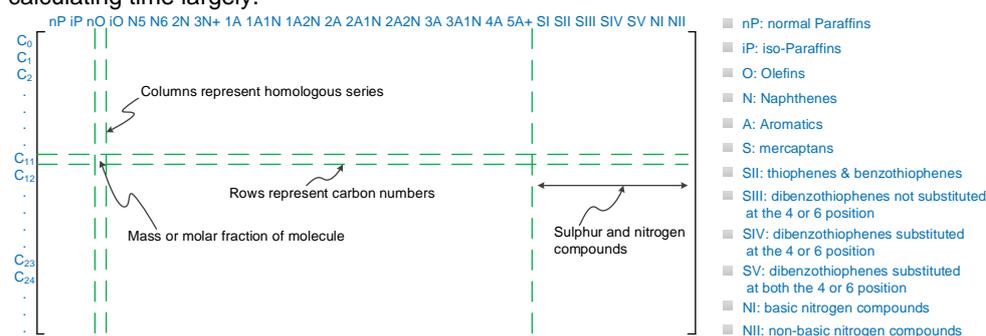


Figure 1: Molecular type-homologous series (MTHS) matrix (adapted from Wu and Zhang (2010)).

In summary, the ranges of statistical distribution parameters and the estimation method are important in molecular reconstruction. If the initial ranges are obtained by parameter regression of existing molecular composition data of sample oils, it will result that a set of parameters can only be applied to specific oils. In order to solve the parameter estimation problem and ensure that the parameters ranges obtained have wide universality and are not specific to any oil characteristics, a novel complete parameter estimation procedure is proposed in this article. The pre-set ranges of parameters are determined from the analysis of the probability density plot of gamma distribution based on the basic available information. Using the range of parameters obtained above, the molecular reconstruction of a naphtha sample shows that the accuracy of the calculated bulk properties and molecular composition is within an acceptable range.

## 2. Parameter estimation

The gamma distribution function  $\text{gampdf}(x; \alpha, \beta)$  is a two-parameter family of continuous probability distributions. The parameter  $\alpha$  is called shape parameter, and the parameter  $\beta$  is called scale parameter:

$$\text{gampdf}(x; \alpha, \beta) = \frac{x^{\alpha-1} e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)} \quad \text{for } x > 0 \text{ and } \alpha, \beta > 0 \quad (1)$$

Where  $x$  is a random variable;  $\Gamma(\alpha)$  is the gamma function.

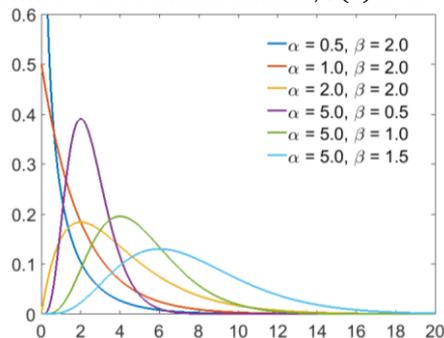


Figure 2: Probability density plots of gamma distributions.

Figure 2 shows the probability density plots of gamma distributions. As can be seen from Figure 2, the parameters  $\alpha$  and  $\beta$  simultaneously control the shape of the gamma distribution. According to the practical issues of molecular reconstruction, the shape of probability density plots of gamma distributions can be restricted according to the characteristics of the model. In this paper, it is assumed that the molecular composition within each homologous series follows a gamma distribution against carbon number, e.g., the random variable  $x$  in Eq(1) stands for the carbon number of each matrix entry. Therefore, according to the common knowledge of naphtha composition distribution, the following constraints can be obtained.

$$\alpha_j > 1 \quad (2)$$

$$\beta_j > 0 \quad (3)$$

$$\text{gampdf}(\text{CN}_j^{\text{min}}; \alpha_j, \beta_j) > 0 \quad (4)$$

$$\text{gampdf}(\text{CN}_j^{\text{max}}; \alpha_j, \beta_j) > 0 \quad (5)$$

$$\text{CN}_j^{\text{min}} < (\alpha_j - 1)\beta_j < \text{CN}_j^{\text{max}} \quad (6)$$

$$\text{gampdf}((\alpha_j - 1)\beta_j, \alpha_j, \beta_j) < 1 \quad (7)$$

Eq(2) and Eq(3) are the lower limits of parameters  $\alpha$  and  $\beta$  of the homologous series  $j$ , respectively. Only when  $\alpha > 0$ , the gamma distribution is a unimodal function, which satisfies the requirements of the model. Eq(4) and Eq(5) control that the function values of the starting point ( $\text{CN}_j^{\text{min}}$ ) and ending point ( $\text{CN}_j^{\text{max}}$ ) of each plot must be larger than 0. These constraints are to ensure that gamma distribution is effective in fitting molecular composition without too many zeros between the range of minimum carbon number ( $\text{CN}_j^{\text{min}}$ ) and maximum carbon number ( $\text{CN}_j^{\text{max}}$ ).  $(\alpha_j - 1)\beta_j$  is the extreme point obtained by deriving the gamma distribution function. Eq(6) defines that the peak position of gamma distribution must fall within the range of maximum and minimum carbon number in order to avoid monotonous increase or decrease in the studied range. Eq(7) indicates that the maximum value of gamma distribution does not exceed 1, to match the fact that the molecular fractions in a homologous series are between 0 and 1. The carbon number range for the five homologous series divided in this paper is as follows: C<sub>4</sub>~C<sub>12</sub> for n-Paraffins and iso-Paraffins, C<sub>4</sub>~C<sub>7</sub> for Olefins, C<sub>5</sub>~C<sub>11</sub> for Naphthenes, C<sub>6</sub>~C<sub>11</sub> for Aromatics.

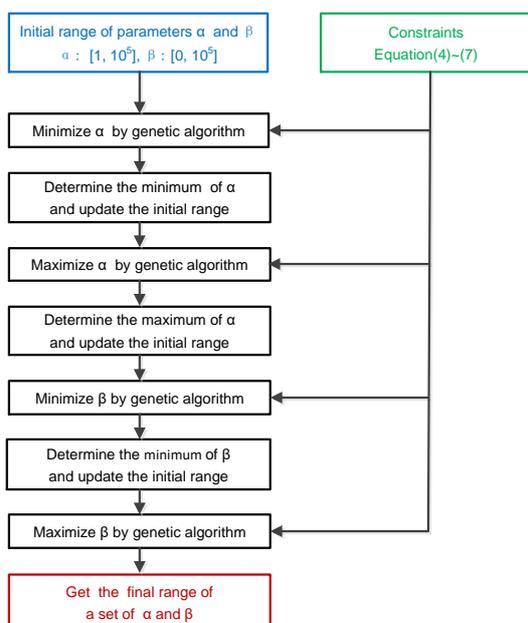


Figure 3: Flowsheet of parameter estimation.

The process of parameter estimation is shown in Figure 3. The genetic algorithm is used to solve this nonlinear programming problem. In the above constraints, there is only a lower limit of the parameters  $\alpha$  and  $\beta$ . To reasonably control the solution space of the model, the upper limit is set to  $10^5$ . Based on such initial range, the objective function is to minimize the parameter  $\alpha$  firstly under the above constraints, e.g., Eq(4) ~ (7). The optimization process is repeated 20 times and the minimum value of  $\alpha$  is taken as the lower limit of parameter  $\alpha$ . Then, the obtained minimum of  $\alpha$  is used to replace the lower limit of  $\alpha$  in the initial range, thereby realizing the update of the initial range. Subsequently, the optimization processes of the maximum of  $\alpha$ , the minimum of  $\beta$ , and the maximum of  $\beta$  are sequentially performed. The operation of each optimization process is the same. Finally, the pre-set ranges of gamma distribution parameters  $\alpha$  and  $\beta$  of each homologous series as shown in Table 1 are obtained.

Table 1: The pre-set range of parameters  $\alpha$  and  $\beta$  of the gamma distribution.

Parameters	Lower limit	Upper limit
$\alpha_{nP}$	1.04	87.00
$\beta_{nP}$	0.05	118.61
$\alpha_{iP}$	1.04	87.01
$\beta_{iP}$	0.05	118.61
$\alpha_O$	1.05	96.02
$\beta_O$	0.05	99.74
$\alpha_N$	1.05	106.39
$\beta_N$	0.05	100.79
$\alpha_A$	1.06	115.63
$\beta_A$	0.05	99.33

### 3. Mathematical model

The formulas to describe the assumption gamma distribution are as follows.

$$y_{i,j} = \frac{CN_{i,j}^{\alpha_j-1} e^{-CN_{i,j}/\beta_j}}{\beta_j^{\alpha_j} \Gamma(\alpha_j)} \quad (8)$$

$$x_{i,j} = X_j \frac{y_{i,j}}{\sum_{i,i} y_{i,i}} \quad (9)$$

where  $y_{i,j}$  in Eq(8) stands for the fraction of the entry  $i, j$  within the  $j$  homologous series;  $CN_{i,j}$  is the carbon number of the entry  $i, j$ ;  $x_{i,j}$  in Eq(9) stands for the mass fraction of the entry  $i, j$  in the whole matrix;  $X_j$  is the total mass fraction of the homologous series  $j$ .

$$\text{obj} = \sum_T (w_{2,T} \frac{V_T^{\text{exp}} - V_T^{\text{cal}}}{V_T^{\text{exp}}})^2 + \sum_{j \in \text{PIONA}} (w_{3,j} \frac{F_j^{\text{exp}} - F_j^{\text{cal}}}{F_j^{\text{exp}}})^2 + \sum_P (w_{1,P} \frac{p^{\text{exp}} - p^{\text{cal}}}{p^{\text{exp}}})^2 \quad (10)$$

where  $\text{obj}$  is the objective function,  $T$  is the distillation curve temperature,  $V$  is the volume fraction at the corresponding distillation curve temperature  $T$ ,  $F$  stands for n-Paraffins, iso-Paraffins, Olefins, Naphthenes, and Aromatics fractions,  $P$  stands for other available bulk properties, and  $w$  is the weight factor.

$$CN_{C8,A} \leq (\alpha_A - 1)\beta_A \leq CN_{C9,A} \quad (11)$$

Eq(11) is a constraint according to refinery reality to constrain the fraction of C8 or C9 for naphtha fraction is the maximum within the aromatic homologous series (Wu and Zhang, 2010).

Particle swarm optimization (PSO) is used to solve the above mathematical model. The whole calculation processes including parameter estimation are executed by MATLAB R2016b on a PC constituted by Intel Xeon E31230 processor 3.20 GHz.

### 4. Result and discussion

A naphtha sample is used to verify the accuracy of the above mathematical model with the obtained parameters ranges in Table 1. The available bulk properties including molecular weight, special-density, PIONA weight fractions and ASTM D 86 distillation data are shown in Table 2.

Table 3 shows the comparison of experimental and calculated values of molecular weight and special-density. The relative errors of these two properties are both less than 1.2 %, indicating that the calculated values agree well with the experimental values. Figure 4 shows the comparison between experimental and calculated data of

simulated distillation curve. It can be seen that after 20 % distillation volume, the simulated distillation curve agrees well with the true boiling point distillation curve with the relative error of less than 5 %.

Table 2: Bulk properties data of the naphtha sample.

Bulk properties			
Molecular weight	101.32	Distillation (ASTM D 86) °C	
		IBP	33.04
Special-density [kg / m <sup>3</sup> ]	715.85	5 %	51.15
		10 %	71.28
n-Paraffins (wt %)	28.89	20 %	80.01
		30 %	88.75
iso-Paraffins (wt %)	39.59	40 %	94.75
		50 %	100.75
Olefins (wt %)	0.26	60 %	110.98
		70 %	121.21
Naphthenes (wt %)	23.54	80 %	138.5
		90 %	155.78
Aromatics (wt %)	7.72	95 %	174.75
		FBP	189.41

Table 3: Comparison of experimental and calculated properties.

Properties	Experiment	Calculation	Relative error
Molecular weight	101.32	102.53	-1.19 %
Special-density kg / m <sup>3</sup>	715.85	717.02	-0.16 %

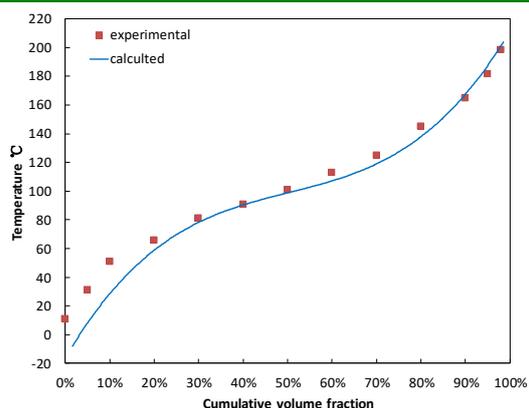


Figure 4: Comparison between experimental and calculated data of simulated distillation curve.

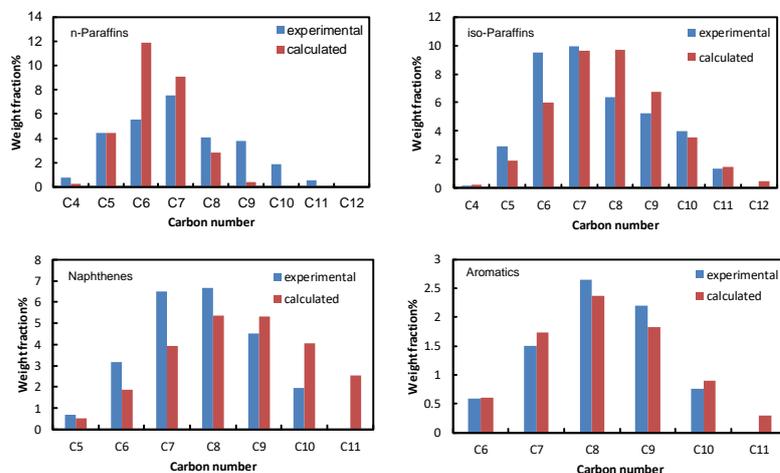


Figure 5: Experimental values and calculated values of the molecular composition of the naphtha sample.

The predicted results of molecular composition are shown in Figure 5. The figure shows the comparison of the experimental and calculated molecular composition of n-Paraffins, iso-Paraffins, Naphthenes and Aromatics. Olefins are neglected for their low content. The mean absolute differences (abbreviated as MAD,  $MAD = \frac{1}{n} \sum_{i=1}^n |x_i^{cal} - x_i^{exp}|$ ) of the calculated mass fractions of all components is 1.1, which is within an acceptable range. The prediction results for aromatics are more accurate due to the additional constraint on the peak position. Although there are errors in the specific content, an approximate molecular composition simulation result is obtained. To some extent, the prediction accuracy of sample oils with compositional distributions more in line with the statistical law of the gamma distribution will be higher.

## 5. Conclusions

A novel parameter estimation method is proposed for the molecular reconstruction model with the statistical distribution introduced. The model is based on the Molecular type-homologous series (MTHS) matrix method where the molecular composition within each homologous series is assumed to follow a gamma distribution against carbon number. The pre-set ranges of gamma distribution parameters are calculated from the analysis of the probability density plot of gamma distribution based on the basic available information, so that the obtained parameter ranges have wide universality and are not specific to any oil characteristics. A naphtha sample is used to verify the model with the obtained parameters range calculated by the proposed parameter estimation method. The relative errors between experimental and calculated values of molecular weight and special-density are less than 1.2 %, and that of the simulated distillation curve is less than 5 % after 20 % distillation volume. The accuracy of the calculated mass fractions of molecular composition is also within an acceptable range with the MAD of 1.1. This work proves that the approximate bulk properties and molecular composition prediction could also be obtained without parameter regression but only through analyzing the features of the statistical distribution plot. In order to obtain better prediction accuracy, more implicit graphical features need to be excavated.

## Acknowledgments

The financial support provided by the Project of National Natural Science Foundation of China (21822809 & 61590925), the National Science Fund for Distinguished Young (21525627) are gratefully acknowledged.

## References

- Cui, C., Billa, T., Zhang, L., Shi, Q., Zhao, S., Klein, M.T., Xu, C., 2018, Molecular Representation of the Petroleum Gasoline Fraction, *Energy and Fuels*, 32, 1525-1533.
- Hong, X., Liao, Z., Lou, Y., Sun, J., Jiang, B., Wang, J., Yang, Y., 2018, Transshipment type model for total site heat integration using non-isothermal utility loop, *Chemical Engineering Transactions*, 70, 1135-1140.
- Hong, X., Liao, Z., Sun, J., Jiang, B., Wang, J., Yang, Y., 2019, Transshipment type heat exchanger network model for intra- and inter-plant heat integration using process streams, *Energy*, 178, 853-866.
- Li, H., Liao, Z., Sun, J., Jiang, B., Wang, J., Yang, Y., 2019, Modelling and simulation of two-bed PSA process for separating H<sub>2</sub> from methane steam reforming, *Chinese Journal of Chemical Engineering*, Accepted.
- Liu, L., 2015, Molecular characterisation and modelling for refining processes, PhD Thesis, University of Manchester Institute of Science and Technology, Manchester, UK.
- Lou, Y., Liao, Z., Sun, J., Jiang, B., Wang, J., Yang, Y., 2019, A novel two-step method to design inter-plant hydrogen network, *International Journal of Hydrogen Energy* 44, 5686-5695.
- Ren, Y., Liao, Z., Sun, J., Jiang, B., Wang, J., Yang, Y., Wu, Q., 2019, Molecular reconstruction: Recent progress toward composition modeling of petroleum fractions, *Chemical Engineering Journal*, 357, 761-775.
- Van de Vijver, R., Devocht, B.R., Van Geem, K.M., Thybaut, J.W., Marin, G.B., 2016, Challenges and opportunities for molecule-based management of chemical processes, *Current Opinion in Chemical Engineering*, 13, 142-149.
- Wang, K., Li, S., 2017, Modified molecular matrix model for predicting molecular composition of naphtha, *Chinese Journal of Chemical Engineering*, 25, 1856-1862.
- Wu, Y., Zhang, N., 2010, Molecular Characterization of Gasoline and Diesel Streams, *Industrial & Engineering Chemistry Research*, 49, 12773-12782.
- Yang, Y., Sun, J., Huang, Z., Hu, D., Liao, Z., Yang, Y., Wang, J., 2018, Dynamic and Steady-State Characterization of the Liquid Spray Zone in an Externally Heated Gas-Solid Fluidized Bed, *Industrial and Engineering Chemistry Research*, 57, 2988-3001.