Taylor & Francis
Taylor & Francis Group

Molecular Physics

An International Journal at the Interface Between Chemistry and Physics

ISSN: 0026-8976 (Print) 1362-3028 (Online) Journal homepage: www.tandfonline.com/journals/tmph20

Second-order group contribution method for T_c, P_c, ω, , and liquid densities of linear and branched alkanes

Ziyan Li, Leonidas Constantinou, Richard Baur, David Dubbeldam, Sofia Calero, Shrinjay Sharma, Marcello Rigutto, Poulumi Dey & Thijs J.H. Vlugt

To cite this article: Ziyan Li, Leonidas Constantinou, Richard Baur, David Dubbeldam, Sofia Calero, Shrinjay Sharma, Marcello Rigutto, Poulumi Dey & Thijs J.H. Vlugt (2025) Second-order group contribution method for T_c, P_c, ω, , and liquid densities of linear and branched alkanes, Molecular Physics, 123:21-22, e2566763, DOI: 10.1080/00268976.2025.2566763

To link to this article: https://doi.org/10.1080/00268976.2025.2566763

MOLECULAR PHYSICS
2025, VOL. 123, NOS. 21–22, e2566763 (16 pages)
https://doi.org/10.1080/00268976.2025.2566763

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

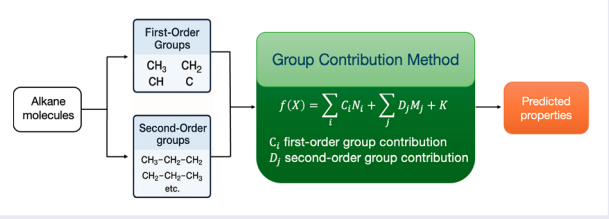
Second-order group contribution method for T_c, P_c, ω, ΔG_f^0, ΔH_f^0 and liquid densities of linear and branched alkanes

Ziyan Li[a,b], Leonidas Constantinou[c], Richard Baur[d], David Dubbeldam[e], Sofia Calero[f], Shrinjay Sharma[a,f], Marcello Rigutto[c], Poulumi Dey[b] and Thijs J.H. Vlugt [a]

[a]Engineering Thermodynamics, Process & Energy Department, Faculty of Mechanical Engineering, Delft University of Technology, Delft, The Netherlands; [b]Department of Materials Science and Engineering, Faculty of Mechanical Engineering, Delft University of Technology, Delft, The Netherlands; [c]Shell Global Solutions International B.V., The Hague, The Netherlands; [d]Shell Global Solutions International B.V., Amsterdam, The Netherlands; [e]Van't Hoff Institute of Molecular Sciences, University of Amsterdam, Amsterdam, The Netherlands; [f]Department of Applied Physics and Science Education, Eindhoven University of Technology, Eindhoven, The Netherlands

ABSTRACT

Accurate prediction of thermodynamic properties of hydrocarbons is essential for chemical process modelling. Conventional group contribution methods often are used to predict these properties. However, these methods often require extensive parameter sets to handle structural complexities. A refined group contribution method for predicting thermodynamic properties of hydrocarbon isomers with reduced complexity and improved accuracy is presented and discussed. By combining the structural framework of Constantinou and Gani (CG94) with a sensitivity-based selection of second-order groups, a reduced yet highly effective set of twelve second-order groups is identified. This reduced set retains the predictive power comparable to more complex models while significantly reducing the number of parameters. Linear regression is applied to model enthalpies and Gibbs free energies of formation for a wide temperature range. To test broader applicability, the model is further extended to properties that require nonlinear regression, including critical temperatures, critical pressures, acentric factors, and liquid densities. For all cases, the proposed model achieves high predictive accuracy, demonstrating its robustness and generalizability. This methodology balances interpretability, efficiency, and performance, making it suitable for both research and industrial thermodynamic modelling.



ARTICLE HISTORY
Received 7 August 2025
Accepted 22 September 2025

KEYWORDS
Group contribution methods; critical parameters; Gibbs free energy; standard enthalpy of formation; long-chain hydrocarbons

1. Introduction

The accurate prediction of thermodynamic properties of hydrocarbons is a fundamental requirement for the design, simulation, and optimisation of chemical processes, as well as innovative products with improved environmental and safety properties, particularly in the context of the global transition toward sustainable fuels and chemicals [1]. Iso-alkanes with high degrees of branching are preferred constituents in sustainable aviation fuels (SAF), lubricants, and phase change materials due to their desirable thermophysical properties such as high energy density, low freezing point, and cold flow properties [2]. Consequently, catalytic processes such as hydroisomerization, which convert linear alkanes into branched isomers inside shape-selective zeolites, are of growing industrial relevance [3]. Experimental determination of thermodynamic properties like the standard Gibbs free energy (ΔG_f^0), standard enthalpy of formation (ΔH_f^0) and entropy (ΔS^0) for the myriad of possible branched alkanes, particularly those with more than ten carbon atoms, is often infeasible due to the large number of isomers and practical limitations of laboratory

CONTACT Thijs J.H. Vlugt t.j.h.vlugt@tudelft.nl Engineering Thermodynamics, Process & Energy Department, Faculty of Mechanical Engineering, Delft University of Technology, Leeghwaterstraat 39, Delft 2628CB, The Netherlands

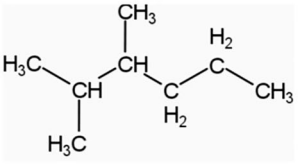
Supplemental data for this article can be accessed online at http://dx.doi.org/10.1080/00268976.2025.2566763.

measurements [4,5]. To address this, group contribution methods (GCMs) have emerged as a widely-used and efficient approach to estimate thermodynamic properties based on molecular structure [6]. These methods predict properties by summing contributions from predefined structural fragments, termed 'groups' which are generally classified as first-order groups that are basic functional units or higher order groups that capture local structural environments and neighbouring atom effects [7].

Classical GCMs such as those of Lydersen [8], and Joback and Reid [9] have provided reasonably accurate predictions for small and moderately branched molecules. The Constantinou and Gani (CG94) method [10] improved thermochemical property predictions by introducing a two-level structure: first-order groups capture basic functional fragments, while another set of groups, i.e. second-order groups, account for local structural effects like branching and conjugation. This methodology managed to improve accuracy and applicability of group contributions and partially capture the isomer effect. In this method, through chemical intuition, the typical first-order groups for alkanes are used, and second-order groups are defined by specifying a central atom or group and its first neighbouring atoms or groups, thereby encoding the local chemical environment more explicitly. This allows for more accurate differentiation between isomers and improves predictions for molecules with complex or branched structures. Unfortunately, the accuracy of predictions still decreases for highly branched long-chain alkanes [11]. This limitation often arises primarily from the reliance on first-order groups and limited inclusion of second-order corrections [12]. Recent research [13–15] has increasingly focussed on refining group definitions, expanding group libraries to integrate more structural effects, and applying new computational advances to improve the prediction of thermodynamic properties of complex isomers.

To overcome the shortcomings of existing GCMs, Sharma et al. [13] proposed a novel linear regression-based second-order group contribution method for alkanes that explicitly captures the interactions between neighbouring atoms. By training the model on a dataset of C_1–C_{10} alkane isomers and systematically incorporating all possible second-order groups, an accuracy beyond 1 kcal/mol was achieved in predicting ΔH_f^0 and ΔG_f^0 for alkanes longer than C_{10}. While highly accurate, the Sharma et al. method has certain limitations that lies in the complexity introduced by the use of 69 distinct second-order groups to represent local atomic environments. Although this comprehensive enumeration improves prediction for long and branched alkanes, it significantly increases the dimensionality of the model, which can make the regression process more complex, and thus reduce interpretability.

This paper presents a novel idea that combines the basic principles of Constantinou and Gani (CG94) [10] and the Sharma et al. methods [13] to maintain the accuracy while reducing the complexity. By identifying and selecting the most relevant second-order groups defined in the Sharma method, and adopting the second-order approximation strategy of CG94 in a data-driven framework, we aim to balance model accuracy and complexity. The proposed method holds potential to predict properties of more structurally complex hydrocarbons that contain additional functional groups beyond those found in alkanes. This paper is organised as follows: first, the theoretical background of linear regression, the CG94 framework, and the Sharma et al. method are presented, followed by details of the methodology for selecting key second-order groups through sensitivity analysis. We analyze the predictive accuracy for both ΔH_f^0 and ΔG_f^0 for a wide temperature range from 0–1500 K, and assess how temperature affects outcomes and model robustness. This study concludes with a summary of key findings and a discussion on the implications for a scalable and interpretable GCM. We specifically focus on ΔH_f^0 and ΔG_f^0 due to the fundamental role in determining chemical equilibrium and thermodynamic feasibility. Other important properties such as the critical constants (T_c, P_c), the molar volume at standard condition (V_m), and the acentric factor (ω) are also included in this study. In the Supporting Information, we provide detailed list of all training data. In SI1.xlsx, the sheet titled DHf0 and DGf0 include training data of ΔH_f^0 and ΔG_f^0 from the Scott tables [16]. The critical temperature, critical pressure, and acentric factors sheets present experimental values of T_c, P_c, and ω from Ref. [17] used for training our model. The liquid density (298K) sheet provides density training data at 298K from Ref. [18], and the molar volume (298K) sheet contains the corresponding molar volume values derived from the liquid density data. SI1.xlsx includes the predictions of these properties using the CG94 first-order group contribution method, the CG94 second-order groups contributions method, the Sharma et al. method, and our method. In SI2.xlsx, the first- and second-order group contributions using different methods for these properties are provided. SI3.py provides the script for identifying the first- and second-order groups in CG94 from SMILES strings and Figure 1 provides such an example for using SI3.py.



SMILES string: 2,3-m-C6

SI3.py

1st order groups	Occurrence (N_i)	2nd order groups	Occurrence (M_j)
CH_3	4	$(CH_3)_2CH$	1
CH_2	2	$(CH_3)_3C$	0
CH	2	$CH(CH_3)$-$CH(CH_3)$	1
C	0	$C(CH_3)_2$-$C(CH_3)_2$	0
		$CH(CH_3)$-$C(CH_3)_2$	0

Figure 1. An example of using SI3.py to identify the first- and second-order groups. One can use the SMILES string of a molecule as input to get the number of first- and second-order groups.

2. Theory

2.1. Linear regression

Linear Regression (LR) is commonly used to predict the thermochemical properties, such as ΔH_f^0 and ΔG_f^0, of alkanes, using the occurrences of first-order or second-order groups as independent variables

$$y = K + \sum_{i=1} C_i N_i \tag{1}$$

where y is the target property, N_i is the occurrence of a first or a second-order group i in the molecule, and C_i is the group contribution of the group i. K serves as fitting residual. To know which variants, or 'groups', are relatively more important, a sensitivity analysis is used [19]. In a LR model of the form

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon, \tag{2}$$

the coefficients β_j indicate the marginal change in the response y per unit change in the predictor x_j keeping all other variables constant. When predictors are measured on different scales or units, as is common in GCMs for thermochemical properties, direct comparison of β_j values can be misleading. To assess sensitivity, all variables are transform into standardised form

$$z_j = \frac{x_j - \bar{x}_j}{s_j}, \quad \text{and} \quad z_y = \frac{y - \bar{y}}{s_y}, \tag{3}$$

where $\bar{x}_j$ and s_j are the mean and standard deviation of predictor x_j, respectively, and similarly for the response y. The standardised regression model becomes

$$z_y = \beta_1^* z_1 + \beta_2^* z_2 + \cdots + \beta_p^* z_p + \varepsilon, \tag{4}$$

where β_j^* is the standardised coefficient of predictor x_j computed from β_j via

$$\beta_j^* = \beta_j \cdot \frac{s_j}{s_y}. \tag{5}$$

The standardised coefficient β_j^* quantifies the number of standard deviations the response will change given a one standard deviation increase in x_j, keeping other variables constant. Therefore, the absolute value $|\beta_j^*|$ gives a direct and interpretable measure of the sensitivity of the output to that predictor [20, 21].

2.2. Constantinou and Gani method (CG94)

The Constantinou and Gani [10] (CG94) method, introduced in 1994, features both first-order groups and second-order groups. first-order groups represent basic functional units like -CH_3 or -CH_2-, while second-order groups serve as correction factors that capture structural dependencies, such as branching, conjugation, and neighbouring group interactions [7]. The definition of the second-order groups was based on the conjugation principle as presented in the open literature. When applied to alkanes in a united-atom representation, only four first-order groups (CH_3, CH_2, CH and C) and five second-order groups (shown in Figure 5(a)) are considered. An innovative element of the CG94 is its two-step property estimation by using the model below:

$$f(X) = \sum_i N_i C_i + W \sum_j M_j D_j + K \tag{6}$$

where $f(X)$ represents the function (linear or non-linear) of estimated value of the target property X, N_i and M_j are the occurrence of first-order groups and second-order groups, and C_i and D_j represent the group contributions. Figure 1 provides an example on how to assign first- and second-order groups for a branched hydrocarbon. Initially, the model fits the contributions of first-order groups by ignoring second-order effects ($W = 0$). Once these base values of C_i and K are established, second-order group contributions are introduced and optimised in a separate regression step ($W = 1$), while keeping C_i and K constant. This ensures that the second-order effects D_j are treated as corrections to the first order

approximation. Note that C_i, D_j, and K are temperature-dependent parameters, allowing the model to capture the temperature variation of the target property. This approach maintains the independence of first-order groups and allows second-order groups to capture subtle topological and interaction-based corrections without excessive adjustable parameters [7]. Despite its advancements over the earlier GCMs, CG94 still requires some improvements in specific areas. For example, the conjugation principle in CG94 does not always account for long-range interactions and overall molecular effects like conformational flexibility or electronic delocalisation, which are important for modelling large or highly interactive molecular systems [22]. Therefore, CG94 can be further supported by molecular-level theories in order to improve the accuracy of the estimation of properties of highly complex organic structures and accurately capture isomer-specific behaviour [23].The CG94 provided the foundation for several other efforts in group contributions aiming to improve GCMs by refining group definitions, expanding group libraries, and incorporating more structural effects. For example, Marrero and Gani [24] added a third order correction to the Constantinou and Gani second order approximation model. However, this introduces a significant number of additional adjustable parameters and implementation complexity. Similarly, Constantinou et al. [25] and later researchers [15, 26–28] explored approaches that integrate ring corrections, stereochemistry, and group interactions beyond nearest neighbours of pure compounds and mixtures. These developments may be perceived as as an intermediate stage, bridging classical dual-level models with modern machine-learning frameworks. A comprehensive critical review of GCMs can be found in Ref. [7].

Table 1. First-order group contributions C_i of our method for ΔH_f^0 at different temperatures.

Temperature/[K]	CH_3	CH_2	CH	C
0	−4.473	−15.034	−24.509	−35.827
200	0.619	−0.645	−0.222	1.522
273	3.751	6.149	11.070	18.372
298	4.874	8.523	13.999	22.572
300	4.938	8.735	15.371	24.771
400	9.645	18.607	31.583	48.773
500	14.559	28.771	48.167	73.197
600	19.637	39.129	64.933	97.765
700	24.804	49.654	81.936	122.516
800	29.822	60.104	98.906	147.302
900	35.080	70.677	115.828	171.833
1000	40.217	81.277	132.910	196.507
1100	44.877	91.980	150.716	222.368
1200	50.419	102.514	167.050	245.820
1300	55.296	113.150	184.229	270.247
1400	60.422	123.745	201.228	294.585
1500	63.390	134.719	219.318	321.960

Table 2. Second-order group contributions D_j of our method for ΔH_f^0 at different temperatures.

Temperature/[K]	$CH_3(C)$	$CH_3(CH)$	$CH_3(CH_2)$	$CH_2(CH)(CH)$	$CH_2(CH)(CH_2)$	$CH_2(CH_2)(CH_2)$	$CH_2(C)(CH_3)$	$CH_2(C)(CH_2)$	$CH_2(CH)(CH_3)$	$CH_2(C)(CH)$	$CH_2(C)(C)$	$CH_2(CH_2)(CH_3)$
0	−0.132	0.677	5.086	−8.676	−4.500	−1.400	−1.819	−8.497	−3.206	−6.441	−1.475	−3.060
200	−0.129	0.734	5.392	−2.840	−0.000	−1.254	−1.777	−3.582	−7.751	−1.343	−3.380	−1.819
273	−0.174	0.761	5.518	−8.077	−6.155	−1.369	−1.640	−9.019	−3.731	−8.142	−1.270	−3.521
298	−0.217	0.775	5.550	−8.012	−5.214	−1.387	−1.648	−0.929	−3.780	−2.823	−1.392	−3.567
300	−0.183	0.775	5.550	−8.012	−5.214	−1.387	−1.648	−0.929	−3.780	−2.823	−1.392	−3.567
400	−0.183	0.775	5.550	−8.012	−5.214	−1.387	−1.648	−0.929	−3.780	−2.823	−1.392	−3.733
500	−0.086	0.861	5.759	−3.963	−5.230	−1.453	−2.172	−0.993	−4.927	−1.753	−3.992	−3.992
600	−0.336	0.916	5.885	−8.999	−5.550	−1.480	−1.077	−0.981	−4.256	−10.117	−0.413	−3.920
700	−0.390	1.026	6.053	−8.608	−6.000	−1.460	−0.806	−0.916	−4.428	−10.613	−0.283	−3.976
800	−0.430	1.026	6.053	−8.608	−6.000	−1.460	−0.806	−0.916	−4.428	−10.613	−0.283	−3.976
900	−0.451	1.169	6.123	−9.084	−6.098	−1.918	−0.949	−0.978	−4.918	−11.064	−0.364	−4.054
1000	−0.507	1.287	6.251	−10.004	−6.538	−1.992	−1.046	−1.012	−4.538	−11.591	−0.456	−4.139
1100	−0.475	1.369	6.339	−10.240	−6.500	−1.431	0.092	−1.059	−4.535	−12.140	−0.542	−4.221
1200	−0.483	1.169	6.317	−10.240	−6.500	−1.431	0.092	−1.059	−4.535	−12.140	−0.542	−4.221
1300	−0.523	1.222	6.336	−10.641	−6.174	−1.309	0.369	−1.022	−3.807	−0.824	−4.322	−4.322
1400	−0.336	1.112	6.411	−10.806	−6.718	−1.411	0.370	−1.160	−4.066	−15.403	−1.579	−3.505
1500	−2.247	1.690	−0.669	−4.930	−8.023	−3.029	10.245	−3.875	3.416	−4.992	1.580	−2.781

2.3. Sharma et al. method

The Sharma et al. method [13] represents a recent advancement in CGMs specifically designed to improve the prediction of thermodynamic properties for long-chain and highly branched alkanes developed considering hydroisomerization as an application. Unlike earlier models that rely primarily on first-order groups, the method uses a comprehensive and systematic enumeration of second-order groups as the sole molecular descriptors. This method exhaustively enumerates all the possible atom combinations surrounding a central atom and forms second-order groups present within a molecule. In this way, 69 second-order groups are defined for branched alkanes. This definition of second-order groups captures the influence of neighbouring group interactions, branching patterns, and local connectivity, which are factors especially crucial in iso-alkanes where small differences in branching can lead to significant changes in thermochemical properties [7]. Unlike CG94 where both first and second-order groups are used in a two-step regression, the Sharma et al. method exclusively considers second-order groups in LR using the data set provided by Scott [16]. Each of the 69 defined second-order groups is treated as an independent variable and its contribution is directly estimated through the regression coefficients. The extensive use of all 69 distinct second-order groups introduces a notable level of complexity. While this richness and exhaust improves the predictive accuracy, it also makes the model harder to interpret, more data-intensive, and less generalisable. Although the Sharma et al. method marks a significant leap in structural sensitivity, its high dimensionality raises challenges for practical implementation and may limit scalability. Therefore, a sensitivity analysis is used to

Table 3. First-order group contributions C_i of our method for ΔG_f^0 at different temperature.

Temperature/[K]	CH_3	CH_2	CH	C
0	−4.473	−15.034	−24.509	−35.827
200	0.619	−0.645	−0.222	1.522
273	3.751	6.149	11.070	18.372
298	4.874	8.523	13.999	22.572
300	4.938	8.735	15.371	24.771
400	9.645	18.607	31.583	48.773
500	14.559	28.771	48.167	73.197
600	19.637	39.129	64.933	97.765
700	24.804	49.654	81.936	122.516
800	29.822	60.104	98.906	147.302
900	35.080	70.677	115.828	171.833
1000	40.217	81.277	132.910	196.507
1100	44.877	91.980	150.716	222.368
1200	50.419	102.514	167.050	245.820
1300	55.296	113.150	184.229	270.247
1400	60.422	123.745	201.228	294.585
1500	63.390	134.719	219.318	321.960

Table 4. Second-order group contributions D_j of our method for ΔG_f^0 at different temperature.

Temperature/[K]	$CH_3(C)$	$CH_3(CH)$	$CH_3(CH_2)$	$CH_2(CH)(CH)$	$CH_2(CH)(CH_2)$	$CH_2(CH_2)(CH_2)$	$CH_2(C)(CH_3)$	$CH_2(C)(CH_2)$	$CH_2(CH)(CH_3)$	$CH_2(C)(CH)$	$CH_2(C)(C)$	$CH_2(CH_2)(CH_3)$
0	−0.132	0.677	5.086	−8.676	−4.500	−1.400	−1.819	−8.497	−3.206	−6.441	−1.475	−3.060
200	−0.129	0.734	5.392	−2.840	−0.000	−1.254	−1.777	−3.582	−7.751	−1.343	−3.380	−1.819
273	−0.174	0.761	5.518	−8.077	−6.155	−1.369	−1.640	−9.019	−3.731	−8.142	−1.270	−3.521
298	−0.217	0.775	5.550	−8.012	−5.214	−1.387	−1.648	−0.929	−3.780	−2.823	−1.392	−3.567
300	−0.183	0.775	5.550	−8.012	−5.214	−1.387	−1.648	−0.929	−3.780	−2.823	−1.392	−3.567
400	−0.183	0.775	5.550	−8.012	−5.214	−1.387	−1.648	−0.929	−3.780	−2.823	−1.392	−3.733
500	−0.086	0.861	5.759	−3.963	−5.230	−1.453	−2.172	−0.993	−4.927	−1.753	−3.992	−3.992
600	−0.336	0.916	5.885	−8.999	−5.550	−1.480	−1.077	−0.981	−4.256	−10.117	−0.413	−3.920
700	−0.390	1.026	6.053	−8.608	−6.000	−1.460	−0.806	−0.916	−4.428	−10.613	−0.283	−3.976
800	−0.430	1.026	6.053	−8.608	−6.000	−1.460	−0.806	−0.916	−4.428	−10.613	−0.283	−3.976
900	−0.451	1.169	6.123	−9.084	−6.098	−1.918	−0.949	−0.978	−4.918	−11.064	−0.364	−4.054
1000	−0.507	1.287	6.251	−10.004	−6.538	−1.992	−1.046	−1.012	−4.538	−11.591	−0.456	−4.139
1100	−0.475	1.369	6.339	−10.240	−6.500	−1.431	0.092	−1.059	−4.535	−12.140	−0.542	−4.221
1200	−0.483	1.169	6.317	−10.240	−6.500	−1.431	0.092	−1.059	−4.535	−12.140	−0.542	−4.221
1300	−0.523	1.222	6.336	−10.641	−6.174	−1.309	0.369	−1.022	−3.807	−0.824	−4.322	−4.322
1400	−0.336	1.112	6.411	−10.806	−6.718	−1.411	0.370	−1.160	−4.066	−15.403	−1.579	−3.505
1500	−2.247	1.690	−0.669	−4.930	−8.023	−3.029	10.245	−3.875	3.416	−4.992	1.580	−2.781

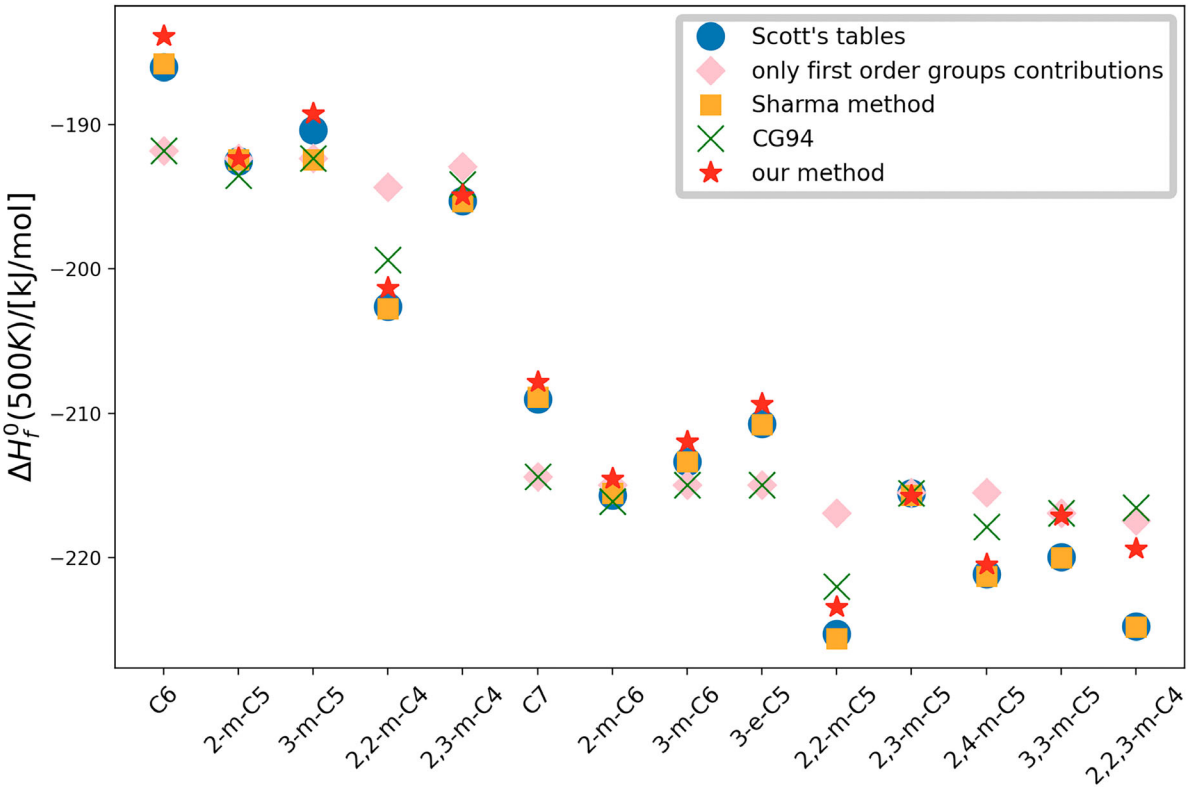


Figure 2. Comparison of predicted values of ΔH_f^0 at 500 K for various iso-alkanes using only first-order group contributions (pink rhomb), Sharma et al. method [13] (yellow squares), CG94 [10] (green crosses), our method (red stars), and the training set from the Scott tables [16] (blue circles). Using only first-order groups provides reasonably accurate predictions for less branched alkanes, the Sharma et al. method shows an excellent agreement with the Scott tables, CG94 a achieves a better accuracy for branched isomers by incorporating local structural corrections via second-order group correction.

determine which groups have more impact on predicting the thermodynamics properties.

3. Results and discussion

Figure 6 shows the sensitivity analysis for ΔH_f^0 at 500 K in the Sharma et al. method, where each second-order group is characterised by its $|\beta_j^*|$ and its occurrence for all molecules provided by the Scott tables [16]. A higher value of $|\beta_j|$ indicates a large sensitivity, meaning the corresponding group has a stronger influence on the predicted thermodynamic property. The circles within the blue ellipses show both high sensitivities, i.e. strong influence on predicted enthalpy and high frequency of occurrence, which indicates that these groups are statistically significant, making them the most important contributors in the model. In sharp contrast, many of the groups concentrated near the origin have either negligible $|\beta_j^*|$ values, low occurrence, or both. These groups contribute little to the overall variance in ΔH_f^0 and may be considered less relevant in terms of predictive power. It is also worth mentioning that some groups have very low occurrence, which may be attributed to the limitation of the training dataset, which includes only C_1–C_{10} isomers and thus lacks highly branched structures only found in heavier alkanes. The combination of high $|\beta_j^*|$ and high occurrence therefore serves as a useful criterion for identifying the most influential structural motifs in the regression model. This trend was consistently observed for all temperatures from 0 K to 1500 K, for both ΔH_f^0 and ΔG_f^0, indicating the robustness of group importance for thermal variations.

Based on the observation, the 12 second groups falling in the blue circles, which are characterised by both high sensitivity and high frequency occurrence, are proposed to be selected as a new representative set and are shown in Figure 5(b). This subset captures the majority of group features while significantly decreasing model complexity by reducing the number of second-order groups needed to fit. This reduced group set (as denoted by: our new model) is then used to develop a new linear regression model, which is systematically compared to the Sharma

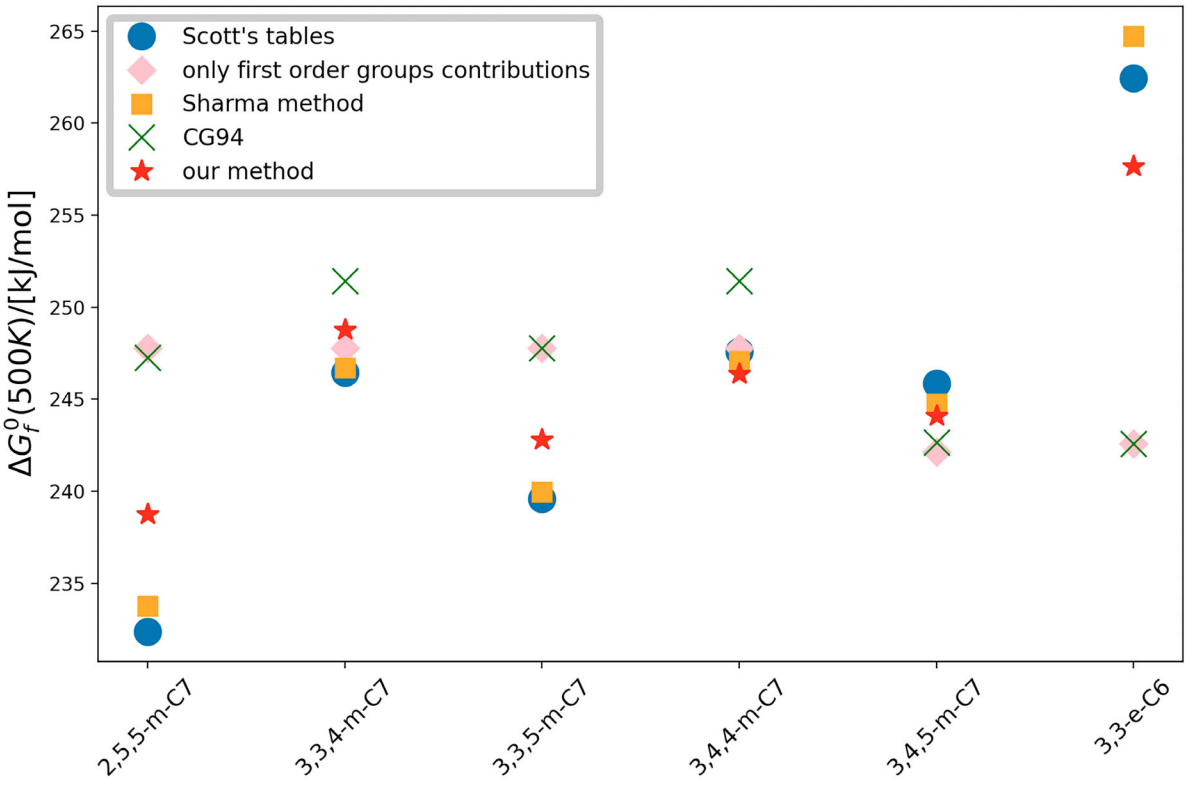


Figure 3. Comparison of predicted values of ΔG_f^0 at 500 K for selected branched iso-alkanes for decane using different GCMs: using only first-order group contributions (pink rhomb), CG94 [10] (green crosses), the Sharma et al. method [13] (yellow squares), and our method (red stars), and compared to reference data from Scott thermochemical tables [16] (blue circles).

et al. method, which includes all 69 second-order groups, and the CG94 method, which incorporates both first and second-order groups. All five second-order groups defined in CG94 (Figure 5(a)) can be fully represented using combinations of the more detailed second-order groups selected in our new method (Figure 5(b)). For example, the CG94 group corresponding to $CH(CH_3)_2$ can be assembled from two $CH_2(CH_3)$ and one $C(CH_3)$ groups. Similarly, the CG94 group $CH(CH_3)CH(CH_3)$ corresponds to two $CH_2(CH_3)$ units connected via a central carbon. This demonstrates that the CG94 groups are a subset or simplified combinations of the second-order groups selected through our sensitivity-based approach. Therefore, our new set preserves the representational capacity of CG94 while offering a finer structural resolution.

While our sensitivity analysis is conducted specifically for ΔG_f^0 and ΔH_f^0, this focus is rooted in the original design of the Sharma et al. method, which was developed and calibrated mainly for these two thermodynamic properties. Since the 69 second-order groups in the Sharma et al. method were trained and validated using ΔG_f^0 and ΔH_f^0 data, the selection of a reduced group set should start from the same context. Interestingly, the selected subset of second-order groups emerging from our analysis shows a high degree of chemical intuitiveness. Many of these groups represent prototypical local environments that reflect key branching and substitution patterns, such as $CH_2(C)(CH_3)$ or $CH_2(CH)(CH)$, which are expected to influence a wide range of thermodynamic and physical properties. This structural logic suggests that the most influential groups for ΔG_f^0 and ΔH_f^0 may also play important roles in other properties like critical parameters and the acentric factor (ω). Therefore, while our method is derived from sensitivity analysis on a limited property domain, its generalizability is empirically plausible and chemically justifiable. In the later sections of this work, we test whether this same set retains promising predictive performance for multiple temperature-dependent properties, providing a first assessment of its broader applicability. The fist- and second-order group contributions C_i adn D_i of these 12 groups used in our method for ΔH_f^0 and ΔG_f^0 can be found in Tables 1–4.

The predicted ΔH_f^0 at 500 K for various C_6–C_7 isomers using first-order group contributions only and CG94 with the reference values from the Scott tables are shown in Figure 2 and compared. For all isomers, the first-order group contribution method yields the poorest performance. As branching increases, the

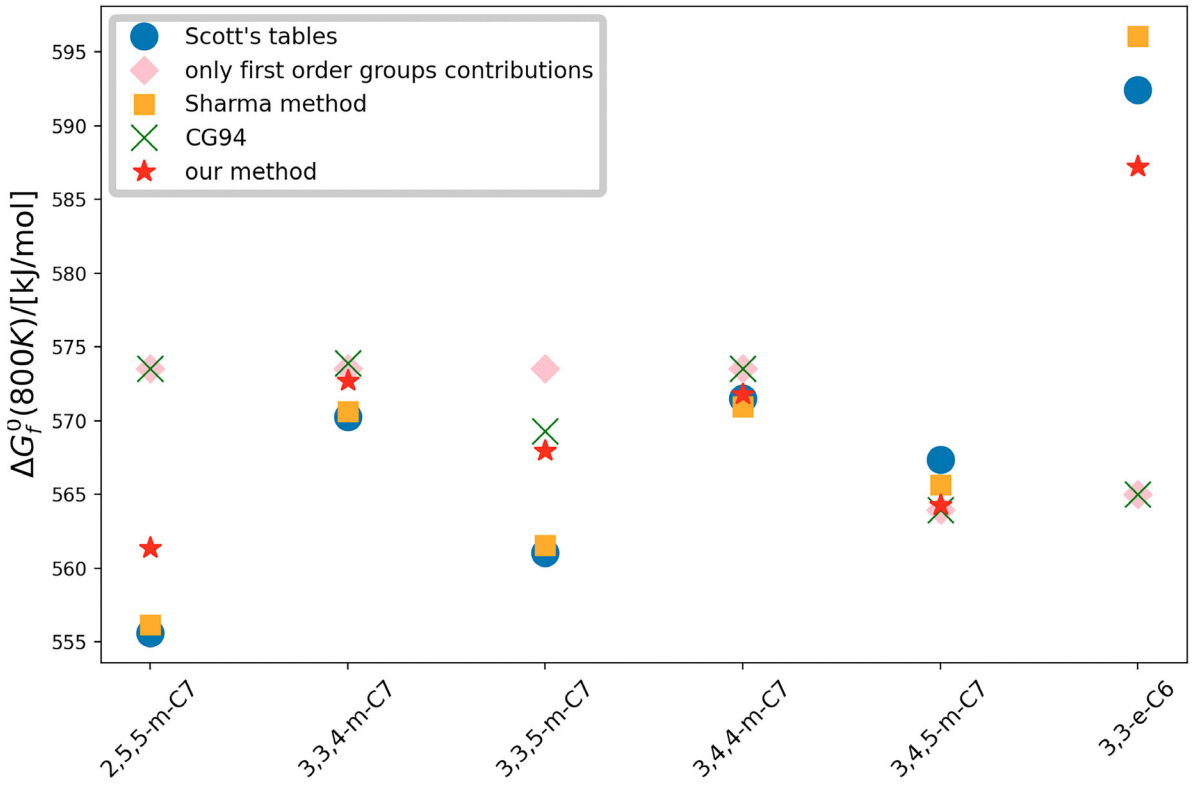


Figure 4. Comparison of predicted values of ΔG_f^0 at 800 K for selected branched iso-alkanes for decane using different GCMs: using only first-order group contributions (pink rhomb), CG94 [10] (green crosses), the Sharma et al. method [13] (yellow squares), and our method (red stars), and compared to reference data from Scott thermochemical tables [16] (blue circles).

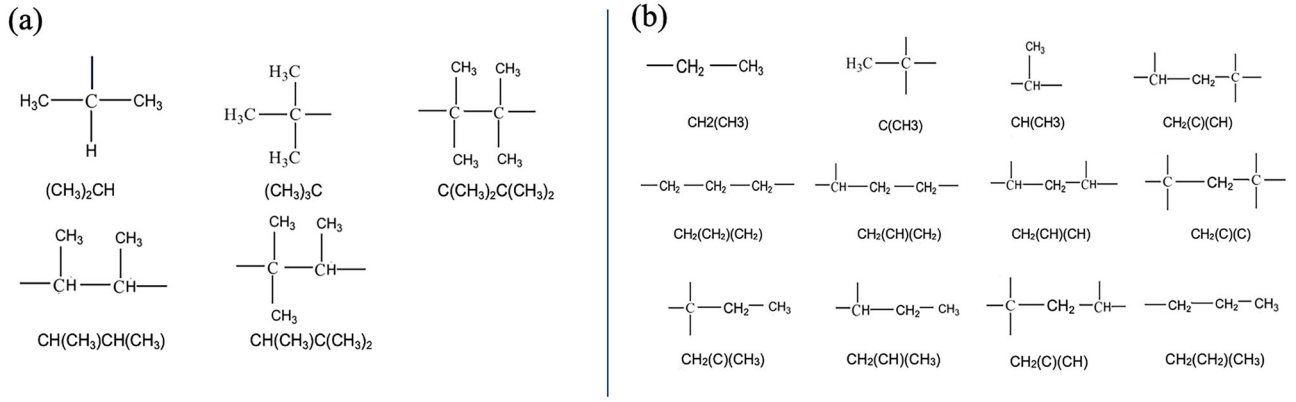


Figure 5. (a) five second-order groups used in CG94 [10] and (b) twelve second-order groups selected through sensitivity analysis. The original CG94 work defined only 5 second-order groups for alkanes, while 12 second-order groups featuring high sensitivity and high occurrence are chosen for our method from the 69 second-order groups defined in the Sharma et al. method [13].

accuracy of the first-order model reduces significantly, while CG94, by incorporating second-order structural correction, improves predictions for some isomers. CG94 still fails to fully capture fine-grained structural effects. In particular, when the number of occurrences of a second-order group is small, the contribution of this group may be undervalued. This still shows the importance of using second-order groups as a correction. Capturing the local structural environment and surrounding atom effects is essential for accurately predicting thermochemical properties of complex branched isomers [29].

At all investigated temperatures, our method consistently outperforms CG94, with first-order group contributions being the least accurate for all isomers. The Sharma method generally yields the best agreement with the Scott tables, but our method closely follows, striking

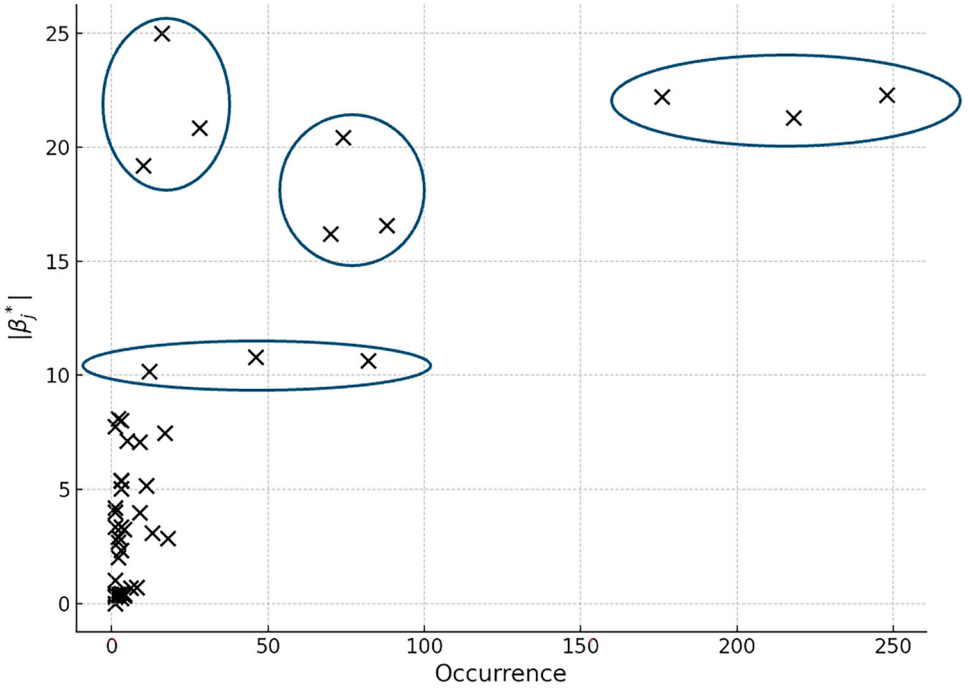


Figure 6. Sensitivity analysis of second-order groups used in the Sharma et al. method [13] for predicting ΔH_f^0 at 500 K. Each point represents a second-order group, with the vertical axis indicating its sensitivity ($|\beta_j^*|$), and the horizontal axis showing its number of occurrences in the dataset. Groups within the blue ellipses are both highly sensitive and frequently occurring, and were thus selected as the 12 most influential groups for our method to construct a reduced group set for further comparison with CG94 [10] and the Sharma model [13]. Notably, only 46 out of the 69 second-order groups in the Sharma et al. method were detected from all the molecules listed in Scott tables [16].

a balance between accuracy and transferability. At 500 K (Figure 3), our predictions for ΔG_f^0 for 3,3,4-m-C_7, 3,4,4-m-C_7, and 3,4,5-m-C_7 are in excellent agreement with both the Sharma method and Scott tables, and this agreement is also maintained at 800 K (Figure 4). Importantly, our model systematically provides results that fall between those of the Sharma method and CG94, offering improved accuracy while relying on a reduced set of only 12 high-sensitivity second-order groups. This compact yet carefully selected set is sufficient to capture structural effects effectively, demonstrating that exhaustive parameterisation is not required to achieve reliable predictions. The new method, by incorporating a few additional fitted parameters beyond CG94, enhances accuracy substantially without introducing the complexity of Sharma's 69 second-order parameters. While some minor trade-offs in accuracy remain, particularly for highly branched isomers such as 3,3-e-C6, the predictive performance of our approach demonstrates clear robustness across a broad range of branched structures. This balance of accuracy, simplicity, and scalability makes our method particularly attractive for extension to hydrocarbons beyond alkanes, where the inclusion of additional second-order groups will be essential but where the exhaustive Sharma approach would become increasingly cumbersome.

To further assess the temperature dependence of model performance, Figure 7 presents the R^2 values for ΔG_f^0 and ΔH_f^0 predictions, respectively, for a range of temperatures from 400 K to 1500 K. For both ΔG_f^0 and ΔH_f^0, the Sharma et al. method consistently maintains the highest R^2 values, exceeding 0.99 at nearly all temperatures, which re-affirms its robustness and accuracy. Our new method exhibits a performance curve that closely follows that of Sharma et al. method, achieving R^2 values above 0.995 over a wide temperature range. Interestingly, its accuracy improves steadily with temperature up to around 1000 K, before a slight decline appears. In contrast, CG94 starts with significantly lower R^2 values, below 0.96 at 400 K and then shows a gradual rise, reaching a plateau around 0.996 at mid-range temperatures, before dropping sharply at 1500 K. The Sharma et al. method again delivers near-perfect R^2, while the our method remains stable around 0.990 with minor fluctuations. CG94 shows improved accuracy with increasing temperature but consistently underperforms relative to the other models. Notably, the gap between our method and Sharma is slightly more pronounced for ΔH_f^0 than for ΔG_f^0, possibly reflecting that enthalpy is more sensitive to specific group contributions. It is also worth noting that all three models exhibit a decline

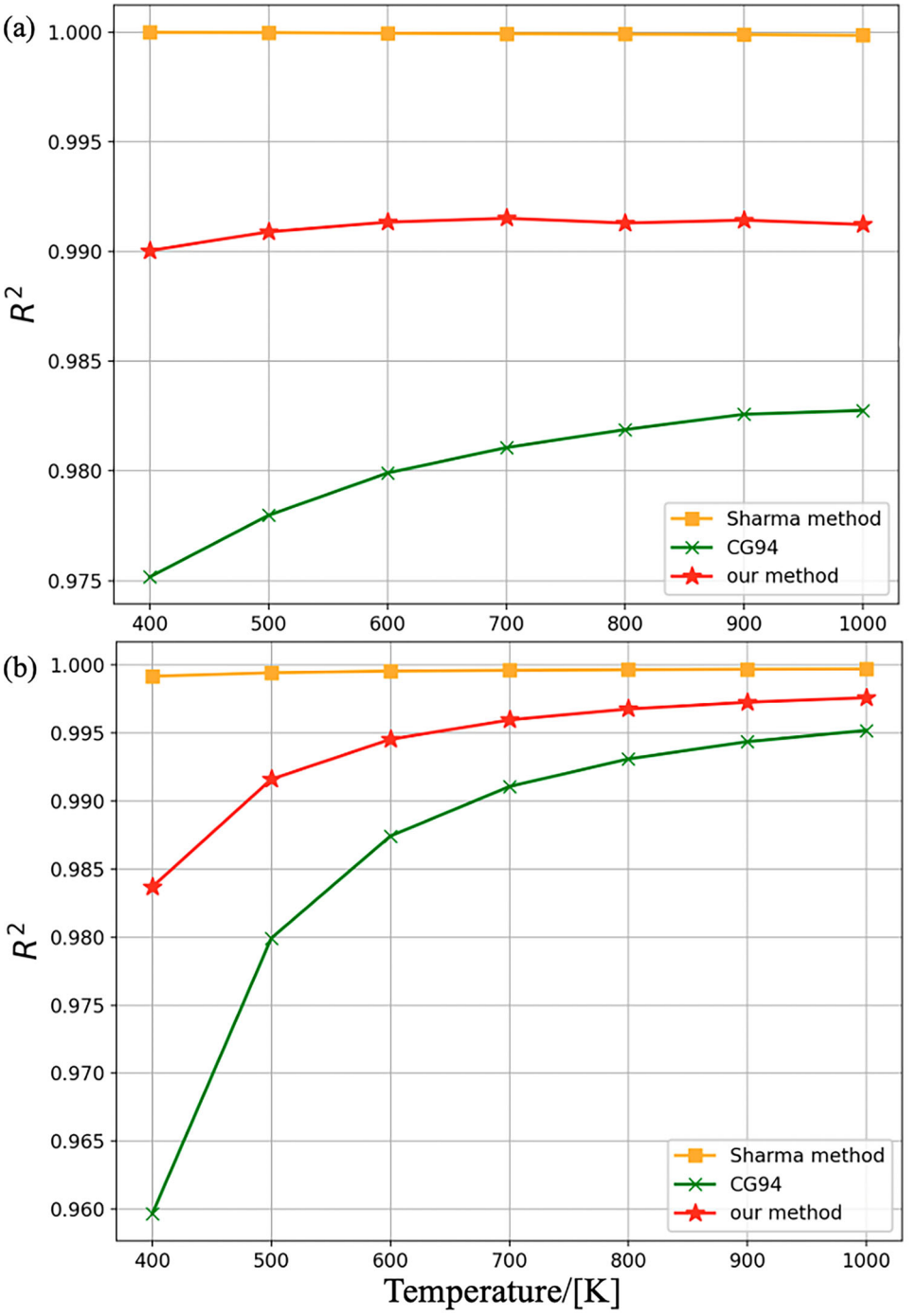


Figure 7. Temperature-dependent coefficient of determination (R^2) for thermochemical property predictions of CG94 (yellow), our method (blue), and the Sharma et al. method (green). (a) R^2 values for ΔH_f^0 predictions for a temperature range of 400–1500 K. (b) R^2 values for ΔG_f^0 predictions. The Sharma et al. method maintains consistently high accuracy for all temperature, while our method exhibits strong performance with minor deviations at high temperatures. In contrast, CG94 shows lower R^2 values, particularly at lower temperatures, reflecting its limited structural resolution.

in R^2 values for ΔG_f^0 at 1500 K. While still maintaining relatively high accuracy, this simultaneous drop for all models suggests that prediction becomes inherently more challenging at extreme temperatures. One possible explanation may be the increasing dominance of entropic contributions at high temperatures [30]. R^2 values for all temperatures can be found in Tables 5 and 6. The absolute and relative root mean square deviations are shown in Tables 7 and 8 for ΔH_f^0 and ΔG_f^0, respectively.

Table 5. Comparison of R^2 values for predicted ΔH_f^0 at various temperatures using three different GCMs: CG94 [10], our method, and the Sharma et al. method [13].

Temperature/[K]	R^2 (CG94)	R^2 (our method)	R^2 (Sharma et al. method)
0	0.943972	0.976717	0.999988
200	0.965040	0.985654	0.999987
273	0.969607	0.987911	0.999993
298	0.970933	0.988487	0.999995
300	0.971017	0.988528	0.999995
400	0.975165	0.990045	0.999997
500	0.977978	0.990901	0.999986
600	0.979905	0.991340	0.999956
700	0.981066	0.991520	0.999935
800	0.981887	0.991299	0.999915
900	0.982585	0.991432	0.999896
1000	0.982759	0.991231	0.999859
1100	0.982753	0.991009	0.999842
1200	0.982715	0.990652	0.999816
1300	0.982599	0.990334	0.999805
1400	0.982220	0.989895	0.999590
1500	0.981967	0.989441	0.999742

Table 6. Comparison of R^2 values for predicted ΔG_f^0 at various temperatures using three different GCMs: CG94 [10], our method, and the Sharma et al. method [13].

Temperature/[K]	R^2 (CG94)	R^2 (our method)	R^2 (Sharma et al. method)
0	0.943976	0.976705	0.999880
200	0.244684	0.671566	0.995275
273	0.791147	0.919019	0.997689
298	0.865462	0.947700	0.998270
300	0.869337	0.949164	0.998313
400	0.959663	0.983684	0.999167
500	0.979919	0.991594	0.999422
600	0.987419	0.994528	0.999543
700	0.991061	0.995952	0.999598
800	0.993089	0.996761	0.999641
900	0.994359	0.997251	0.999675
1000	0.995183	0.997582	0.999695
1100	0.995812	0.997809	0.999690
1200	0.996180	0.997989	0.999719
1300	0.996533	0.998127	0.999730
1400	0.996438	0.997936	0.999590
1500	0.969542	0.970780	0.975832

The R^2 analysis for a wide temperature range confirms that Sharma et al. method shows the most accurate and consistent performance, which is expected as it used all 69 type of second-order group as fitted parameters. In sharp contrast, our new method, despite using only four first-order and twelve 'important' second-order groups, manages to achieve R^2 values above 0.995 at most temperatures examined. This indicates that the reduced model is not only significantly simpler but also highly efficient in capturing the essential structural effects. Compared to CG94, which shows notably lower R^2 values, particularly at lower and higher temperatures, our new method shows a clear advantage in balancing model complexity with predictive reliability.

Table 7. Absolute and relative root mean square deviations for predicted ΔH_f^0 at various temperatures using our GCM.

Temperature/[K]	Absolute RMSD	Relative RMSD
0	4.08	0.0242
200	4.084	0.0202
273	3.992	0.0185
298	3.968	0.0183
300	3.967	0.0182
400	3.922	0.017
500	3.907	0.0162
600	3.945	0.0164
700	3.948	0.0152
800	4.013	0.0152
900	4.011	0.0151
1000	4.043	0.0152
1100	4.073	0.0153
1200	4.106	0.0157
1300	4.126	0.016
1400	4.164	0.0165
1500	4.192	0.017

Table 8. Absolute and relative root mean square deviations for predicted ΔG_f^0 at various temperatures using our GCM.

Temperature/[K]	Absolute RMSD	Relative RMSD
0	4.08	0.0243
200	4.229	0.0958
273	4.403	2.41
298	4.456	0.858
300	4.464	0.4793
400	4.774	0.0522
500	5.027	0.0269
600	5.507	0.0194
700	6.449	0.0162
800	6.379	0.0133
900	6.862	0.0119
1000	7.352	0.0109
1100	7.874	0.0103
1200	8.393	0.0096
1300	9.457	0.0095
1400	10.182	0.0095
1500	41.381	0.048

To better compare the model performance in practical applications, Figure 8 presents ΔH_f^0 at 298 K for linear alkanes of C_4 to C_{20} using these three different GCMs. The predictions for C_4 to C_{10} represent the fitted results, while the experimental data for C_{11} to C_{20} are extrapolated values, intended to evaluate generalizability of each model beyond the training range. When extrapolated to longer alkanes, significant differences in performance emerge. The Sharma et al. method exhibits the best consistency with the experimental data, followed by our method, while the CG94 shows the largest deviations, particularly for higher carbon numbers. This comparison shows the improved extrapolation capability of our method over CG94.

Having demonstrated the strong performance of the reduced second-order group set in predicting ΔG_f^0 and ΔH_f^0 for a wide temperature range, we next explored whether this our method also retains predictive power for other key thermodynamic properties. Specifically, we applied the same group framework to estimate critical

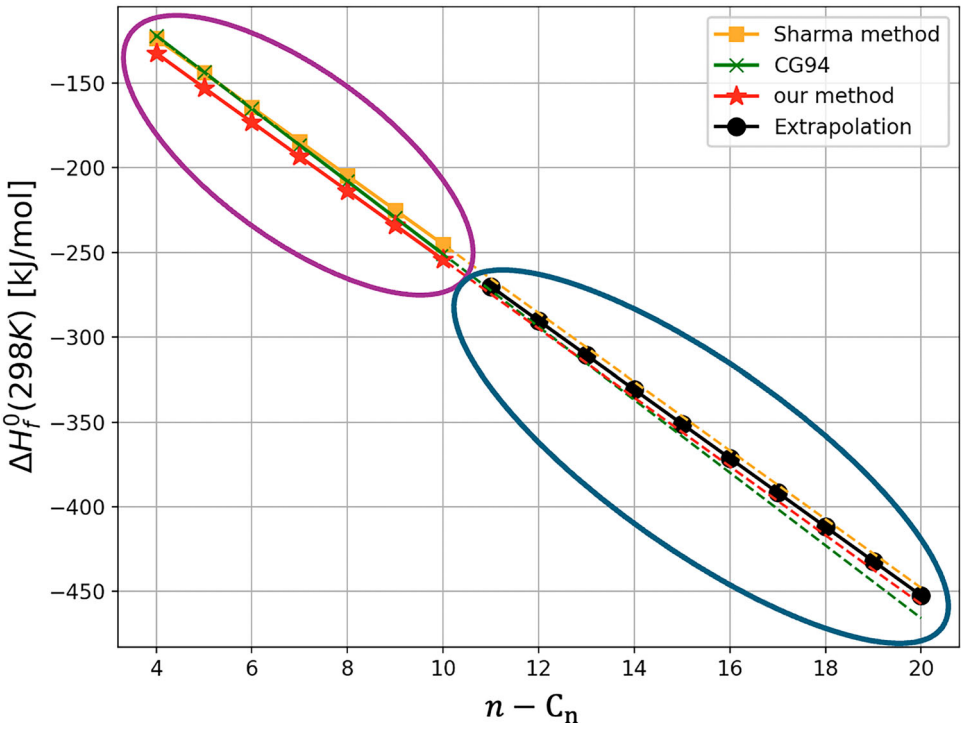


Figure 8. Prediction of ΔH_f^0(298 K) for linear alkanes using three GCMs: Sharma et al. method [13] (yellow squares), CG94 [10] (green crosses), and our method (red stars). The black circle line represents extrapolated experimental values of C_{11}–C_{20}. The region on the left (purple ellipse) shows the fitted range C_4–C_{10}, while the right region (blue ellipse) indicates the extrapolation zone. The dashed lines correspond to linear trendlines for each method. Among the three models, the Sharma et al. method shows the best extrapolation performance, followed by our method, while CG94 exhibits the largest deviation from the experimental data.

Table 9. First-order group contribution C_i for T_c, P_c, ω and V_m using our method.

Group	T_c	P_c	ω	V_m
CH_3	1.571	0.456	0.538	19.725
CH_2	1.681	0.405	0.005	15.942
CH	1.676	−0.623	−0.531	10.714
C	1.967	−1.282	−1.069	4.506

Table 10. Second-order group contribution D_j for T_c, P_c, ω and V_m using our method.

Group	T_c	P_c	ω	V_m
CH_3(C)	−0.030	0.195	1.124×10^{-4}	0.214
CH_3(CH)	0.114	0.187	-1.488×10^{-4}	−0.303
$CH_3(CH_2)$	0.223	0.001	-5.382×10^{-4}	−1.320
CH_2(CH)(CH)	−0.909	−0.388	1.296×10^{-3}	3.082
CH_2(CH)(CH_2)	−0.486	−0.390	1.207×10^{-3}	1.976
$CH_2(CH_2)(CH_2)$	−0.028	−0.391	5.621×10^{-5}	0.554
CH_2(C)(CH_3)	−0.688	−0.386	1.053×10^{-3}	1.295
CH_2(C)(CH_2)	0.246	−0.204	-4.237×10^{-4}	−0.426
CH_2(CH)(CH_3)	−0.096	−0.203	1.036×10^{-4}	0.414
CH_2(C)(CH)	−1.109	−0.384	8.907×10^{-4}	3.622
CH_2(C)(C)	−1.051	−0.378	2.045×10^{-3}	2.241
$CH_2(CH_2)(CH_3)$	0.073	−0.210	-2.176×10^{-4}	−1.037

temperatures (T_c), critical pressures (P_c), acentric factors (ω), and liquid densities at standard conditions (ρ_l), to assess the broader applicability and structural relevance of these selected groups. Unlike before, these properties require functional forms that can accommodate diminishing returns or saturation as molecular size increases, which shows non-linear beheviors [31]. This justifies fitting curves such as power-laws or logarithmic relation rather than relying on a simple additive linear model. Such an approach aligns with prior works [23, 32–34] in the field, where GCMs using nonlinear regression have successfully improved accuracy for critical properties of hydrocarbons. The fitting equations for T_c, P_c and ω are as follows:

Table 11. Parameters fitted trough non-linear regression using Equations (7)–(9).

T_0/[K]	P_0/[bar]	a/[-]	α/[-]	β/[-]
218.880	7.301	−5.364	2.938	0.575

$$e^{\frac{T_c}{T_0}} = \sum_i N_i C_i + W \sum_j M_j D_j + K \tag{7}$$

$$P_c = P_0 + \left(\sum_i N_i C_i + W \sum_j M_j D_j + K \right)^a \tag{8}$$

$$\omega = \alpha \left[\ln \left(\sum_i N_i C_i + W \sum_j M_j D_j + K \right) \right]^\beta \tag{9}$$

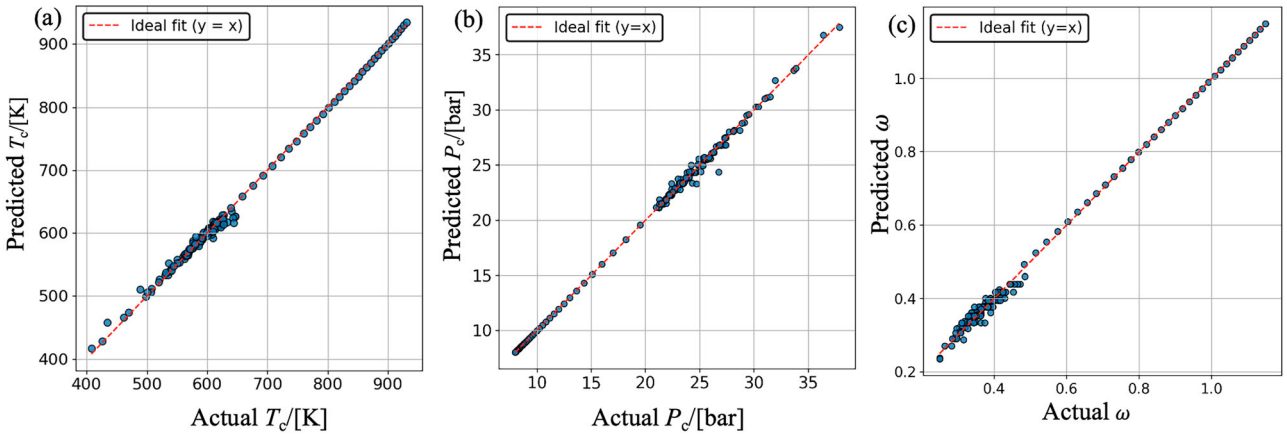


Figure 9. Parity plots comparing predicted and experimental values for (a) critical temperatures (T_c), (b) critical pressures (P_c), and (c) acentric factors (ω) using the proposed nonlinear regression. The red dashed line represents the ideal correlation ($y = x$). All three properties show strong agreement between predicted and actual values, highlighting the accuracy and robustness of the model for a diverse range of hydrocarbon structures.

These equations are fitted through non-linear regression using the `curve_fit` function from the `scipy.optimise` module in Python. All training data used in this non-linear regression were obtained from Yaws' Handbook [17]. Similarly, the regression was conducted in a two-step procedure consistent with the philosophy illustrated in Equation (6). First, only first-order group parameters C_i were fitted with $W = 0$. Once the contribution values for first-order groups C_i were established, second-order group effects D_i were introduced and optimised in a separate regression step by setting $W = 1$. It is important mentioning that all the parameters, including T_0, P_0, a, α, β and K, in Equations (7)–(9) were fitted together with C_i (when $K = 0$) and these fitted parameters can be found in Table 11. The group counts N_i and M_j are determined from SMILES string [35] using Python. The first- and second-order groups contributions C_i and D_j for T_c, P_c and ω can be found in Tables 9 and 10. All the fitted parameters can be found in the file SI2.xlsx in the Supporting Information.

Figure 9 shows the parity plots of the predictive performance of our proposed model for T_c, P_c and ω. For all three properties, the predicted values exhibit a strong linear correlation with experimental data, as evidenced by the close alignment of the data points along the ideal $y = x$ reference line. The prediction of ω, T_c, and P_c also demonstrates excellent accuracy, although minor deviations appear in more complex or highly branched compounds. This high degree of agreement reflects the ability and robustness of our model to incorporate non-linear structural effects through tailored functional forms. Quantitative performance metrics including R^2, the mean absolute error (MAE), the average relative deviation (ARD), the absolute root mean square deviation, and the relative root mean square deviation for each property are summarised in Table 12, which shows a decent accuracy for these three properties.

Table 12. The mean absolute error (MAE), the average relative deviation (ARD), R^2, and absolute and relative root mean square deviation (RMSD) for T_c, P_c, ω, V_m, and ρ_l using nonlinear regression (our method).

Property	MAE	ARD	R^2	Absolute RMSD	Relative RMSD
T_c	3.73 K	0.63%	0.9967	5.978	0.0105
P_c	0.17 bar	0.70%	0.9974	0.312	0.0124
ω	0.0091	2.43%	0.9968	0.022	0.0844
V_m	0.95 ml/mol	0.50%	0.9968	0.964	0.0052
ρ_l	0.80 g/ml	0.51%	0.9667	0.004	0.0052

To analyze the liquid densties (ρ_l) at 298K and 1 bar pressure of alkanes, we followed an indirect regression procedure. First, we compiled a dataset of experimental ρ_l values from literature [18], which covers of a wide range of linear and branched alkanes. Next, the values of ρ_l were converted into molar volumes (V_m) using

$$V_m = \frac{M}{\rho_l} \tag{10}$$

where M is the molar mass. Both V_m and ρ_l depend on molecular size and structure, with ρ_l being inversely proportional to V_m. In our approach, linear regression is applied to V_m, and the predicted molar volumes are subsequently converted back to ρ_l for direct comparison with the training dataset. The first- and second-order group contributions (C_i and D_j) for V_m are reported in Tables 9 and 10. Figure 10 presents parity plots of the predictive performance of the proposed model for both V_m and ρ_l. For V_m, the predictions show near-perfect alignment with the experimental values, closely following the ideal correlation line ($y = x$). In sharp contrast, the

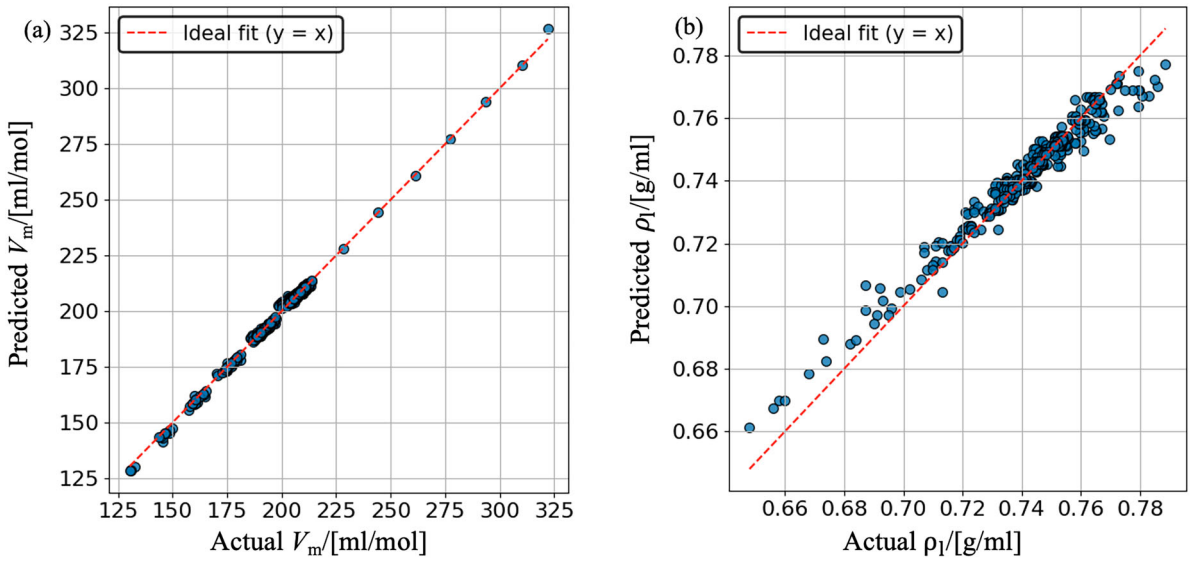


Figure 10. Parity plots comparing predicted and experimental values of (a) V_m and (b) the values of ρ_l. V_m were directly fitted using linear regression, while ρ_l were obtained by converting the predicted V_m values. The excellent agreement in (a) shows the suitability of V_m for linear regression CGMs while the slightly larger deviations in (b) reflect the additional complexity inherent in density predictions.

converted predictions of ρ_l display somewhat larger deviations. This can be attributed to the narrower range of experimental ρ_l values (0.66–0.78 g/mL) compared to the broader range of V_m (125–300 mL/mol). As ρ_l is inversely related to V_m, even small errors in the predicted V_m are amplified on conversion, particularly at higher densities. While the overall correlation remains strong, the scatter around the ideal line is visibly larger for ρ_l than for V_m. This deviation reflects the nonlinear transformation between volume and density, which inherently magnifies minor discrepancies in the underlying V_m predictions. Combined with the quantitative performance shown in Table 12, these results confirm the advantage of modelling V_m as the primary regression target. This approach not only yields more accurate and stable predictions, but also better reflects the physical relationship between molecular structural and thermodynamic properties. Together with the high predictive performance for T_c, P_c and ω shown earlier, these results validate the applicability of our method to both linear and nonlinear thermodynamic properties.

4. Conclusions

In this work, we proposed a simplified CGM that applies the second-order approximation approach of the Constantinou and Gani method [10] with a sensitivity-guided selection of second-order groups inspired by the Sharma et al. method [13]. By identifying twelve most impactful second-order groups based on sensitivity, we were able to develop a new model that strikes a balance between predictive accuracy and model simplicity. Our method shows a promising predictive performance for both ΔH_f^0 and ΔG_f^0 of alkane isomers for a wide temperature range from 0–1500 K. It retains accuracy comparable to the Sharma et al. method, which uses 69 second-order groups as fitting parameters, while using only 16 parameters. This shows that only a reduced subset of second-order groups can be essential for capturing the key structural variations relevant to thermochemical properties. Beyond linear regression of ΔH_f^0 and ΔG_f^0, we tested the broader applicability of this reduced group set by fitting experimental T_c, P_c, ω, and ρ_l at 298K using nonlinear regression. The results showed excellent agreement with experimental data, e.g. $R^2 > 0.996$ for T_c, P_c, and ω, confirming the effectiveness of our approach in modelling thermodynamic properties that require nonlinear fitting procedures. These results collectively demonstrate that our methodology is not only efficient but also broadly applicable to both linear and nonlinear regression tasks in group contribution modelling. The high accuracy for diverse property types validates the robustness of our approach, and shows its potential use in industrial applications where interpretability, scalability, and efficiency are essential. This new methodology has been implemented only to alkanes. Encouraged by the excellent results, future work will expand its implementation to a wide range of pure organic compounds and properties (thermodynamic, transport, environmental-related, safety related, etc.) that would allow the availability of a powerful tool for process optimisation and design of molecules with properties of environmental importance.

Supporting information

The Supporting Information consists of the file SI1.xlsx, SI2.xlsx and SI3.py. All training data are listed in SI1.xlsx together with the first- and second-order model predictions. In SI2.xlsx, the contributions, C_i and D_i, of each group for each property and other fitted parameters, including K, T_0, P_0, a, α and β in Equations (7)–(9), are listed in the sheets containing 'contributions', and the predictions of each property are listed in the sheets containing 'predictions'. On SI2.xlsx, the sheets starting with 'CG94' show the group contributions and properties predictions for CG94 [10], the sheets starting with 'Sharma' show the group contributions and properties predictions for the Sharma et al. method [13], and the sheets starting with 'New method' show the group contributions and properties predictions for our method. The code to convert to the SMILES string and the code to count the numbers of first- and second-order groups for CG94 are in SI3.py.

Acknowledgments

This work is part of the Advanced Research Center for Chemical Building Blocks, ARC-CBBC, which is cofunded and cofinanced by The Netherlands Organization for Scientific Research (NWO) and The Netherlands Ministry of Economic Affairs and Climate Policy. The authors also acknowledge the use of computational resources of the DelftBlue supercomputer, provided by Delft High Performance Computing Center (https://www.tudelft.nl/dhpc) [36]. We are also grateful for the support by NWO Domain Science for the use of supercomputer facilities, with financial support from the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (Netherlands Organisation for Scientific Research, NWO).

Data availability statement

All raw data is provided in the Supporting Information of this paper.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work received funding from the Advanced Research Center for Chemical Building Blocks, ARC-CBBC, which is cofunded and cofinanced by The Netherlands Organization for Scientific Research (NWO) and The Netherlands Ministry of Economic Affairs and Climate Policy.

ORCID

Thijs J.H. Vlugt http://orcid.org/0000-0003-3059-8712

References

[1] S. van Bavel, S. Verma, E. Negro and M. Bracht, ACS Energy Lett. **5**, 2597–2601 (2020). doi:10.1021/acsenergylett.0c01418
[2] T.M. Letcher editor, *Chemical Thermodynamics for Industry*, 1st ed., (Royal Society of Chemistry, Cambridge, UK, 2004).
[3] L. Tao, J.J. Jacobson, L. Zhang, M.A. Jackson, D.B. Hodge, C. Kinchin and M. Wang, Biofuels Bioprod. Biorefining **11**, 965–980 (2017).
[4] F.M. Fraser and E.J. Prosen, J. Res. Natl. Bur. Stand. **55**, 329–333 (1955). doi:10.6028/jres.055.040
[5] E. Prosen, K. Pitzer and F. Rossini, Natl. Bur. Stand. **34**, 403–412 (1945). doi:10.6028/jres.034.022
[6] R. Gani, Curr. Opin. Chem. Eng. **23**, 184–196 (2019). doi:10.1016/j.coche.2019.04.007
[7] Z. Li, L. Constantinou, R. Baur, D. Dubbeldam, S. Calero, S. Sharma, M. Rigutto, P. Dey and T.J.H. Vlugt, Mol. Phys. (2025). In Press.
[8] A.L. Lydersen, Engineering Experiment Station Report 3, College of Engineering, University of Wisconsin, Madison, Wisconsin, 1955.
[9] K.G. Joback and R.C. Reid, Chem. Eng. Commun. **57**, 233–243 (1987). doi:10.1080/00986448708960487
[10] L. Constantinou and R. Gani, AIChE J. **40**, 1697–1710 (1994). doi:10.1002/aic.v40:10
[11] J. Abildskov, L. Constantinou and R. Gani, Fluid. Phase Equilib. **118**, 1–12 (1996). doi:10.1016/0378-3812(95)02846-3
[12] K.K. Yalamanchi, V.C. Van Oudenhoven, F. Tutino, M. Monge-Palacios, A. Alshehri, X. Gao and S.M. Sarathy, J. Phys. Chem. A **123**, 8305–8313 (2019). doi:10.1021/acs.jpca.9b04771
[13] S. Sharma, J.J. Sleijfer, J. Op de Beek, S. van der Zeeuw, D. Zorzos, S. Lasala, M.S. Rigutto, E. Zuidema, U. Agarwal, R. Baur, S. Calero, D. Dubbeldam and T.J.H. Vlugt, J. Phys. Chem. B **128**, 9619–9629 (2024). doi:10.1021/acs.jpcb.4c05355
[14] S. Hwang and J. Kang, Int. J. Thermophys. **43**, 9–42 (2022). doi:10.1007/s10765-022-03060-7
[15] J. Gmehling, J. Chem. Thermodyn. **41**, 731–747 (2009). doi:10.1016/j.jct.2008.12.007
[16] D.W. Scott, J. Chem. Phys. **60**, 3144–3165 (1974). doi:10.1063/1.1681500
[17] C.L. Yaws, *Yaws' Handbook of Thermodynamic and Physical Properties of Chemical Compounds*, 1st ed., (Knovel, Houston, USA, 2003).
[18] C.L. Yaws, *Thermophysical Properties of Chemicals and Hydrocarbons*, 1st ed., (William Andrew Inc., Norwich, USA, 2008).
[19] C.H. Achen, *Interpreting and Using Regression*, 1st ed., (Sage Publications, Beverly Hills, CA, 1982).
[20] J.W. Johnson and J.M. LeBreton, Organ. Res. Methods **7**, 238–257 (2004). doi:10.1177/1094428104266510
[21] R. Azen and D.V. Budescu, Psychol. Methods **8**, 129–148 (2003). doi:10.1037/1082-989X.8.2.129
[22] E. Stefanis and C. Panayiotou, Int. J. Thermophys. **29**, 568–585 (2008). doi:10.1007/s10765-008-0415-z
[23] L. Constantinou, R. Gani and J.P. O'Connell, Fluid Phase Equilib. **103**, 11–22 (1995). doi:10.1016/0378-3812(94)02593-P

[24] J. Marrero and R. Gani, Ind. Eng. Chem. Res. **40**, 5256–5267 (2001).
[25] L. Constantinou and V. Vassiliades, in *Chemical Product Design: Towards a Perspective through Case Studies* (Elsevier, Amsterdam, 2007), p. 34.
[26] A. Tihic, N. von Solms, M.L. Michelsen, G.M. Kontogeorgis and L. Constantinou, Fluid Phase Equilib.**281**, 60–69 (2009). doi:10.1016/j.fluid.2009.04.003
[27] A. Groniewsky and B. Hégely, Fluid Phase Equilib. **577**, 113990 (2024).
[28] Y. Nannoolal, J. Rarey and D. Ramjugernath, Fluid Phase Equilib. **252**, 1–27 (2007). doi:10.1016/j.fluid.2006.11.014
[29] R. Rajesh, R. Morales-Rodríguez and R. Gani, Ind. Eng. Chem. Res. **61**, 17469–17493 (2022).
[30] N.M. Kuznetsov and S.M. Frolov, Energies **14**, 2641 (2021). doi:10.3390/en14092641
[31] Y. Nannoolal, J. Rarey, D. Ramjugernath and W. Cordes, Fluid Phase Equilib. **226**, 45–63 (2004). doi:10.1016/j.fluid.2004.09.001
[32] X. Wen and Y. Qiang, Ind. Eng. Chem. Res. **40**, 20–32 (2001).
[33] W. Wakeham, G. Cholakov and R. Stateva, J. Chem. Eng. Data. **47**, 559–570 (2002). doi:10.1021/je010308l
[34] V. Villazón-León, A. Bonilla-Petriciolet, J. Tapia-Picazo, J. Segovia-Hernández and M. Corazza, Chem. Eng. Res. Des. **185**, 458–480 (2022). doi:10.1016/j.cherd.2022.07.033
[35] D. Weininger, J. Chem. Inf. Comput. Sci. **28**, 31–36 (1988). doi:10.1021/ci00057a005
[36] Delft High Performance Computing Centre (DHPC), *DelftBlue Supercomputer (Phase 2)* (2024). (accessed Jul 26, 2024).