Molecular Reconstruction Method of Petroleum Based on Dual Two-Dimensional Homologous Series Distributions

Chen Zhang,a,b Yijiong Hu,c Tong Qiu,a,b

a Institute of Process Systems Engineering, Department of Chemical Engineering, Tsinghua University, Beijing 100084, China

b Beijing Key Laboratory of Industrial Big Data System and Application, Tsinghua University, Beijing 100084, China

c PetroChina Planning and Engineering Institute, Beijing 100083, China

qiutong@tsinghua.edu.cn

Abstract

Molecular reconstruction methods, based on rapidly measurable petroleum properties, leverage knowledge of molecular homologous compositions and distribution patterns to determine the detailed molecular composition of petroleum. These methods have received widespread attention and research. In this study, utilizing molecular distribution data obtained from experimental analysis, molecular distribution patterns under the dimension of double bond equivalent (DBE) are proposed. By quantitatively expressing the proposed qualitative molecular distribution patterns using Jensen-Shannon Divergence (JSD) and integrating them into the existing two-dimensional homologous series distribution (2D-HSD) method, the Dual-2D-HSD method was constructed. On the molecular composition level, the Dual-2D-HSD method provides more realistic and reasonable molecular distribution results compared to the traditional 2D-HSD method. This highlights the importance of incorporating molecular distribution patterns in the molecular reconstruction process.

**Keywords**: Molecular reconstruction, Homologous series distribution, Double bond equivalent.

* 1. Introduction

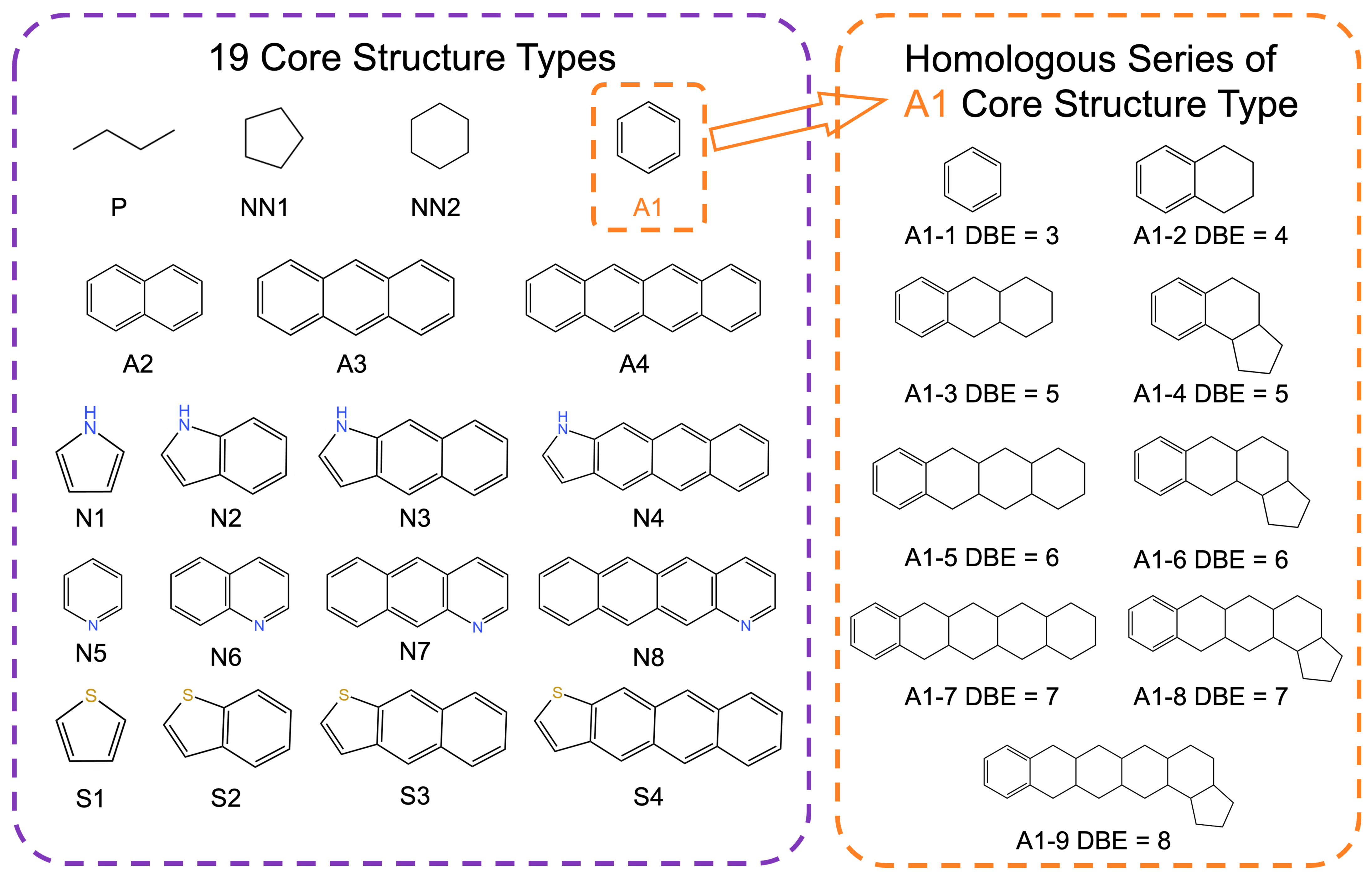
As the petroleum refining industry moves towards digitization and automation, there's a need for a detailed digital representation of the refining process for effective molecular management. The cornerstone of this process is accurately identifying every molecule in petroleum, primarily through experimental analysis like High-Resolution Mass Spectrometry (HRMS), which offers a quantitative view of molecules based on double bond equivalent (DBE) distributions (Zhang et al., 2014). However, such experimental methods are costly and time-consuming. Alternatively, molecular reconstruction methods, utilizing petroleum's rapidly measurable physical properties, provide a quicker and cost-effective way to determine petroleum's molecular composition, essential for molecular management. This approach has gained significant research interest and practical application in the industry.

As research progresses, molecular reconstruction models have evolved to incorporate increasing amounts of knowledge about molecular composition and distribution. The integration of this knowledge has significantly improved the accuracy of computational results. Early models focused on variables like molecule mole fractions without considering molecular distribution. Later, Stochastic Reconstruction (SR) methods were developed, where molecules are constructed using specific methods and probability density functions (PDFs) to describe structural attributes (Hudebine and Verstraete, 2004). The molecular type-homologous series (MTHS) method further enhanced these models by introducing homologous series, thus reducing optimization variables and increasing computational efficiency (Wu and Zhang, 2010). Recent advancements include the integration of two-dimensional homologous series distributions and the cloud model (CM) for better handling molecular distribution uncertainty (Zhang et al., 2022). This has led to more precise molecular libraries and distribution fluctuation descriptions. High-Resolution Mass Spectrometry (HRMS) data now provide DBE distribution data for vacuum gas oil (VGO) and heavier oil products in reconstruction processes (Guan et al., 2023). However, there is still limited research on DBE distribution patterns of petroleum components, and current models have not fully utilized these patterns from HRMS.

This article presents a new molecular reconstruction method utilizing dual two-dimensional homologous series distributions (Dual-2D-HSD), focusing on core structure-side chain distribution and DBE-carbon number distribution. The first 2D-HSD, based on core structure-side chain distribution, creates a molecular library essential for reconstruction. The second 2D-HSD, based on the DBE distribution pattern from HRMS, uses the JSD method to measure distribution consistency, thereby quantitatively representing the proposed distribution patterns and enhancing the alignment of molecular distribution results with actual conditions. Combining these dual 2D-HSD approaches in the optimization process enhances property prediction accuracy and aligns with real molecular DBE distribution patterns. This method marks a significant advancement in molecular reconstruction, particularly in efficiently handling complex molecular structures and addressing challenges in scenarios with limited data.

* 1. Methods
     1. Building a Molecular Library Using the First 2D-HSD.

The initial step of the molecular reconstruction method involves creating a molecular library using the first 2D-HSD and the structure-oriented lumping (SOL) method with 21 structural increments. This two-step process first establishes core structures, such as the A1-1 core structure in Figure 1, and then attaches side chains of varying lengths to each core, forming the complete molecular library. This library is crucial for later stages of molecular reconstruction. The first 2D-HSD mapped the molecular library and molar fractions of complex petroleum components, identifying 19 core structure types and 172 core structures. Side chains with 1 to 50 carbon atoms were added to each core structure, forming a library of 8600 molecules. The first dimension HSD detailed the distribution of core structures per type, while the second dimension HSD showed side chain length distribution per core structure, using the gamma distribution for characterization.



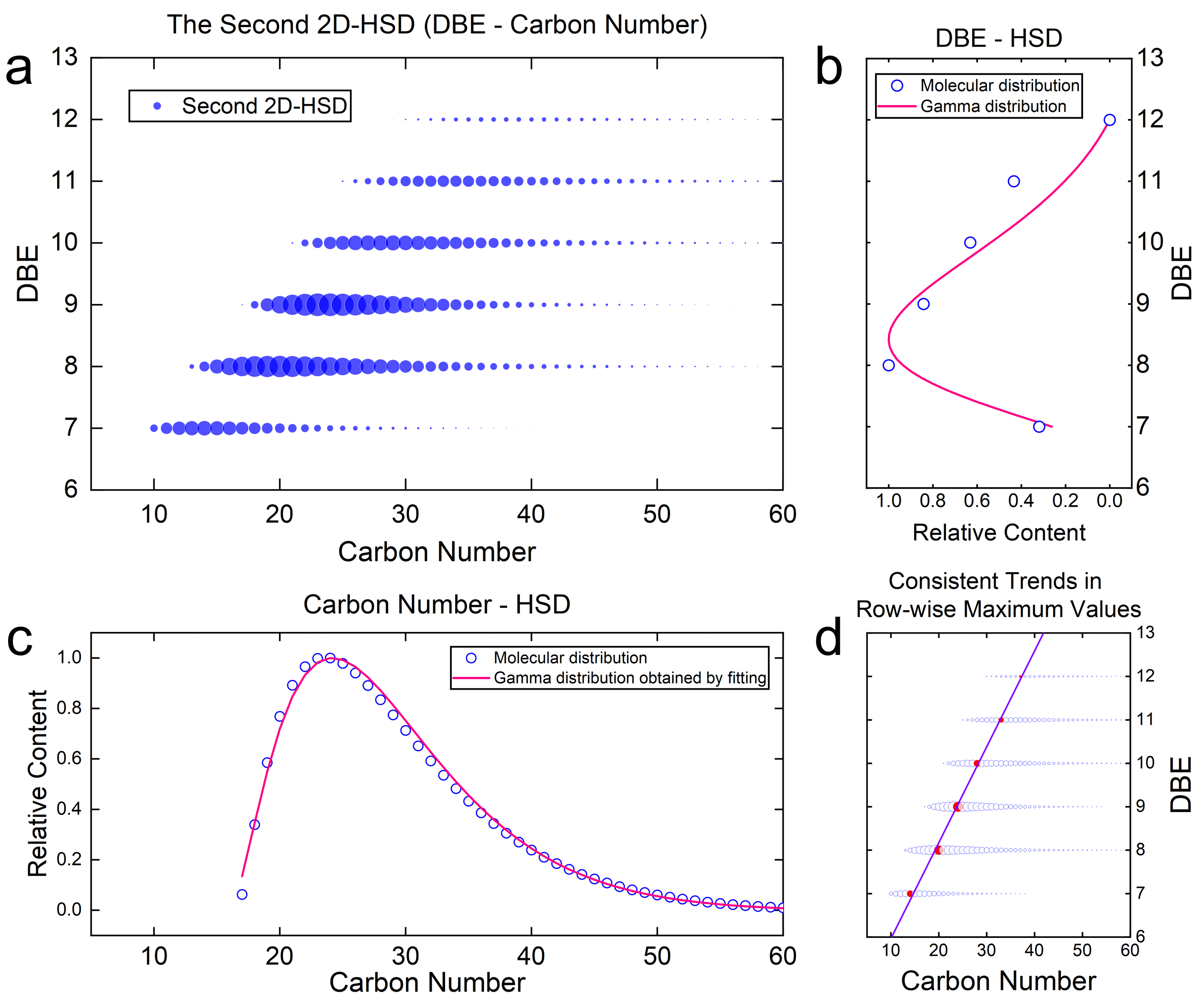
**Figure 1.** 19 Core structure types and homologous series of the A1 core structure type.

* + 1. The Second 2D-HSD and the DBE Distribution Patterns

Due to limitations in the precision of analytical techniques, existing HRMS can only provide molecular abundance distribution data represented in terms of the DBE dimension for complex petroleum components. As a result, the distribution patterns derived from HRMS are applicable in the DBE dimension. The DBE value for each molecule can be calculated from its structural formula using the Eq. (1), where C is carbon, H is hydrogen, X is halogen atoms, and N is nitrogen.

|  |  |
| --- | --- |
|  | (1) |

Using the molecular library from the first 2D-HSD and given molar fractions, two-dimensional molecular abundance graphs in DBE and carbon number dimensions are created. Figure 2a shows the DBE dimension molecular abundance for N6 core structure type. Figure 2b, the vertical distribution from 2a, displays the relative abundance for varying DBE values, with blue points summing up molecular abundances per DBE row in 2a, and red lines indicating fitted gamma distribution curves. Figure 2c shows relative abundance for DBE = 9 in 2a, depicting carbon number distribution for molecules with identical DBE values. Figure 2d marks the maximum values in each row from 2a with red dots and a purple trend line, illustrating the change trend in maximum values per row.



**Figure 2.** (a) The second 2D-HSD diagram (in core structure type N6). (b) Distribution of molecular relative abundance along the DBE axis (vertical direction in figure 2a). (c) Distribution of molecular relative abundance along the carbon number axis (horizontal direction for DBE=9 in figure 2a). (d) The trend of the maximum points of each row.

Based on HRMS data (Guan et al., 2021) on molecular distribution in petroleum components, the DBE representation reveals: (a) Summing molecules with identical DBE values shows their cumulative molecular abundance in the DBE dimension follows a gamma distribution, depicted in Figure 2b. (b) For a specific DBE value, molecular abundance distribution in the carbon number direction also follows a gamma distribution, as shown in Figure 2c. (c) In the DBE versus carbon number two-dimensional graph, each row (representing the same DBE value) has maximum abundance points that follow a consistent trend, shifting gradually towards the upper-right direction, as demonstrated in Figure 2d.

* + 1. Optimization Process for Molecular Reconstruction Based on Dual-2D-HSD

To optimize molecular reconstruction in the first 2D-HSD level, the goal is to minimize discrepancies between calculated and experimental properties. This involves using a weighted average of the absolute relative deviations of physical properties, as outlined in Eq. (2), with the optimization variables being the distribution's parameters.

|  |  |
| --- | --- |
|  | (2) |

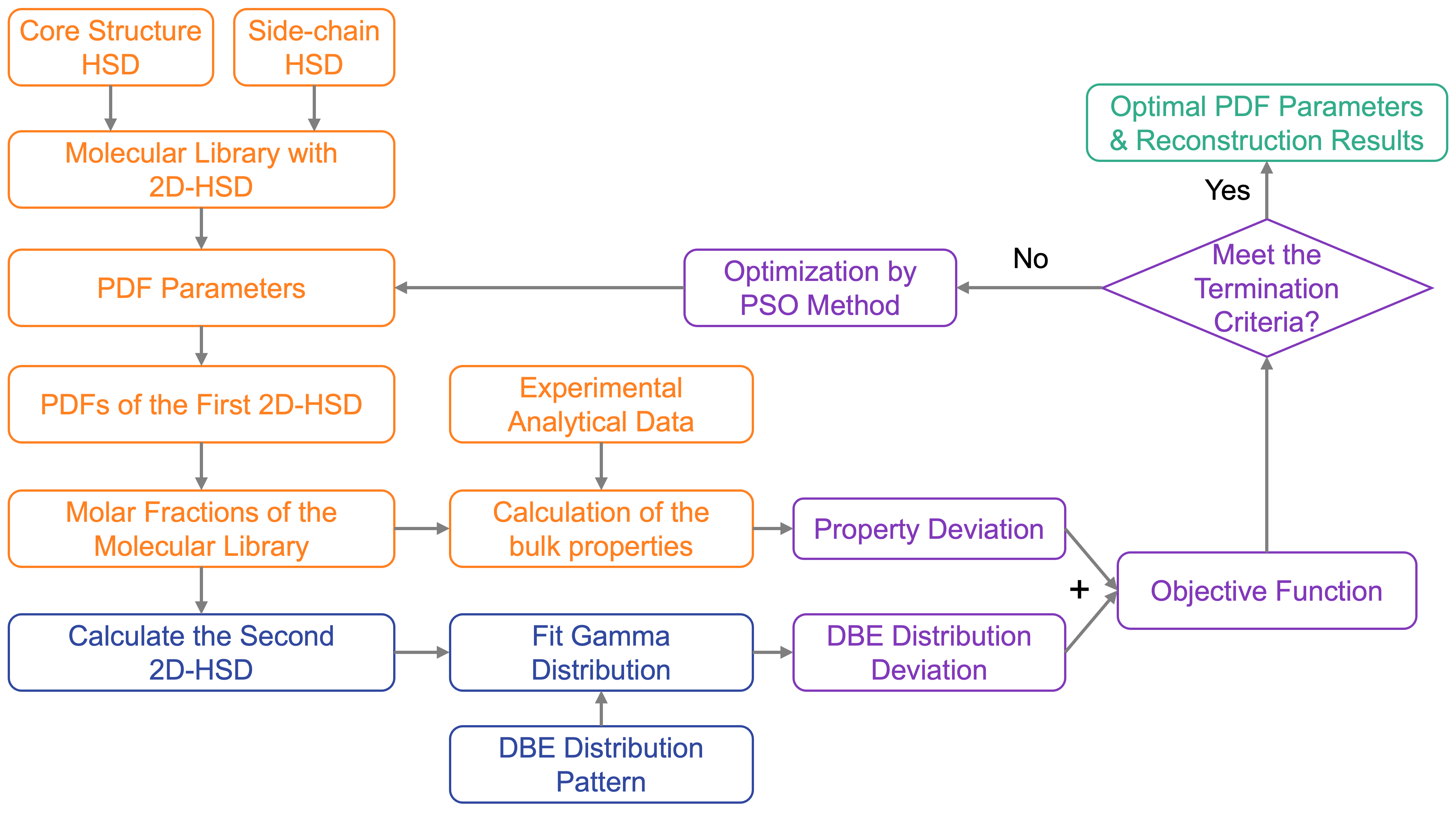
In the second 2D-HSD level for molecular reconstruction, the aim is to quantitatively match the molecular abundance distribution within each core structure type with existing distribution patterns. The objective function evaluates the difference between the actual molecular abundance and the proposed patterns. In specific calculations: (1) For the proposed molecular distribution patterns (a) and (b), corresponding gamma distributions are fitted based on the actual distribution data. The dissimilarity is measured using the Jensen-Shannon Divergence (JSD), derived from the Kullback-Leibler Divergence (KLD), as detailed in Eq. (3) to (5). Here, P is the gamma distribution from fitting, and Q is the actual distribution. A smaller JSD indicates a closer match to the proposed patterns.

|  |  |
| --- | --- |
|  | (3) |
|  | (4) |
|  | (5) |

(2) For molecular distribution pattern (c) in the second 2D-HSD level, linear correlation (R2) evaluates the trend consistency of maximum points in each row. Higher R2 means more consistent trends, indicating a closer fit to the proposed pattern. The objective function, detailed in Eq. (6), includes three terms: the first two assess agreement with the molecular distribution in horizontal and vertical directions, while the third term measures the trend consistency in maximum value points across rows.

|  |  |
| --- | --- |
|  | (6) |

The objective function for the optimization problem is obtained by summing the respective objective functions for the first and second 2D-HSD levels, and the optimization objective is to minimize the objective function. The particle swarm optimization (PSO) algorithm is used as the optimization algorithm, and the overall computation and optimization workflow are depicted in Figure 3.



**Figure 3.** the optimization process and the calculation of the objective function based on the Dual 2D-HSD.

* 1. Results and discussion

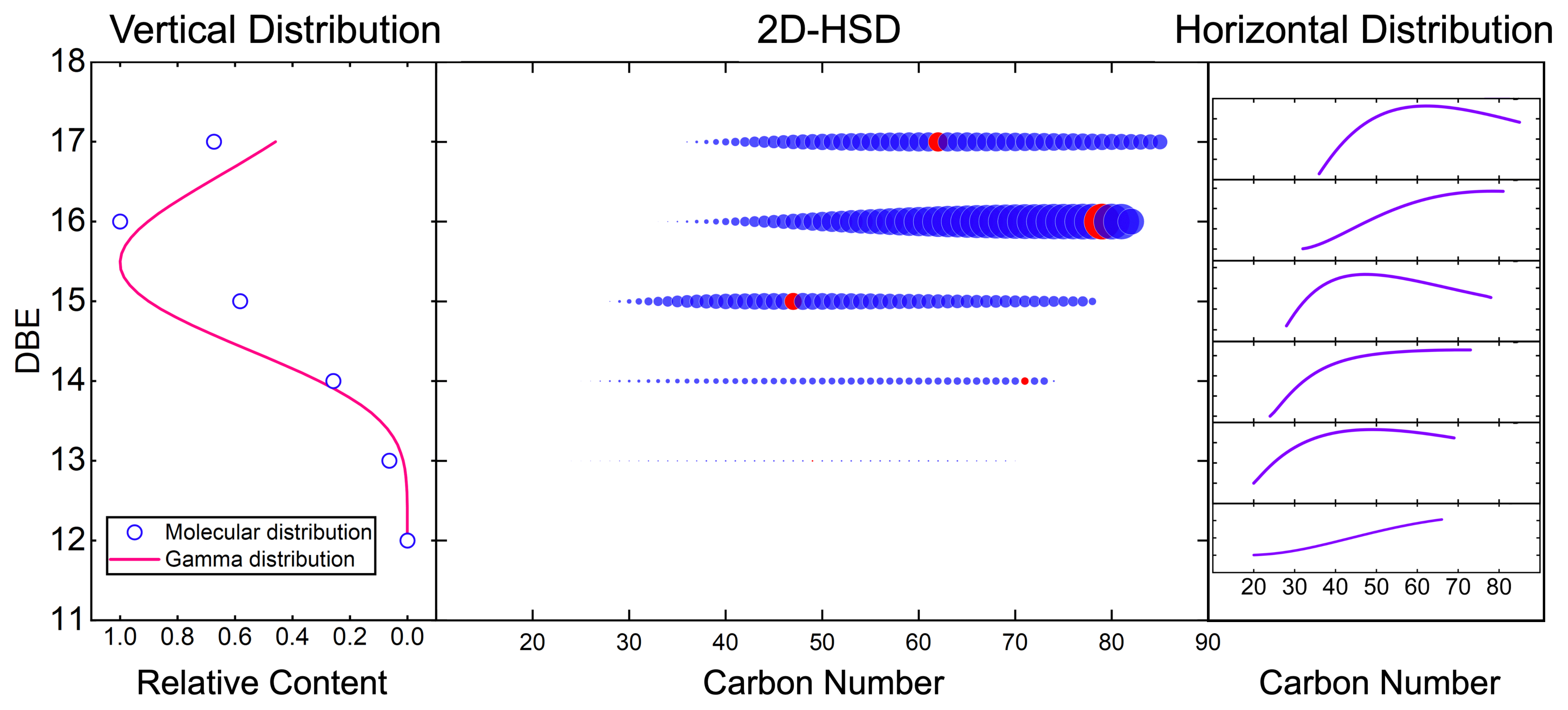
The Dual-2D-HSD molecular reconstruction model was applied to VGO's molecular reconstruction, alongside a traditional 2D-HSD model for comparison. The optimization problem was solved using MATLAB 2023a and the PSO algorithm with a particle swarm of 100. Results from both methods are in Table 1, with the absolute relative error (ARE) measuring computational deviations for each physical property.

**Table 1.** Calculation results of the 2D-HSD and the Dual-2D-HSD

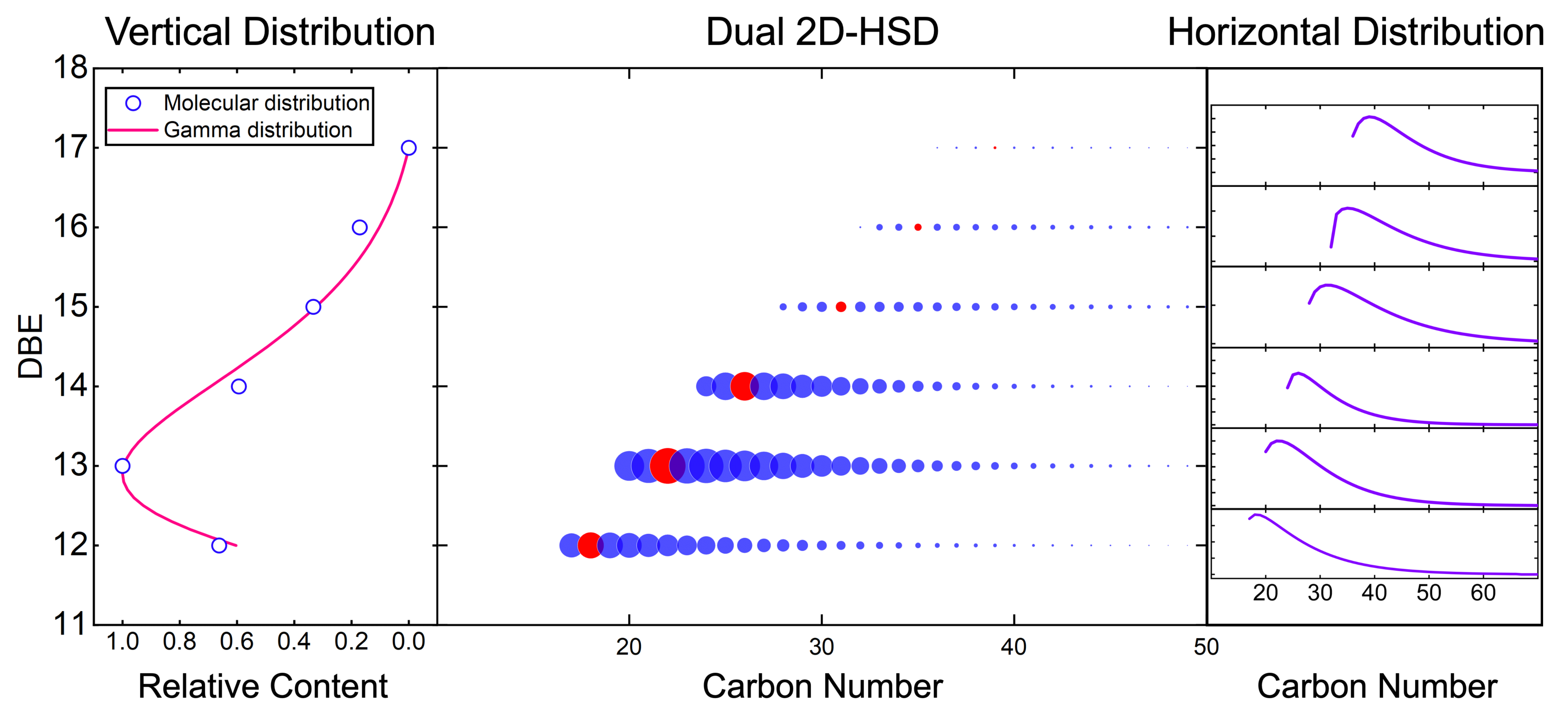
|  |  |  |
| --- | --- | --- |
|  | 2D-HSD | Dual 2D-HSD |
| Optimization Time | 769s (100%) | 316s (41.1%) |
| Property | ARE of the Calculation Result | |
| Carbon | 0.23 % | 0.24 % |
| Hydrogen | 1.64 % | 1.53 % |
| Sulfur | 0.00 % | 0.00 % |
| Nitrogen | 0.00 % | 0.00 % |
| Paraffin | 0.00 % | 0.00 % |
| Naphthene | 0.00 % | 0.00 % |
| Aromatic | 0.00 % | 0.00 % |
| 10 vol% BP | 0.56 % | 2.13 % |
| 30 vol% BP | 1.49 % | 1.89 % |
| 50 vol% BP | 0.38 % | 0.96 % |
| 70 vol% BP | 0.01 % | 1.71 % |
| 90 vol% BP | 1.77 % | 3.09 % |
| 95 vol% BP | 2.11 % | 2.74 % |
| MARE | 0.63 % | 1.10 % |

The Dual-2D-HSD method showed a mean absolute relative error (MARE) of 1.10%, compared to 0.63% for the 2D-HSD method. The computation time for Dual-2D-HSD was only 41.1% of that required for 2D-HSD. This indicates that including molecular distribution knowledge reduces the optimization space and speeds up the algorithm, improving the potential of molecular reconstruction algorithms for industrial use.

To evaluate how the proposed molecular distribution patterns affect optimization, we extracted molecular abundance data for the core molecule type S4 from both 2D-HSD and Dual-2D-HSD methods. The results are shown in Figures 4 and 5, respectively. Each figure features a central molecular abundance distribution chart with DBE and carbon number coordinates. The left side shows the vertical molecular abundance distribution, and the right side displays the horizontal distribution for each DBE value.



**Figure 4**. Molecular distribution obtained from the 2D-HSD method.



**Figure 5**. Molecular distribution obtained from the Dual-2D-HSD method.

Figure 4 shows that without molecular distribution pattern knowledge, the 2D-HSD method's molecular reconstruction optimization lacks clear distribution patterns in both horizontal and vertical directions. There are significant variances in molecular abundance and distribution across different rows, with a higher concentration of molecules at larger DBE values, contradicting experimental findings. Additionally, the maximum abundance points in each row are sporadically placed, not aligning with the proposed distribution pattern (c). In contrast, Figure 5 illustrates that incorporating molecular distribution pattern knowledge in the Dual-2D-HSD method results in a clearer gamma distribution both horizontally and vertically. The molecular distribution skews towards lower DBE values, matching real-world distributions more closely. Furthermore, the maximum abundance points in each row show a consistent upward-rightward trend. Thus, the Dual-2D-HSD method achieves more realistic molecular abundance distributions, aligning better with the distribution patterns than the 2D-HSD method.

* 1. Conclusions

This study enhances the existing 2D-HSD molecular reconstruction method by adding a second 2D-HSD representation in the DBE dimension, creating the Dual-2D-HSD method. By utilizing JSD to measure the deviation between distributions, it quantitatively incorporates molecular distribution patterns derived from HRMS data into this new representation. The Dual-2D-HSD method not only yields accurate computational results more quickly but also provides more realistic molecular distributions than the traditional 2D-HSD method. This approach's efficiency, accuracy, and rationality have been proven in reconstructing VGO molecules, showing its potential for wider industrial application to improve factory process simulations.

References

D. Guan, G. Cai, L. Zhang, 2023. Dual-objective optimization for petroleum molecular reconstruction based on property and composition similarities. AIChE Journal 69, e18108.

D. Guan, Z. Chen, X. Chen, Y. Zhang, Q. Qi, Q. Shi, S. Zhao, C. Xu, L. Zhang, 2021. Molecular-level heavy petroleum hydrotreating modeling and comparison with high-resolution mass spectrometry. Fuel 297, 120792.

Hudebine, D., Verstraete, J.J., 2004. Molecular reconstruction of LCO gasoils from overall petroleum analyses. Chemical Engineering Science 59, 4755–4763.

Y. Wu, N. Zhang, 2010. Molecular Characterization of Gasoline and Diesel Streams. Ind. Eng. Chem. Res. 49, 12773–12782.

C. Zhang, K. Bi, T. Qiu, 2022. Molecular Reconstruction of Crude Oil: Novel Structure-Oriented Homologous Series Lumping with a Cloud Model. Ind. Eng. Chem. Res. 61, 18810–18820.

Y. Zhang, L. Zhang, Z. Xu, N. Zhang, K.H. Chung, S. Zhao, C. Xu, Q. Shi, 2014. Molecular Characterization of Vacuum Resid and Its Fractions by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry with Various Ionization Techniques. Energy Fuels 28, 7448–7456.