Review of group contribution methods for prediction of thermodynamic properties of long-chain hydrocarbons

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

DOI
10.1080/00268976.2025.2563020

Licence
CC BY-NC-ND

Publication date
2025

Document Version
Final published version

Published in
Molecular Physics

Citation (APA)
Li, Z., Constantinou, L., Baur, R., Dubbeldam, D., Calero, S., Sharma, S., Rigutto, M., Dey, P., & Vlugt, T. J. H. (2025). Review of group contribution methods for prediction of thermodynamic properties of long-chain hydrocarbons. *Molecular Physics*, *123*(21-22), Article e2563020. https://doi.org/10.1080/00268976.2025.2563020

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

Review of group contribution methods for prediction of thermodynamic properties of long-chain hydrocarbons

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) Review of group contribution methods for prediction of thermodynamic properties of long-chain hydrocarbons, Molecular Physics, 123:21-22, e2563020, DOI: 10.1080/00268976.2025.2563020

To link to this article: https://doi.org/10.1080/00268976.2025.2563020

MOLECULAR PHYSICS
2025, VOL. 123, NOS. 21–22, e2563020 (20 pages)
https://doi.org/10.1080/00268976.2025.2563020

Taylor & Francis
Taylor & Francis Group

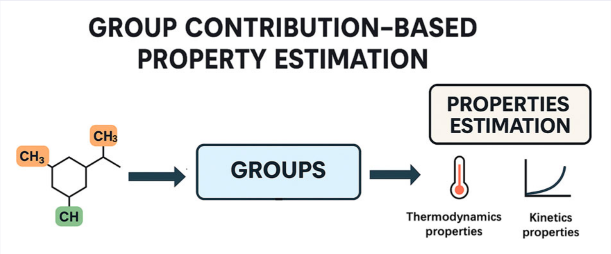
THERMODYNAMICS 2024 CONFERENCE (BY INVITATION ONLY)

Review of group contribution methods for prediction of thermodynamic properties of long-chain hydrocarbons

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

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

ABSTRACT
Group contribution methods (GCMs) provide a practical and computationally efficient approach for predicting thermodynamic properties of hydrocarbons, especially when experimental data are scarce. This review evaluates the evolution of GCMs from classical first-order schemes (e.g. Lydersen method, Joback method) to more advanced second-order frameworks (e.g. CG94, Sharma method), hybrid extensions, and emerging machine learning integrations. While first-order models are simple and widely used, these models struggle with branched and long-chain molecules. Second-order approaches significantly improve structural sensitivity and predictive accuracy, achieving deviations below 2–3% for critical properties and within 1 kcal/mol for formation enthalpies of branched alkanes. Nevertheless, challenges remain in extrapolating to highly complex molecules, underrepresented functional groups, and extreme conditions. Promising directions include reinforcement of second-order GCMs with molecular theory, systematic expansion of experimental and quantum-based datasets, and hybrid GCM–machine learning models that retain interpretability while improving generalisability. We recommend prioritising models that balance accuracy, robustness, simplicity, and transferability to accelerate sustainable process and product designs, particularly in applications such as fuel upgrading including hydroisomerisation, separation processes, and green chemical development.



ARTICLE HISTORY
Received 20 May 2025
Accepted 10 September 2025

KEYWORDS
Group contribution methods; thermodynamic properties; long-chain hydrocarbons; molecular modelling

1. Introduction

With the increasing demand for transition to renewable energy and products, the prediction of thermodynamic properties of organic chemicals is critical for the design, simulation, and optimisation of chemical processes as well as the design of new products with improved environmental properties [1]. Accurate and reliable property predictions are crucial for applications ranging from reaction and separation processes to energy storage and sustainable chemical production [2, 3]. For example, sustainable aviation fuel (SAF), as an alternative for reducing greenhouse gas emissions from the aviation sector, requires various production pathways, such as hydroisomerisation and hydrocracking of long-chain alkanes, for tailoring the properties to meet stringent performance and environmental requirements [4]. To enable accurate modelling and process design for SAF production, it is essential to understand the thermodynamic properties of

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

both linear and branched hydrocarbons involved in these catalytic transformations [5]. The experimental determination of thermodynamic properties of branched hydrocarbons can be costly, time-intensive, and impractical for many compounds, particularly for large branched molecules [6].

Group contribution methods (GCMs) have become a widely-used approach for estimating thermodynamic properties based on the total summation of contributions of molecular substructures [7]. In these methods, the properties of a compound are estimated as a summation of the contributions of simple aspects of the structural groups, which are named as first order groups. In this way, properties can be gained fast by just examining the molecular structure. GCMs provide the important advantage of quick estimates without requiring substantial computational resources [8]. These methods have evolved significantly and incorporated first order and more complex groups, namely higher order groups, to enhance accuracy and account for molecular complexities such as branching and isomerism [6]. GCMs are versatile and capable of predicting a wide range of thermodynamic and transport properties. Commonly estimated properties are the critical temperature (T_c), pressure (P_c), volume (V_c), boiling (T_b) and melting points (T_m), Gibbs energies of formation ($G^0 - H^0(0\,\mathrm{K})$), enthalpies of formation ($H^0 - H^0(0\,\mathrm{K})$), absolute entropies (S^0), Gibbs free energies of formation (ΔG_f), enthalpies of formation (ΔH_f), and constant pressure heat capacities (c_p^0). The notations $G^0 - H^0(0\,\mathrm{K})$ and $H^0 - H^0(0\,\mathrm{K})$ follow the conventions used in the Scott tables [9]. In these tables, enthalpy at 0 K is chosen as the reference state. The temperature over which GCMs can be applied normally covers a range of 298 to 1500 K but GCMs made specifically for temperature outside this range are scarce as GCMs are developed based on experimental data [10].

Despite the success in short-chain organic compounds and mixtures, classical methods, in which a compound is estimated as a summation of the contributions of first order groups that can occur in the molecular structure, often face limitations in predicting properties for highly branched compounds, long-chain molecules, or systems with incomplete group libraries [11]. Recent advancements [12–14] have focussed on improving these methods by refining group contributions, integrating statistical models, and enhancing computational power. A key innovation has been the development of approaches that use higher order group interactions, which refers to interactions beyond isolated functional group like the influence of neighbouring atoms or branches on a central group, to capture molecular intricacies more effectively and accurately [5, 8]. These methods have demonstrated enhanced accuracy for complex systems, often exceeding the precision of earlier models, while reducing the reliance on extensive parameterisation.

In this paper, to support the urgent need for more advanced GCMs, the research status of GCM was reviewed and some of the most representative GCMs will be examined for a better reference for the application at different scenarios. The review is organised into different sections, beginning with the fundamental principles of GCMs, followed by a detailed evaluation of classical and modern methods (Section 3), including Lydersen [15], Joback [16], CG94 [8], and Sharma models [5]. The review also introduces applications in pure components (Section 3.5), mixtures (Section 4), and process design (Section 5), before concluding with recent advancements and future research directions (Section 6). Our findings are summarised in Section 7.

2. Fundamentals of group contribution methods

2.1. Principles and approaches

The fundamental principle of a GCM is that the physical properties of a molecule are determined by its molecular structure and the interactions and chemical bonds between its atoms. In an ideal scenario, if a complete understanding of the interactions acting on each atom is known, all the physical and chemical properties of a molecule can be accurately determined. However, approximations are necessary because of the complexity of the intramolecular and intermolecular interactions. Intramolecular interactions, such as steric hindrance, resonance, and ring strain, mainly influence properties related to the internal energy and flexibility of the molecular structure, such as the absolute entropy (S^0). In sharp contrast, intermolecular interactions significantly affect properties like the critical temperature (T_c), the normal boiling point (T_b), and the enthalpy of vaporisation, as the strength of interactions between molecules determines phase behaviour.

Since the energy of atoms is mainly affected over relatively short distances, it is generally assumed that an atom is primarily influenced by its immediate neighbouring atoms. This implies that if an atom has the same local environment in different molecules, it should behave similarly, regardless of the overall molecular structure. For instance, a carbon atom bonded to three hydrogen atoms and one nitrogen atom exhibits the same contribution to the properties of a molecule whether it is part of N-methylpyrrolidine or trimethylamine. This approximation forms the basis of GCMs. It is worth mentioning that GCMs provide an empirical approximation to the full quantum mechanical (QM) solution by capturing average structural contributions rather than explicitly solving the Schrödinger equation for each molecule. Therefore, GCMs are typically more effective for estimating intramolecular properties like ΔG_f and ΔH_f, but may be less accurate for properties that strongly depend on intermolecular interactions, such as T_b and T_m [17].

In GCMs, specific atomic arrangements, namely ”groups” that can consist of one or more central carbon atoms, are assigned numerical values representing their contributions to a given molecular property. By adding up the contributions of all functional groups within a molecule, the total sum correlates with or directly determines the overall physical properties of the molecule. This approach simplifies the property prediction and is widely used in computational chemistry and molecular design [16]. To determine the numerical values of group contributions, linear or non-linear regression analysis is commonly applied in the development of GCM. A set of experimental property data for known molecules is used to fit a linear (or non-linear) model, where the contribution of each functional group is treated as a regression coefficient. The general form of the equation used is:

$$P = \sum_{i} G_i N_i + \epsilon \tag{1}$$

where P is the target physical property or a function of the target property, G_i represents the contribution of each functional group, N_i is the number of times that group appears in the molecule, and ϵ is a fitting parameter that resembles the residue P. By applying statistical methods, such as least-squares regression, the optimal values of G_i are obtained. One of the main difficulties in this approach is that an extremely large number of groups would be required to describe all possible combinations of atoms and nearest neighbours. Even if these groups could be identified, the lack of sufficient experimental data would prevent statistically significant estimates of their contribution values. To address this, further approximations are made. One key assumption is that the contributions of many groups are not highly dependent on the exact identity of all nearest neighbours as organic compounds often include different types of atoms [16]. For instance, in the case of a carbon atom bonded to three hydrogen atoms and one nitrogen atom, its contribution to molecular properties could be approximated as a carbon bonded to three hydrogen atoms and any non-hydrogen atom, rather than specifically a nitrogen atom. This simplification reduces the number and complexity of group definitions and increases the statistical reliability of contribution values [18].

2.2. *First order group contributions*

In GCMs, first order groups represent basic functional groups in a molecule that contribute to the estimation of physical properties. These groups typically consist of single atoms or small clusters of atoms that define the primary structure of a molecule, without considering more complex interactions. First order groups are directly assigned contribution values, which are summed up to estimate pure component properties like T_b, T_c, and $H^0 - H^0(0\,\mathrm{K})$. These groups are independent of the molecular environment and contribute the same value to the properties of a molecule, regardless of the rest of the molecular structure [19]. Examples of first order groups can be a methyl group ($-CH_3$), a methylene group ($-CH_2-$), a hydroxyl group ($-OH$) and a carboxyl group ($-COOH$). Each of these groups has a predefined numerical value G_i (see Equation (1)) that contributes to the calculation of various thermodynamic and physical properties. The simplicity of first order groups lies in the use of pre-determined values, eliminating the need for complex calculations. Property estimation is achieved by directly summing the contributions of all first order groups, without relying on additional QM calculations or extensive experimental data [20]. While first order groups are highly useful for simple molecules, the limitation in capturing molecular interactions, such as hydrogen bonding, isomers or steric effects, results in the failure of being accurate for polar compounds and complex molecular structures. To improve the accuracy and applicability of GCM, higher order group corrections that capture more complex interactions are often necessary.

2.3. *Second order group and higher order contributions*

The limitations of early GCMs arise because first order groups treat each functional group without accounting for the influence of neighbouring atoms or groups, which can lead to a decreased predictive accuracy in systems where intergroup interactions, steric hindrance, conjugation, or electronic delocalisation significantly affect thermophysical properties. A simple example is the comparison between 2-methylhexane and 2-ethylpentane with the same molecular formula (C_6H_{14}). Although these two compounds have different structures, first order GCMs do not adequately distinguish because only the number of first order groups (here: CH_3, CH_2, CH, and C) are considered and the connectivities of these groups are ignored. Second order groups solve this issue by considering not just the group itself, but also the nature of its neighbouring atoms, which allows the model to better account for branching and other structural effects, thereby improving the prediction of isomer-specific properties.

Second order group contributions are introduced to address these limitations and extend the first order approximation by capturing the interactions between adjacent first order groups. Importantly, second order groups are not merely refinements and are based on underlying molecular theories, which consider how the local chemical environment modifies group behaviour. For example, in the group $CH_2(CH_3)(CH)$, the central CH_2 is influenced by both neighbouring substituents CH_3 and CH, capturing branching effects more accurately than first order representations. Such groups can be identified using QM calculations [20]. One of the earliest methods to consider corrections beyond simple first order additivity was proposed by Benson [10], which introduced systematic rules for estimating thermodynamic properties, including the heat of formation, using group contributions while explicitly incorporating symmetry and steric effects as corrections to the standard group additivity framework. This can be viewed as an initial conceptual foundation for higher order contributions, especially in the context of heat of formation, by recognising the importance of molecular topology and electronic structure in modifying group behaviour.

Multiple definitions of second order groups of hydrocarbons exist in the literature, depending on the structural phenomena being captured–such as steric bulk (e.g. the tert-butyl group $-C(CH_3)_3$), symmetry, or resonance [21]. These groups are sometimes parameterised through ab initio calculations or detailed structural enumeration, offering corrections to first order GCMs and helping reduce systematic bias [22]. In addition, statistical methods like sensitivity analysis can be used to evaluate the impact of each second order group on the overall prediction performance by analyzing how small changes in the presence or frequency of specific groups affect the error of the model [23]. In this way, which groups which have the largest influence on the predicted properties can be identified.

Higher order group based on second order groups contributions go even further by considering long range intramolecular interactions, molecular topology, three-body effects, and even connectivity indices [8]. These are particularly valuable for modelling properties in large, branched hydrocarbons or macromolecules, where simple additive models break down [24]. Higher order terms based on molecular theories help capture the nonlinearity and cooperative effects that first and second order GCMs may miss, making them useful in fields such as drug design, polymer science, and advanced material development [25]. Nevertheless, it is worth noting that property estimation methods beyond second order groups (third order, fourth order, etc.) introduce a large number of model parameters which can satisfactorily regenerate the experimental data used to be introduced, but have questionable extrapolating behaviours. Despite the potential to improve accuracy, second and higher order GCMs come at the cost of increased complexity and are more difficult to parameterise. As the number of interactions and structural variations grows, so too does the number of required parameters and the size of the training dataset needed for reliable calibration. To overcome this, recent advances advocate for hybrid models that combine GCM frameworks with data-driven techniques such as machine learning and Artificial Intelligence (AI) [26]. These approaches allow for flexible parameter tuning, automated feature discovery, and scalable correction models that preserve interpretability while improving performance [12, 27]. At the same time, it should be emphasised that high interpolation accuracy does not guarantee reliable extrapolation performance.

3. Applications of group contribution methods for pure component properties of hydrocarbons

The number of possible molecular structures increases exponentially with carbon number (Table 1), making it practically impossible to synthesise and experimentally measure all compounds [27]. For instance, C_8, C_{12}, C_{16}, and C_{20} have 18, 355, 10,359, and 366,319 structural isomers, respectively [28]. This complexity is particularly relevant in the context of hydroisomerisation processes, where accurate thermodynamic property estimation is essential for designing optimal reaction conditions [5]. GCMs have developed rapidly over the past half decade and many mature models have emerged as a convenient way to provide estimations of thermodynamics properties for pure organic molecules which lack experiment data [7]. More details are provided on three important group contribution methods [8, 15, 16], which, according to the literature [29], have played a central role in the development of other group contribution approaches and have been extensively applied. A GCM developed specifically for alkanes [5] is also examined, which exhibits improved predictive performance for branched isomers and offers valuable insights for advancing the development of structurally sensitive GCM frameworks.

3.1. Lydersen method

The Lydersen [15] method is one of the earliest GCMs used to predict properties for organic compounds where critical parameters T_c, V_c and P_c were investigated. The

Table 1. The number of structural isomers for linear and branched alkanes as a function of chain length.

Molecular Formula	Chain Length	Total No. of Structural Isomers
C_3H_8	3	1
C_4H_{10}	4	2
C_5H_{12}	5	3
C_6H_{14}	6	5
C_7H_{16}	7	9
C_8H_{18}	8	18
C_9H_{20}	9	35
$C_{10}H_{22}$	10	75
$C_{11}H_{24}$	11	159
$C_{12}H_{26}$	12	355
$C_{13}H_{28}$	13	802
$C_{14}H_{30}$	14	1858
$C_{15}H_{32}$	15	4347
$C_{16}H_{34}$	16	10, 359
$C_{17}H_{36}$	17	24, 894
$C_{18}H_{38}$	18	60, 523
$C_{19}H_{40}$	19	148, 284
$C_{20}H_{42}$	20	366, 319
$C_{21}H_{44}$	21	910, 726
$C_{22}H_{46}$	22	2, 278, 658

Notes: The exponential growth in isomer number with increasing chain length shows the combinatorial complexity of molecular structures [30].

equations of groups contributions of these properties were given as:

$$V_\mathrm{c}\ /[\mathrm{cm}^3/\mathrm{mol}] = 40 + \sum G_i N_i \tag{2}$$

$$P_\mathrm{c}\ /[\mathrm{bar}] = \frac{M_w}{\left(0.34 + \sum G_i N_i\right)^2} \tag{3}$$

$$T_\mathrm{c}\ /[\mathrm{K}] = \frac{T_b}{0.567 + \sum G_i N_i - \left(\sum G_i N_i\right)^2} \tag{4}$$

where G_is are the numerical values of group contributions (see Equation (1)) which are shown in Table 2. N_i is the occurrence of the groups and M_w is the molar mass of molecule in [g/mol]. It is worthy mentioning that equation for T_c was gained based on Guldberg Rule [31] that T_b expressed in Kelvin is approximately two-thirds of T_c. The Lydersen Method is simple and widely used for estimating critical properties when experimental data is unavailable. Comparisons with experimental data show that the calculated values deviated by less than 1% for T_c and 3.8% for P_c [16]. However, the method lacks high accuracy when it came to highly branched or heteroatomic compounds where molecular interactions and steric effects shall be considered [16]. One major disadvantage of this method is that the experimental value of T_b is required to estimate T_c, which may not always be available. Nevertheless, Lydersen method has been considered as the prototype for and ancestor of many new models like Joback [16], Constantinou and Gani [8] and others [32, 33].

Table 2. Group contribution parameters used in the Lydersen Method [15] for the estimation of T_c, P_c, and V_c as used in Equations (2)–(4).

Group	T_c /[K]	P_c /[bar]	V_c /[cm^3/mol]
$-CH_3$, $-CH_2-$	0.020	0.227	55.0
$>CH$	0.012	0.210	51.0
$-C<$	–	0.210	41.0
$=CH_2$	0.018	0.198	45.0
$=C<$, $=C=$	–	0.198	36.0
$-CH_2-$ (ring)	0.013	0.184	44.5
$>CH-$ (ring)	0.012	0.192	46.0
$>C<$ (ring)	−0.007	0.154	31.0
$-OH$	0.082	0.060	18.0
$-O-$	0.021	0.160	20.0
$-COOH$	0.085	0.400	80.0
$-NH_2$	0.031	0.095	28.0
$-CN$	0.060	0.360	80.0
$-NO_2$	0.055	0.420	78.0
$-S-$	0.015	0.270	55.0
$=S$	0.003	0.240	47.0

3.2. Joback method

The Lydersen method is refined in the Joback Method [16] to improve accuracy and applicability through the incorporation of additional molecular features that had previously not been considered. While the Lydersen method was built on first order groups that assumed the group contribution of each functional as independent contribution to the whole molecule, Joback introduced topological corrections for cyclic systems, conjugated bonds, and hydrogen bonding effects, which is a significant improvement over the original model of Lydersen where simple additivity was assumed without considering more detailed molecule features. In addition to improving the estimation for the critical parameters, Joback method expands to 12 thermodynamics properties, namely T_c, P_c, V_c, T_b, T_m, ideal gas heat capacity C_p, ΔH_f, ΔG_f, heat of vaporisation (ΔH_vap), entropy of vaporisation (ΔS_vap), heat of fusion (ΔH_fus) and some transport properties like the liquid dynamic viscosity (η_L). Joback investigated a broader and more detailed set of first order groups that covered more heteroatoms, such as nitrogen and halogens, and more functionalised carbon groups, like aromatic carbon. Groups specific to cyclic structures, conjugated systems, and double/triple bonds were also considered. The Joback method provides special attention to polar groups like $-\mathrm{OH}$, $-\mathrm{NH}_2$ and $-\mathrm{COOH}$, which strongly influence intermolecular interactions like hydrogen bonding. These groups have more complex contributions to properties affected by intermolecular forces, particularly to η_L, heat of vaporisation ΔH_vap and T_b. Although the Joback method is still based on first order group contributions, the design of the group values implicitly reflects common neighbouring effects without formally including second order groups, based on the assumption that groups are defined narrowly enough that the surrounding environment is partially accounted for and adjusted group definitions to capture the influence of ring strain in small cyclic molecules, resonance effects in conjugated systems such as aromatics, and branching penalties by distinguishing between groups in branched and linear environments. These considerations make the method more structurally aware and sensitive to complex molecular geometries and interactions, without explicitly introducing higher-order interaction terms. This 'pseudo-second-order' effect within a first-order framework opened a new pathway for defining higher-order groups in future models.

Some of the thermodynamics properties equations derived by Joback are shown below:

$$T_{\mathrm{m}}/[\mathrm{K}] = 122.5 + \sum G_i N_i \tag{5}$$

$$T_{\mathrm{b}}/[\mathrm{K}] = 198 + \sum G_i N_i \tag{6}$$

$$T_{\mathrm{c}}/[\mathrm{K}] = T_{\mathrm{b}}\left[0.584 + 0.965 \sum G_i N_i - \left(\sum G_i N_i\right)^2\right]^{-1} \tag{7}$$

Unlike the Lydersen method, which requires experimental T_{b} values for computing $\mathrm{T_c}$, the Joback method additionally provides an equation for estimating T_{b} when such data are unavailable. Nevertheless, it is worth noting that reliable estimation of the critical temperature for Lydersen and Joback methods require the experimental value of the boiling point. Estimated values of the boiling point used for the prediction of the critical temperature may lead to unreliable extrapolating behaviour as demonstrated in the original paper of Constantinou and Gani method [8]. The corresponding expressions for critical volume and pressure are:

$$V_{\mathrm{c}}\ /[\mathrm{cm^3/mol}] = 17.5 + \sum G_i N_i \tag{8}$$

$$P_{\mathrm{c}}\ /[\mathrm{bar}] = \left[0.113 + 0.0032 N_a - \sum G_i N_i\right]^{-2} \tag{9}$$

where N_i is the number of groups of type i in the molecular structure. This equation improved on Lydersen method by incorporating the impact of molecular structure, ensuring that branched and cyclic molecules were handled with better accuracy.

As mentioned before, Joback also investigated some new properties like ΔH_{vap} and ΔH_{fus}. Notably, the method was extended to include certain transport properties, such as liquid viscosity η_{L} at standard temperature and pressure, broadening the applicability of GCMs beyond just thermodynamic properties:

$$\Delta H_{\mathrm{vap}}/[\mathrm{kJ/mol}] = 15.30 + \sum G_i N_i \tag{10}$$

$$\Delta H_{\mathrm{fus}}/[\mathrm{kJ/mol}] = -0.88 + \sum G_i N_i \tag{11}$$

$$\eta_{\mathrm{L}}/[\mathrm{Pa\cdot s}] = M_w \exp\left[\frac{\left(\sum \eta_a N_i - 597.82\right)}{T} + \sum \eta_b N_i - 11.202\right] \tag{12}$$

where M_{w} is the molecular weight. The values G_i for Joback method are shown in Table 3. The Joback method extended group contribution methods by enabling the estimation of a broader set of thermodynamic properties. Compared to the Lydersen method, the Joback method introduced improved group definitions and took the structural effects like branching, and ring strain into account to enhance the accuracy for a wider variety of organic compounds. Nevertheless, while the Joback Method improved performance for moderately complex molecules, newer developments in GCMs have revealed its limitations in extrapolating to large and complex molecules, as well as in handling strongly associating groups [16]. These shortcomings are likely due to the relatively small and limited database used by Joback and Reid to derive the group parameters [34]. It also has unrealible extrapolating behaviour for properties such as T_{c} and P_{c} [8]. Despite these limitations, Joback Method remains a widely used group contribution method due to its wide applicability and simplicity, and one of the most influential improvements of the original framework of Lydersen.

3.3. Constantinou and Gani method (CG94)

The Constantinou-Gani Method [8] (CG94) was proposed in 1994. Based on the foundational work of earlier models such as Lydersen and Joback methods, which relied on first order group contributions and assumed additive behaviour of independent functional groups, the CG94 was developed to overcome the limitations of earlier GCMs. While Lydersen and Joback methods provided decent and reasonable accuracy for small, simple molecules, these methods struggle with molecules containing multiple functional groups, isomeric variations, and extended conjugated systems where localised interactions and structural dependencies could no longer be neglected [33]. Constantinou and Gani introduced a more systematic and theoretically grounded group contribution framework capable of accounting for the effects of molecular topology, neighbouring group interactions,

Table 3. Joback method [16] group contributions values (G_i) for various thermodynamic properties such as T_b, T_m, T_c, P_c, V_c, ΔH_f, ΔG_f, ΔH_vap, ΔH_fus, and liquid viscosity coefficients (η_a, η_b) (see Equation (9)).

Group	T_b/[K]	T_m/[K]	T_c/[K]	P_c/[bar]	V_c/[cm^3/mol]	ΔH_f/[kJ/mol]	ΔG_f/[kJ/mol]	ΔH_vap/[kJ/mol]	ΔH_fus/[kJ/mol]	η_a	η_b
$-CH_3$	23.58	−5.10	0.0141	−0.0012	65	−76.45	−43.96	2.373	0.908	548.29	−1.719
$-CH_2-$	22.88	11.27	0.0189	0.0000	56	−20.64	8.42	2.226	2.590	94.16	−0.199
-CH−	21.74	12.64	0.0164	0.0020	41	29.89	58.36	1.691	0.749	−322.15	1.187
-C-	18.25	46.43	0.0067	0.0043	27	82.23	116.02	0.636	−1.460	−573.56	2.307
$=CH_2$	18.18	−4.32	0.0113	−0.0028	56	−9.63	3.77	1.724	−0.473	495.01	−1.539
= CH−	17.34	13.72	0.0065	−0.0010	41	12.95	34.72	1.000	0.000	*n.a.*	*n.a.*
≡ CH	9.20	−11.18	0.0027	−0.0008	46	79.30	77.71	1.155	2.322	*n.a.*	*n.a.*
= C =	26.15	17.78	0.0026	0.0028	36	142.14	136.70	2.661	4.720	*n.a.*	*n.a.*
= C-	24.14	11.14	0.0117	0.0011	38	83.99	92.36	2.138	3.063	*n.a.*	*n.a.*
$-NH_2$	73.23	66.89	0.0243	0.0109	38	−22.02	14.07	*n.a.*	*n.a.*	*n.a.*	*n.a.*
-NH (non-ring)	50.17	52.66	0.0295	0.0077	35	53.47	89.39	*n.a.*	*n.a.*	*n.a.*	*n.a.*
−OH	63.56	20.09	0.0031	0.0084	63	−17.33	−22.99	*n.a.*	*n.a.*	*n.a.*	*n.a.*
−S− (non-ring)	68.78	34.40	0.0119	0.0049	54	41.87	33.12	*n.a.*	*n.a.*	*n.a.*	*n.a.*
−S− (ring)	52.10	79.93	0.0019	0.0051	38	39.10	27.76	*n.a.*	*n.a.*	*n.a.*	*n.a.*
$-NO_2$	152.54	127.24	0.0437	0.0064	91	−66.57	−16.83	*n.a.*	*n.a.*	*n.a.*	*n.a.*
-F	−0.03	−15.78	0.0111	−0.0057	27	−251.92	−247.19	−0.670	1.398	*n.a.*	*n.a.*
-Cl	38.13	13.55	0.0105	−0.0049	58	−71.55	−64.31	4.532	2.512	625.45	−1.814
-Br	66.86	43.43	0.0133	0.0057	71	−29.48	−38.06	6.582	3.603	738.91	−2.038
-I	93.84	41.69	0.0068	−0.0034	97	21.06	5.74	9.52	2.724	809.55	−2.224

Notes: Each value of G_i represents the additive contribution of the corresponding structural group to the estimated property [16].

and the effect of specific bonding arrangements on thermodynamic properties. The method was designed to estimate a broad range of critical and thermodynamic properties, including T_c, P_c, V_c, ΔH_f and ΔG_f, directly from molecular structure with improved accuracy and wider applicability. By the incorporation of a hierarchical group definition system and the more detailed consideration of molecular complexity, CG94 was seen as an important step forward in the evolution of group contribution methodologies and established a versatile and reliable tool for property prediction for a diverse set of organic compounds, to date cited more than 1000 times. For P_c, T_c, V_c, T_b and T_m, ΔH_f, and ΔG_f, CG94 method outperforms Joback method. For T_c, Joback method and CG94 method provide similar accuracies with average absolute errors of 4.8 K and 4.85 K, respectively. The comparisons between the two methods are shown in Table 4. The most notable refinement made in CG94 is its two-level group contribution framework. In CG94, first order groups were defined representing basic functional fragments of molecules like those used in Lydersen and Joback methods, which captured the primary contributions of individual atoms and small functional units. Constantinou and Gani introduced second order groups to account for structural dependencies, local interactions, and topological effects based on the principle that molecular properties are not solely dictated by isolated groups but are also significantly influenced by the surrounding molecular environment. These second order groups served as correction factors that modify the contributions of first order groups based on specific arrangements of neighbouring groups, the presence of conjugation, and the molecular context in which a functional group exists. The proposed equation for CG94 equals:

$$f(X) = \sum_i N_i C_i + W \sum_j M_j D_j \tag{13}$$

where $f(X)$ represents the function 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. W is assigned as follows:

$$W = \begin{cases} 1, & \text{both first order and second order groups are used(full model)} \\ 0, & \text{only first order groups are used (basic model for a first order approximation)} \end{cases} \tag{14}$$

The determination of the adjustable parameters and second order groups N_i and M_j is achieved via a two-step regression procedure designed to ensure accuracy and

Table 4. Average absolute errors (AAE) for different thermochemical properties predicted using Joback [16] and CG94 [8] methods.

	P_c/ [bar]	T_c/ [K]	V_c/ [m^3/mol]	T_b/ [K]	T_m/ [K]	ΔH_f/ [kJ/mol]	ΔG_f/ [kJ/mol]
Joback and Reid	2.1	4.8	7.5	12.9	22.6	8.4	8.4
CG94	1.13	4.85	6.0	5.35	14.03	3.71	3.24

Notes: These values are obtained from the original article by Constantinou and Gani [8].

independence between the two levels of group contributions:

1. (1) Regression analysis on experimental data is performed exclusively on the first order groups, with the factor W set to zero, effectively deactivating the influence of second order groups. This allows for the precise estimation of C_i based solely on the primary functional groups present in the molecule without interference from structural corrections.
2. (2) The previously determined first order contributions C_i are held constant, and the second order approximation is activated and fitted by setting $W = 1$. With the first order contributions fixed, regression is then used to determine the values of the second order group contributions D_j. This sequential approach ensures that the second order groups serve as optimisation or corrections to the initial first order approximation, capturing structural and topological effects such as neighbouring group interactions and conjugation.

By dividing the regression into two distinct phases, the method maintains the independence of first order contributions while enhancing overall accuracy through the systematic incorporation of second order corrections. The selection of second order groups in the fitting method shall be considered carefully for a better accuracy of the whole model behavior. As mentioned before, the definition of first order group in CG94 was similar to the previous methods, which involved small functional units such as $-\mathrm{CH}_3$ and $-\mathrm{CH}_2-$. To be consistent with the group contributions for mixtures, Constantinou and Gani used the UNIFAC [35] first order groups. Rather than representing a unique atom or bond type, second order groups were characterised as a structural relation or interaction between adjacent groups or atoms. These include systems of conjugated double bonds, cyclic structures, and specific branching patterns. The incorporation of second order groups enables the model to detect and correct for structural motifs that influence physical properties beyond the sum of isolated functional groups. One of the most important structural effects accounted for in second order groups definition is conjugation, which refers to the delocalisation of π-electrons across alternating single and multiple bonds and then creates a system where electron density is spread over several atoms [36]. This delocalisation stabilises molecules and change the physical and thermodynamic properties [37]. In CG94, Constantinou and Gani incorporated conjugation by specific second order groups that recognise conjugated patterns within the molecular structure based on the ABC approach (Atoms, Bonds, and Conjugates) [38], where conjugation was treated as a core structural principle, knowing that many molecules exist as hybrids of multiple conjugate forms. In this way, the classes of conjugate forms having the strongest conjugation effects can be identified by analyzing the contributions of the associated operators. The group identification focuses on the operators which correspond to the important conjugate forms, that is, the operators with significantly higher contribution than others. Using these principles, Constantinou and Gani have defined 30 second order groups. For parameter estimation and model calibration, Constantinou and Gani used an extensive and diverse dataset composed of 370 organic compounds, varying from hydrocarbons, alcohols, ketones, acids, esters, ethers, and amines, as well as more structurally complex compounds such as aromatics, conjugated dienes, and cyclic systems. Eight pure component properties including T_b, T_m and critical parameters are investigated:

$$\exp\left[\frac{T_\mathrm{c}/[\mathrm{K}]}{t_{c0}}\right] = \sum_i N_i C_i + W \sum_j M_j D_j \tag{15}$$

$$\left(P_\mathrm{c}/[\mathrm{bar}] - p_{c1}\right)^{-0.5} - p_{c2} = \sum_i N_i C_i + W \sum_j M_j D_j \tag{16}$$

$$V_c/[\mathrm{cm}^3/\mathrm{mol}] - v_{c0} = \sum_i N_i C_i + W \sum_j M_j D_j \tag{17}$$

with selected parameters shown in Tables 5 and 6. By incorporating second order corrections to account for structural complexities such as conjugation and branching, the model achieved a high level of consistency for a diverse range of properties. For the critical properties, the method produced an average absolute percentage

Table 5. First order group contribution values C_i for the estimation of T_{c1}, P_{c1}, and V_{c1} in CG94 [8].

Group	$T_{c1}/[\mathrm{K}]$	$P_{c1}/[\mathrm{bar}]$	$V_{c1}/[\mathrm{m^3/kmol}]$
CH_3	1.6781	0.019904	0.07504
CH_2	3.4920	0.010558	0.05576
CH	4.0330	0.001315	0.03153
C	4.8823	−0.010404	−0.00034
$CH_2=CH$	5.0146	0.025014	0.11648
$CH=CH$	7.3691	0.017865	0.09541
$CH_2=C$	6.5081	0.022319	0.09183
$CH=C$	8.9582	0.012590	0.07327
$C=C$	11.3764	0.002044	0.07618
$CH_2=C=CH$	9.9318	0.031270	0.14831
CHO	10.1986	0.014091	0.08635
CH_3O	6.4737	0.020440	0.08746
ACH	3.7337	0.007542	0.04215
AC	14.6409	0.002136	0.03985
$ACCH_3$	8.2130	0.019360	0.10364
$ACCH_2$	10.3239	0.012200	0.10099
ACCH	10.4664	0.002769	0.07120
OH	9.7292	0.005148	0.03897
ACOH	25.9145	−0.007444	0.03162
CH_3CO	13.2896	0.025073	0.13396
CH_2CO	14.6273	0.017841	0.11195
CHO	10.1986	0.014091	0.08635
CH_3COO	12.5965	0.029020	0.15890
CH_2COO	3.8116	0.021836	0.13649
HCOO	11.6057	0.013797	0.10565
CH_3O	6.4737	0.020440	0.08746
CH_2O	6.0723	0.015135	0.07286
CH–O	5.0663	0.009857	0.05865
FCH_2O	9.5059	0.009011	0.06858
CH_2NH_2	12.1726	0.012558	0.13128
$CHNH_2$	10.2075	0.010694	0.07527
CH_3NH	9.8544	0.012589	0.12152
CH_2NH	10.4677	0.010390	0.09956
CHNH	7.2121	−0.000462	0.09165
CH_3N	7.6924	0.015874	0.12598
CH_2N	5.5172	0.004917	0.06705
$ACNH_2$	28.7570	0.001120	0.06358
C_5H_4N	29.1528	0.029565	0.24831
C_5H_3N	27.9464	0.025653	0.17027
CH_2CN	20.3781	0.036133	0.15831
COOH	23.7593	0.011507	0.10188

Notes: These values represent the contributions (G_i) of each group to the respective thermodynamic property and are used in linear summation models to predict critical properties.

deviation of 0.89% for T_c, 2.89% for P_c and 1.79% for V_c, which shows a better performance than Joback method.

In summary, CG94 represented a significant advancement in the field of GCMs by introducing a systematic and comprehensive framework of second order groups, which in a way addressed the limitations of earlier models. By the integration of both first order and second order groups, this method successfully captured not only the fundamental contributions of individual functional groups but also the critical influence of structural features such as conjugation, branching, and ring strain. A clear set of principles for defining second order groups made sure that the method can accurately represent localised molecular interactions essential for reliable property estimation. With its hierarchical design and consistent group structure applied across multiple thermodynamic and physical properties, CG94 achieved a superior accuracy and capability at that time. Its performance over a broad dataset and its ability to maintain internal consistency across diverse properties established itself as one of the most influential frameworks in GCMs [39]. Despite its significant contributions and improved accuracy over earlier methods, the CG94 model has several limitations. First, its development was based on the experimental data available at the time, which imposed a constraint on the diversity of compounds used for model calibration [40]. For instance, the experimental data set used for the regression included a limited number of complex aromatic, cyclic, and highly branched molecules, which may be less accurate to such structures, despite its strong extrapolating performance. While the model introduced second order groups to better account for local interactions, its physical foundation is somewhat limited–only the second order terms were derived based on structural theory and the first order contributions remained in the form of a conventional additive group term (as per the group contribution principle) [41]. Due to the sparsity of reliable data, several groups were not assigned any contribution values during regression, leading to gaps in model generality. The method also struggles to fully describe long-range interactions and global molecular effects such as conformational flexibility or electronic delocalisation, which are critical for large or highly interactive systems. As a result, CG94 may fail to capture the complexity of modern molecular design challenges without further refinement or integration with hybrid, data-driven approaches especially when distinguishing between branched alkane isomers [5]. Therefore, CG94 requires further reinforcement with molecular theories to reliably capture the properties of very complex organic structures and isomeric behaviours. Nevertheless, CG94 still remained one of the most reliable and transferable approaches available and has laid the foundation for the development of even more sophisticated GCMs. It introduces the second order approximation and demonstrates how molecular theories can be embedded into group contributions through second order groups without sacrificing the simplicity and applicability of group contributions [42]. It has also been applied to one of the most extensive sets of property estimations, including pure component properties, polymer properties, temperature-dependent properties, and mixture properties [43, 44]. Finally, it was shown to have robust extrapolating behaviour [45]. Building on the principles of CG94, Marrerro and Gani [20] added a correction to the original model by introducing a third order term. Using an extended database, the updated approach improved applicability in interpolation behaviour. Nevertheless, it may not be able to distinguish among the complex isomer structures, to extensively encounter molecular symmetry

Table 6. Second order group contribution values D_j for the estimation of T_{c2}, P_{c2}, and V_{c2} in CG94 [8].

Group	$T_{c2}/[\mathrm{K}]$	$P_{c2}/[\mathrm{bar}^{-0.5}]$	$V_{c2}/[\mathrm{m}^3/\mathrm{kmol}]$
$(CH_3)_2CH$	−0.5334	0.000488	0.00400
$(CH_3)_2C$	−0.5143	0.001410	0.00572
$CH(CH_3)CH(CH_3)$	1.0699	−0.001849	−0.00398
$CH(CH_3)C(CH_3)_2$	1.9886	−0.005198	−0.01081
$C(CH_3)_2C(CH_3)_2$	5.8254	−0.013230	−0.02300
3 membered ring*	−2.3305	0.003714	0.00401
4 membered ring*	−1.2978	0.001171	−0.00851
5 membered ring*	−0.6785	0.000424	−0.00866
6 membered ring*	0.8479	0.002257	0.01636
7 membered ring*	3.6714	−0.009799	−0.02700
$CH_n = CH_m\text{-}CH_p = CH_k$	0.4402	0.004186	−0.00781
$CH_3\text{-}CH_m = CH_n$	0.0167	−0.000183	−0.00098
$CH_2\text{-}CH_m = CH_n$	−0.5231	0.003538	0.00281
$CH\text{-}CH_m = CH_n$ or $C\text{-}CH_m = CH_n$	−0.3850	0.005675	0.00826
Alicyclic side chain $C_{cyclic}C_m$ ($m > 1$)	2.1160	−0.002546	−0.01755
CH_3CH_3	2.0427	0.005175	0.00227
CHCHO or CCHO	−1.5826	0.003659	−0.00664
CH_3COCH_2	0.2996	0.001474	−0.00510
CH_3COCH or CH_3COC	0.5018	−0.002303	−0.00122
$C_{cyclic}(=O)$	2.9571	0.003818	−0.01966
ACCHO	1.1696	−0.002481	0.00664
CHCOOH or CCOOH	−1.7493	0.004920	0.00559
ACCOOH	6.1279	0.000344	−0.00415
CH_3COOCH or CH_3COOC	−1.3406	0.000659	−0.00293
$COCH_2COO$ or COCHCOO or COCCOO	2.5413	0.001067	−0.00591
CO–O–CO	−2.7617	−0.004877	−0.00144
ACCOO	−3.4235	−0.000541	0.02605
CHOH	−2.8035	−0.004393	−0.00777
COH	−3.5442	0.000178	0.01511

*Stress-strain ring corrections treated as second order terms.

*Notes: These corrections are added to account for structural effects such as branching, ring strain, and conjugation that are not captured by first order groups.

and steric effects. The third order term introduced additional complexity and a very big number of adjustable parameters in the regression disproportional to the number of experimental data were used [46]. Finally, Stefanis and Panagiotou [47] successfully adopted the CG94 method to the estimation of the Hansen solubility parameters.

3.4. Sharma method

Recent GCMs have aimed to systematically include higher order structural effects to improve predictive accuracy particularly for large and highly branched molecules, which is essential to industrial processes such as hydroisomerisation, hydrocracking, and fuel production where precise knowledge of properties like ΔG_f and ΔH_f is essential for predicting reaction equilibria, optimising catalyst performance, and designing energy-efficient processes [48]. In isomerisation applications, where small structural differences critically affect thermodynamic stability and equilibrium, conventional GCMs fall short due to the inability to distinguish long-chain isomers as the reliability of first order group counts ignores the positional and topological context of each group. As a result, molecules with identical group compositions but different connectivities often yield less accurate predictions [49]. To address these limitations, Sharma et al. [5] introduced an advanced group contribution framework that fully relies on a comprehensive set of second order groups into a linear regression scheme for the accurate prediction of ΔG_f and ΔH_f of hydrocarbons. Unlike earlier models that often relied on predefined functional or simple group contributions and limited structural corrections, this method came up with an exhaustive enumeration of only second order groups to account for the detailed local environments within a molecule, which can provide greater structural sensitivity and enhanced predictive performance. The main feature of the Sharma method is its comprehensive and systematic use of second order groups to accurately represent the local structural environment for long-chain and highly branched alkanes. Unlike previous group contribution methods that apply a limited or heuristic selection of second order corrections, this model exhaustively enumerates all the possible atom combinations surrounding a central atom and forms second order groups present within a molecule. It captures the effects of adjacent group interactions, branching patterns, and local connectivity, which are particularly important in iso-alkanes where subtle variations in branching can significantly affect the thermochemical properties. This complete representation of second order structural features allowed the model to distinguish between molecules with identical first order group compositions but different topologies, addressing a major limitation of conventional GCMs when applied to large isomeric systems. The definition of second order groups were built by identifying all the possible atom combinations surrounding a central atom. These pairs account for direct connections between neighbouring carbon atoms with the respective substitution patterns to reflect the branching and connectivity. The procedure required enumerating every bonded pair of carbons, recording the nature of surrounding substituents of each pair to capture the subtle differences between linear, branched, and highly branched structures. Through this exhaustive identification of bonded pairs across the entire molecule, the model generates a complete set of second order descriptors, ensuring that no significant local structural variation is overlooked. This rigorous method allows the model to systematically capture the influence of localised branching, steric effects, and chain topology, all of which are critical for accurately estimating the thermochemical properties of long-chain and highly branched alkanes. The developers of the Sharma method are also co-authors of this manuscript.

The model was calibrated using a dataset of 3325 long-chain alkane isomers [5]. These were selected from iso-alkanes ranging from C_1 to C_{14} to ensure diversity in chain length and branching and then the model extrapolated up to C_{20}. This calibration was enabled by the thermodynamic tables compiled by Scott [9]. These tables cover a broad range of linear and branched alkanes based on high-level quantum chemical calculations and provide consistent and reliable thermochemical data. Examples of second order group defined in the Sharma method can be seen in the treatment of local environments surrounding a central CH_2 unit, as shown in Figure 1. In both cases (a) and (b), the central united atom is CH_2, and if only first order group contributions would be used, these two environments would be treated identically, which may lead to inaccurate property predictions because the influence of neighbouring groups would be neglected. By defining second order groups that explicitly incorporate the identities of adjacent groups, the model can distinguish between these environments. This differentiation distinguishes the variations in branching and local connectivity, which can significantly affect thermochemical properties. By considering not only the central group but also its immediate neighbours, the second order group contribution framework can improve structural sensitivity and enhance predictive accuracy, particularly for isomeric systems where subtle differences in local structure have pronounced thermodynamic effects. Following this idea, 69 kinds of second order groups are defined and can be found in Table 7. As a result, the introduced methodology enable higher accuracy in predicting thermochemical properties of alkanes compared to the existing general group contribution methods in literature, and was developed with hydroisomerisation applications in mind [5].

Another important feature of the Sharma method is its implementation of temperature-dependent group contributions through polynomial regression. The model uses a quadratic polynomial function to each second order group, enabling the accurate prediction of properties for a temperature range from 0 to 1000 K based on the fact that the influence of structural features on thermodynamic properties varies with temperature. This approach allowed the model to move beyond the fixed, temperature-independent parameters of traditional GCMs, and provided a more flexible and physically consistent framework that accounts for the thermal behaviour of complex hydrocarbons. The resulting model is therefore highly tailored to the specific challenges of long-chain alkanes. The model demonstrated excellent predictive accuracy for key thermochemical properties, achieving chemical accuracy within 1 kcal/mol for properties such as ΔG_f and ΔH_f. Despite these advances, the Sharma method remains limited to alkanes and its applicability to other classes of compounds is uncertain. This is because extending the second order group definitions to molecules containing heteroatoms (such as oxygen, nitrogen, or sulfur) would require an explosion in the number of group types, which significantly increases model complexity and reducing practicality. While the single usage of second order groups effectively captured local interactions, the method does not explicitly account for long-range intramolecular effects or global conformational changes, which can affect the properties of very large or highly flexible molecules. In summary, Sharma et al. [5] have provided a new thought for GCMs by only defining a set of second order groups. Through its systematic second order group framework, temperature-dependent corrections, and robust regression strategy, the method has successfully presented a significant progress in predicting thermochemical properties for complex, branched hydrocarbons.

3.5. Applications for the prediction of thermodynamics properties of pure compounds

Presently, a large number of GCMs have been widely used for prediction and estimations of thermodynamic properties of pure compounds in both academics and industry. One example is the work of Nannoolal et al. [34], who developed a refined group contribution approach

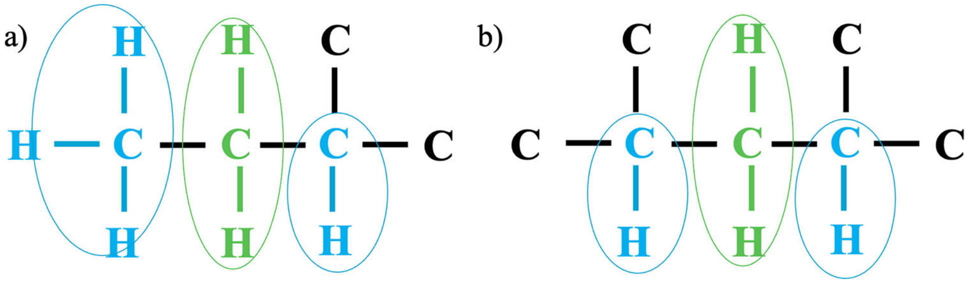


Figure 1. Typical examples of second order groups (a) $CH_2(CH_3)$-(CH) and (b) $CH_2(CH)(CH)$ with CH_2 as the central united atom. Each second order group includes the central atom (green) and its two adjacent bonded fragments (blue) [5].

Table 7. Second order group definitions used in the Sharma method for branched alkanes [5].

Category	Groups
No neighbouring groups (CH_4)	CH_4
1 neighbouring group (CH_3 as the centre atom)	$CH_3(C)$, $CH_3(CH)$, $CH_3(CH_2)$, $CH_3(CH_3)$
2 neighbouring groups (CH_2 as the centre atom)	$CH_2(C)(C)$, $CH_2(C)(CH)$, $CH_2(C)(CH_2)$, $CH_2(C)(CH_3)$, $CH_2(CH)(CH)$, $CH_2(CH)(CH_2)$, $CH_2(CH)(CH_3)$, $CH_2(CH_2)(CH_2)$, $CH_2(CH_2)(CH_3)$, $CH_2(CH_3)(CH_3)$
3 neighbouring groups (CH as the centre atom)	$CH(C)(C)(C)$, $CH(C)(C)(CH)$, $CH(C)(C)(CH_2)$, $CH(C)(C)(CH_3)$, $CH(C)(CH)(CH)$, $CH(C)(CH)(CH_2)$, $CH(C)(CH)(CH_3)$, $CH(C)(CH_2)(CH_2)$, $CH(C)(CH_2)(CH_3)$, $CH(C)(CH_3)(CH_3)$, $CH(CH)(CH)(CH)$, $CH(CH)(CH)(CH_2)$, $CH(CH)(CH)(CH_3)$, $CH(CH)(CH_2)(CH_2)$, $CH(CH)(CH_2)(CH_3)$, $CH(CH)(CH_3)(CH_3)$, $CH(CH_2)(CH_2)(CH_2)$, $CH(CH_2)(CH_2)(CH_3)$, $CH(CH_2)(CH_3)(CH_3)$, $CH(CH_3)(CH_3)(CH_3)$
4 neighbouring groups (C as the centre atom)	$C(C)(C)(C)(C)$, $C(C)(C)(C)(CH)$, $C(C)(C)(C)(CH_2)$, $C(C)(C)(C)(CH_3)$, $C(C)(C)(CH)(CH)$, $C(C)(C)(CH)(CH_2)$, $C(C)(C)(CH)(CH_3)$, $C(C)(C)(CH_2)(CH_2)$, $C(C)(C)(CH_2)(CH_3)$, $C(C)(C)(CH_3)(CH_3)$, $C(C)(CH)(CH)(CH)$, $C(C)(CH)(CH)(CH_2)$, $C(C)(CH)(CH)(CH_3)$, $C(C)(CH)(CH_2)(CH_2)$, $C(C)(CH)(CH_2)(CH_3)$, $C(C)(CH)(CH_3)(CH_3)$, $C(C)(CH_2)(CH_2)(CH_2)$, $C(C)(CH_2)(CH_2)(CH_3)$, $C(C)(CH_2)(CH_3)(CH_3)$, $C(C)(CH_3)(CH_3)(CH_3)$, $C(CH)(CH)(CH)(CH)$, $C(CH)(CH)(CH)(CH_2)$, $C(CH)(CH)(CH)(CH_3)$, $C(CH)(CH)(CH_2)(CH_2)$, $C(CH)(CH)(CH_2)(CH_3)$, $C(CH)(CH)(CH_3)(CH_3)$, $C(CH)(CH_2)(CH_2)(CH_2)$, $C(CH)(CH_2)(CH_2)(CH_3)$, $C(CH)(CH_2)(CH_3)(CH_3)$, $C(CH)(CH_3)(CH_3)(CH_3)$, $C(CH_2)(CH_2)(CH_2)(CH_2)$, $C(CH_2)(CH_2)(CH_2)(CH_3)$, $C(CH_2)(CH_2)(CH_3)(CH_3)$, $C(CH_2)(CH_3)(CH_3)(CH_3)$, $C(CH_3)(CH_3)(CH_3)(CH_3)$

for estimating T_c, V_c and P_c. This method was particularly applied to create a comprehensive and consistent database of critical properties for a wide range of organic compounds for process design and simulation in the petrochemical industry where accurate critical property data are required for phase equilibrium modelling, distillation column design, and safety assessments of hydrocarbons. Marrero and Gani [20] presented a group contribution framework to estimate multiple thermophysical properties for the application of solvent screening for separation processes, where hundreds of potential solvents must be evaluated for the volatility, thermal stability, and compatibility with target solutes. In such design scenarios, GCMs enable rapid pre-screening of candidates based on predicted properties before experimental validation, which can save significant time and resources. Another example is from Rarey et al. [50], who focussed on predicting the thermal conductivity of organic liquids through group contributions. This model has been particularly relevant in thermal management of heat transfer fluids, where the thermal conductivity of potential components was investigated for designing efficient heat exchangers, cooling systems, and other thermal processes in the chemical and energy sectors. Recently, Csemány et al. [13] evaluated various material property estimation methods, including several GCMs, for alkanes with a focus on modelling droplet evaporation processes in combustion applications. The study showed the importance of accurate prediction of critical properties, vapor pressure, and transport properties (such as thermal conductivity and viscosity) where experimental data are scarce, demonstrating the continuing relevance of GCMs in the energy sector. Groniewsky and Hégely [51] proposed an extension of CG94 through an automated group conversion procedure to broaden the applicability of GCMs to predict vapor pressure and improve acentric factor estimation, which is especially significant for applications in process engineering where accurate vapor-liquid equilibrium (VLE) data are essential, but traditional GCM frameworks face limitations due to incomplete group databases. Overall, these methods show a wide range of ways group contribution methods that are used across different industries, including petrochemical process design, separation technologies, thermal systems, and sustainable chemical development. In all these fields, GCMs are valuable because of the quick, reliable, and broadly applicable property predictions, helping researchers and engineers make informed decisions when experimental data are limited or unavailable.

4. Applications to mixture property estimation

In industrial practice, most operations, like distillation, extraction, absorption, and reaction, deal with complex mixtures rather than individual pure components [52]. As a result, understanding how pure component properties combine to define the behaviour of mixtures is essential for accurate process simulation and equipment design. In these situations, GCMs are commonly combined with excess Gibbs energy models to predict mixture properties that are intermolecular like activity coefficients and phase equilibria [53]. These models use group contribution concepts to estimate the non-ideal interactions between different functional groups in a mixture [54]. Such predictions are especially important when dealing with multicomponent systems where experimental data are unavailable or incomplete, as is often the case for new solvents, specialty chemicals, or bio-based feedstocks. A typical example of mixture property estimation using GCMs is in solvent recovery and recycling, where knowledge of VLE behaviour helps optimise distillation sequences and minimise energy consumption [55]. Another important application is in extractive distillation, where the interaction between solvent and solute molecules governs the separation efficiency [56]. Fuller et al. [57] developed one of the earliest and most widely adopted GCMs for estimating binary gas-phase diffusion coefficients, which introduced the concept of diffusion volumes assigned to molecular fragments and calculated the diffusion coefficient via structural information from each component. The method was calibrated using a large dataset of 340 binary systems and remained favoured in engineering applications due to its balance of simplicity, broad applicability, and reasonable accuracy. Similarly, in the design of liquid-liquid extraction processes, GCM-based mixture models can help identify suitable solvent systems by predicting phase separation and component distribution without the need for exhaustive experimental testing [58]. A significant advancement in predicting mixture properties, especially for systems involving complex and associating molecules, has been the integration of GCMs with molecular-based equations of state. One widely recognised framework is the Statistical Associating Fluid Theory (SAFT) [59], which models fluids based on rigorous statistical thermodynamics by considering molecular size, shape, polarity, and hydrogen bonding interactions explicitly. Variants such as PC-SAFT (Perturbed Chain Statistical Associating Fluid Theory) [60] extend the SAFT framework by modelling molecules as chains of spherical segments with explicit association sites, which enables the accurate description of long-chain, associating, and polar compounds [43]. Papaioannou et al. [61] propose a novel Group Contribution PC-SAFT methodology to predict the thermodynamic behaviour of complex heteronuclear molecules and multicomponent mixtures where group contributions are defined to fit PC-SAFT parameters systematically, enabling the estimation of key thermodynamic properties such as phase equilibria and densities even in the absence of extensive experimental data. The approach proved particularly effective for mixtures involving polar, associating, and long-chain components, thereby broadening the applicability of predictive molecular thermodynamics to challenging chemical systems. GCMs can also be powerful tools for the estimation of activity coefficients, excess enthalpies, and phase equilibria in binary and multicomponent systems [44]. In mixtures, GCMs like UNIFAC [62] (UNIQUAC [63] Functional-group Activity Coefficients) and its variants (e.g. Effective UNIFAC) can capture both combinatorial effects (due to molecular size and shape) and residual interactions (due to energy differences among functional groups), enabling accurate modelling of vapor–liquid, liquid–liquid, and solid–liquid equilibria [63]. For instance, these methods can be used to predict azeotropic behaviour, miscibility gaps, and solubility of solids in solvents–all crucial in process design. When enhanced with more precise group definitions or hybridised with data-driven models, GCMs offer a versatile platform for handling the increasing complexity of real-world mixture behaviour [17]. In addition, the FeOs framework [64] offers a modern, open-source platform implementing PC-SAFT equations of state and classical density functional theory to model both homogeneous and inhomogeneous systems. Recently, Hayer et al. [65] proposed UNIFAC 2.0, a hybrid machine learning-enhanced group contribution method trained on over 224,000 experimental data points that systematically completes missing interaction parameters using matrix completion techniques, leading to significantly improved prediction accuracy and a much broader applicability domain compared to UNIFAC 1.0. Isamu Nagata and Jitsuo Koyabu [66] presente a modified GCM–termed the Effective UNIFAC model–designed to improve the estimation of activity coefficients for predicting phase equilibria. Detailed group volume and area parameters, group-interaction matrices, and validation examples such as binary alcohol-hydrocarbon systems and ternary systems like ethanol–water–ethyl acetate are included and predictions for solubility and excess enthalpy are provided. This method is rooted in an extension of the effective UNIQUAC equation and is tailored to handle vapor–liquid, liquid–liquid, and solid–liquid equilibria, especially in complex binary and multicomponent mixtures. The model relies on group-interaction parameters derived from experimental data and achieves improved prediction accuracy, particularly for systems involving alcohols, hydrocarbons, and polar components.

In summary, GCMs have proven to be a tool for the estimation of mixture properties, for accurate predictions of phase equilibria, activity coefficients, and other thermodynamic behaviours in multicomponent systems. By extending the principles of pure component property estimation to mixtures, GCMs are capable for the modelling of complex interactions between different functional groups without requiring extensive experimental data. By the integration of GCMs with established excess Gibbs energy models, such as UNIFAC and its variants, these methods provide reliable and transferable predictions that support process design, optimisation, and scale-up across a wide range of chemical industries.

5. Application to process and product design, optimisation and reaction kinetics

Besides the estimation of pure component and mixture properties, GCMs can also be used in the broader context of process and product design and optimisation [67]. In modern chemical engineering, process development increasingly relies on predictive models that can evaluate not only the feasibility of individual compounds but also the performance of entire process systems. GCMs provide a critical foundation in this regard by enabling the rapid estimation of key thermodynamic and physical properties that feed directly into process simulation, flowsheet design, energy integration, and economic evaluation. By coupling GCMs with process modelling environments, engineers can efficiently explore a wide range of design scenarios, even when working with hypothetical molecules or novel compounds for which no experimental data exist [68]. This capability is particularly important in the early stages of process synthesis, where screening thousands of possible chemical pathways requires fast and reliable property predictions [21]. GCMs are also deeply integrated into Computer-Aided Molecular Design (CAMD) frameworks to support the simultaneous optimisation of both molecular structure and process performance [69, 70]. In such workflows, GCMs allow the property estimation of candidate molecules to be directly linked to process objectives such as minimising energy consumption, maximising product yield, or reducing environmental impact [71]. Harper et al. [72] developed an extended CAMD methodology by integrating traditional GCMs with molecular modelling tools and including a structured approach that progressively refines candidate molecules across four levels, from group vector assembly to atomic-level structure generation and 3D molecular modelling using Chem3D. Two industrial application examples were presented in Ref.[72]: the first involved the design of alternative solvents to replace toluene for phenol removal from wastewater, where candidate solvents were generated, screened, and optimised based on multiple property constraints including environmental impact and separation performance. The second example focussed on designing solvents for extractive distillation to separate a close-boiling organic acid mixture, demonstrating the ability of the CAMD method to handle complex molecular architectures and phase behaviour modelling. This work shows how combining GCMs with detailed molecular modelling can improve property prediction accuracy, especially when traditional GCMs are insufficient, and the potential for CAMD in industrial solvent design, process optimisation, and material innovation. A key advantage of using GCMs in process optimisation is the ability to support multi-scale decision-making, bridging molecular design, mixture behaviour, and full-process performance. For instance, in the development of alternative refrigerants, green solvents, or bio-based feedstocks, group contribution-based models are used not only to predict the properties of new substances but also to evaluate the performance within the overall process, accounting for separation efficiency, energy demands, and lifecycle emissions [73]. One example is the work of Gmehling [74], where GCMs were applied to optimise complex distillation processes, including the design of distillation columns and the selection of suitable solvents for azeotropic separations. These applications enabled efficient process simulation and flowsheet development, particularly when experimental data are limited or when working with multicomponent systems that exhibit strong non idealities. This study also emphasised the potential for further improvement of GCMs to handle increasingly complex mixtures and systems encountered in industrial practice. Traditionally, GCMs have been predominantly used for the prediction of thermodynamic and physical properties. The application of GCMs has been successfully extended into the field of chemical kinetics. In this context, GCMs are used to estimate kinetic parameters such as reaction rate constants (k) and activation energies (E_{act}) by decomposing complex molecules into functional groups whose individual contributions can be systematically summed [75]. This approach enables the rapid prediction of reaction behaviour for large sets of molecules, significantly aiding in the design and environmental assessment of chemical processes where experimental kinetic data may be sparse or unavailable [16]. Minakata et al. [75] applied a group contribution method to predict aqueous phase hydroxyl radical reaction rate constants for various organic compounds, supporting environmental assessments of pollutant degradation. Saeys et al. [76] propose an ”ab initio group contribution method” for estimating activation energies of hydrogen abstraction reactions, which combined quantum mechanical knowledge with group contributions to improve reaction rate predictions. In these studies, GCMs have proven effective in providing kinetic parameters across diverse reaction systems, from biological networks to environmental processes and catalytic mechanisms. This fast, transferable, and systematic way of estimating reaction kinetics based on molecular structure is becoming a handy tool for both fundamental research and industrial process development.

6. Future directions and research opportunities

Recent advances in GCM have significantly improved the accuracy and generalisability of property prediction, yet several challenges and opportunities remain. Second order group contributions appear to be a particularly promising direction, which provided an effective balance between structural complexity and model accuracy, capturing critical local interactions such as branching, conjugation, and hydrogen bonding [8, 42]. Current studies suggest that while second order groups considerably improve model fidelity, pursuing higher order group contributions offers limited additional benefit and often leads to diminishing returns with increased model complexity [77, 78]. Therefore, the focus remains on refining second order contributions rather than extending to higher orders [50]. A key priority is to reinforce the group contribution framework with molecular-level theories while maintaining the second order approximation. Quantum chemical calculations and statistical mechanics may help ground the definition of second order groups in physically meaningful terms. This integration can support the consistent treatment of effects like molecular symmetry, which is especially important for accurately modelling isomeric species and complex structural motifs [79]. Such refinement would also help overcome one of the major current limitations: the inability of many GCMs to distinguish between structurally similar but functionally different molecules. Despite the rapid adoption of machine learning and AI in this field, human expertise remains essential, particularly in thermodynamics, process design, and model interpretation, as GCMs are grounded in chemical theory and human expertise is essential to define meaningful group structures, ensure thermodynamic consistency, and interpret results–tasks that machine learning alone cannot reliably perform. Chemical insight is still required to define meaningful molecular fragments, detect thermodynamic inconsistencies, and guide extrapolation in underexplored chemical domains [27]. Future efforts should also focus on strengthening extrapolation capabilities. While GCMs perform well in interpolation, errors and failures occur when extended beyond original training domains. To address this, better utilisation of existing databases–for both interpolation and extrapolation–is needed [7]. Group identification tools and robust similarity metrics can play a key role in mapping new molecules to known group environments. Simultaneously, the availability and coverage of property data must be improved, which in turn supports more reliable process simulation and inverse property estimation frameworks. With these strengthened foundations, GCMs will be better positioned to contribute to sustainable process and product development. Reliable property prediction enables more informed decisions in solvent selection, material screening, and lifecycle design, ultimately accelerating innovation in green chemistry and circular manufacturing.

One of challenges in GCMs is the ability to extrapolate to branched large hydrocarbons reliably beyond the range of data used for parameterisation. While GCMs have proven effective for predicting the properties of well characterised molecules within established chemical families, the accuracy decreases when applied to compounds with novel structural motifs, extreme operating conditions, or larger molecular sizes. This limitation arises because traditional GCMs are often calibrated using datasets that primarily cover small to medium-sized molecules, with limited diversity in functional groups and topologies [35]. Future work must focus on the development of robust extrapolation strategies that can extend the applicability of GCMs to broader chemical spaces. One promising direction involves the use of hierarchical models, where first order and second order contributions are complemented by additional correction terms that capture higher order interactions and long range structural effects [80]. Hybrid approaches that integrate group contribution frameworks with physics-based models, such as equations of state or molecular simulations, may provide better extrapolative capabilities by grounding empirical contributions in more fundamental thermodynamic principles [81, 82]. As traditional GCMs largely rely on empirical correlations derived from experimental data, most models often struggle with accuracy when applied to compounds with unusual electronic structures or rare functional groups. Recent research [7, 14] has explored the integration of physical chemistry principles and quantum mechanical calculations into GCMs. This approach aims to enhance the fundamental understanding of molecular interactions, thus improving both the predictive accuracy and the transferability of GCMs to systems beyond the range of available data. By combining

Table 8. Comparison between enthalpies of formation, ΔH_f at 298.15 K for C_{14} isomers using group contribution methods by Benson et al. [10], Constantinou and Gani (CG94) [8], Sharma et al. [5], and group contribution-based graph convolution network by Hwang and Kong [12].

Isomer	SMILES	Benson et al. [kJ/mol]	CG94 [kJ/mol]	Sharma et al. [kJ/mol]	Hwang and Kong [kJ/mol]
C_{14}	CCCCCCCCCCCCCC	−332.06	−329.36	−332.03	−341.08
2-m-C_{13}	CC(C)CCCCCCCCCCC	−337.57	−337.72	−338.83	−346.26
7-m-C_{13}	CCCCCCC(C)CCCCCC	−334.22	−337.30	−336.67	−341.34
2,2-m-C_{12}	CC(C)(C)CCCCCCCCCC	−345.73	−342.90	−349.29	−359.98
2,10-m-C_{12}	CC(C)CCCCCCCC(C)CC	−339.73	−345.66	−343.27	−347.09
2,11-m-C_{12}	CC(C)CCCCCCCCC(C)C	−343.08	−346.08	−345.63	−351.45
4,9-m-C_{12}	CCCC(C)CCCCC(C)CCC	−336.38	−345.24	−341.31	−341.99
5,5-m-C_{12}	CCCCC(C)(C)CCCCCCC	−339.03	−342.90	−344.71	−349.52
5,6-m-C_{12}	CCCCC(C)C(C)CCCCCC	−336.38	−339.18	−336.12	−337.25
2,2,3-m-C_{11}	CC(C)(C)C(C)CCCCCCCC	−344.54	−342.77	−344.83	−353.83
2,2,4-m-C_{11}	CC(C)(C)CC(C)CCCCCCC	−347.89	−350.84	−346.55	−354.39
2,2,6,7-m-C_{10}	CC(C)(C)CCCC(C)C(C)CCC	−350.05	−352.72	−353.38	−356.39
2,2,6,8-m-C_{10}	CC(C)(C)CCCC(C)CC(C)CC	−350.05	−358.79	−358.65	−360.03
3-e-5,7-m-C_{10}	CCC(CC)CC(C)CC(C)CCC	−335.19	−353.19	−344.00	−337.91
3-e-5,8-m-C_{10}	CCC(CC)CC(C)CCC(C)CC	−335.19	−353.19	−343.52	−339.19
5,5-e-C_{10}	CCCCC(CC)(CC)CCCCC	−325.63	−342.90	−334.16	−339.56
5,6-e-C_{10}	CCCCC(CC)C(CC)CCCC	−329.68	−345.24	−331.44	−332.11
4,4-p-C_8	CCCC(CCC)(CCC)CCCC	−325.63	−342.90	−335.17	−337.11
4,5-p-C_8	CCCC(CCC)C(CCC)CCC	−329.68	−345.24	−331.84	−329.97
3-tb-2,2,4-m-C_7	CC(C)(C)C(C(C)(C)C)C(C)CCC	−353.67	−372.33	−330.51	−338.33
3-tb-2,2,5-m-C_7	CC(C)(C)C(C(C)(C)C)CC(C)CC	−353.67	−372.33	−338.51	−343.27

quantum mechanical, such as electron density distributions, orbital interactions, and non-covalent forces, phenomena like hydrogen bonding, resonance stabilisation, and polarisability features that are difficult to capture using only first order additive group contributions can be better described. Hybrid models, which combine quantum chemical calculations with group contribution schemes, offer a pathway to generate group parameters grounded in first-principles calculations rather than relying exclusively on experimental fitting. For example, Gani [7] discussed how integrating QM-derived parameters into property estimation models helps to address cases involving highly polar, reactive, or strained molecules that challenge conventional GCMs, the development of semiempirical quantum mechanical methods, as shown by Christensen et al. [14], computationally efficient ways to calculate molecular interactions and thermodynamic properties can be incorporated into GCMs to enhance the accuracy for systems involving non-covalent interactions. The reinforcement of molecular theories to group contributions should not be done at the expense of simplicity and wide applicability. Therefore, embedding those theories into the definition of first and/or second order groups will maintain the ability of group contributions to support any process and product design framework in a simple and effective manner.

Another limitation of current GCMs is that the reliance on the availability of high quality experimental data for parameter development and validation. Most existing GCMs are built on databases dominated by small, well-characterised organic molecules containing common functional groups [45]. As a result, these models often fail when applied to complex compounds, such as large biomolecules, polymers, ionic liquids, pharmaceuticals, and heavily branched hydrocarbons, for which experimental property data are lacking. Without robust experimental data covering a diverse range of chemical structures, the extrapolation of GCMs to new molecular spaces becomes unreliable, compromising the accuracy and generalisability of property predictions. A key future direction is the systematic expansion of experimental data libraries, focussing specifically on underrepresented compound classes and functional groups. This includes generating accurate thermophysical and thermochemical property data (such as critical properties, phase behaviour, heat capacities, viscosities, and enthalpies of formation) for structurally complex molecules. High-throughput experimental techniques, along with advanced calorimetry, spectroscopy, and chromatographic methods, are now making it more feasible to measure such properties across larger chemical spaces [83]. Collaborative efforts to develop standardised and open-access databases are essential for ensuring broad usability of new data for GCM development [84]. Recent initiatives such as the NIST ThermoData Engine [85] and Dortmund Data Bank [54] have played critical roles in expanding accessible datasets, but further efforts are required to include more heteroatomic species, ionic species, biodegradable compounds, and environmentally relevant pollutants. These enriched datasets will not only strengthen the accuracy of existing GCMs but also enable the creation of next-generation models capable of handling the increasing molecular diversity encountered in areas like biotechnology, green chemistry,

Table 9. A summary of GCMs applied to the prediction of thermophysical and thermochemical properties, including rate constants (k) and activation energies (E_{act}) for hydroxyl radical reactions and hydrogen abstraction reactions, along with representative references.

Property	Method/Author	Reference
T_{b}, T_{m}	Joback	[16]
	CG94	[8]
	Marrero-Pardillo	[20]
	Hwang	[12]
	Nannoolal, et al.	[80]
	He et al.	[92]
	Xue et al.	[93]
	Hou et al.	[94]
	Simamora et al.	[95]
P_c, V_c, T_c	Lydersen	[15]
	Joback	[16]
	CG94	[8]
	Marrero-Pardillo	[20]
	Wilson-Jasperson	[96]
	Nannoolal et al.	[34]
	Lan et al.	[97]
	He et al.	[92]
	Mann et al.	[2]
	Xue et al.	[93]
ΔH_f, ΔG_f	Benson	[10]
	CG94	[8]
	Holderbaum-Gmehling	[35]
	Mann et al.	[89]
	Sharma method	[5]
	Hwang	[12]
	Liu et al.	[98]
	Li et al.	[99]
C_p	Gardas	[100]
	Pietro et al.	[101]
	Han et al.	[102]
	Haghbakhsh et al.	[103]
	Ahmadi et al.	[104]
	Albert et al.	[105]
	CG94	[8]
	Villazón-León et al.	[106]
k, E_{act}	Minakata et al.	[75]
	Saeys et al.	[76]
	Liu et al.	[107]

and pharmaceutical design [86, 87]. Without sufficient experimental data, even the most sophisticated GCMs are constrained in predictive power. Therefore, continuous expansion and curation of high-quality, diverse experimental data libraries remain essential to the future advancement of GCMs.

The future of GCMs is closely related to advanced computational techniques like ML, AI and big data analytics. Traditional GCMs rely on linear and additive models, which is effective for many systems yet can fail to capture the non-linear, multi-scale interactions present in complex molecular systems. By combining GCM frameworks with modern computational strategies, it is now possible to overcome these limitations, improve predictive performance, and automate the development of new group parameters [88]. ML algorithms, such as neural networks, decision trees, and support vector machines, can analyze large molecular datasets to identify hidden patterns and complex relationships between structural features and properties. These data-driven models complement GCMs by refining group contributions, correcting systematic errors, and providing corrections where traditional group-additive assumptions break down. One example can be found in the work of Hwang [12] where the Group Contribution Graph Convolution Neural Network (GC-gcn), a hybrid model that combines the conventional GCM framework with graph convolution networks (GCNs) was introduced to estimate pure component thermodynamic properties. Instead of representing molecules with detailed atomic graphs, the GC-gcn model uses functional groups as nodes, drastically reducing the number of adjustable parameters and making it feasible to train accurate models even with limited thermophysical data. This showed an opportunity to combine machine learning and modern computational science with GCMs. Similarly, Mann et al. [89] propose a new perspective to address the limited flexibility in capturing complex molecular phenomena of traditional GCMs by introducing the integration of deep learning models to dynamically define and adjust group contributions, allowing more nuanced structural features to be encoded. In addition, the framework emphasised the use of explainable AI (XAI) to ensure that the resulting models retain physical interpretability, a crucial factor for reliable property prediction. This hybrid approach aims to improve both the predictive accuracy and generalisability of GCMs, especially for novel and diverse chemical spaces.

7. Conclusions

GCMs have been developed as a practical alternative tool to experimental methods that are often costly, time-consuming, or even impossible for complex and novel molecules for the prediction of thermodynamic properties of organic compounds. In this review, we have examined the development and evolution of GCMs, starting from early models like the Lydersen and Joback methods, which focussed on first order groups, to more advanced approaches, such as CG94 and the Sharma method, which used second order group corrections to capture intricate molecular interactions, branching effects, and conjugation patterns. These innovations significantly improved the accuracy and applicability of GCMs, particularly for large, highly branched, or isomeric molecules. Table 8 presents a comparison of standard enthalpies of formation, ΔH_{f}, for C_{14} isomers obtained using group contribution methods by Benson et al. [10], Constantinou and Gani (CG94) [8], Sharma et al. [5], and the group contribution-based graph convolution network (GC-GCN) developed by Hwang and Kang [12]. The results show that all methods provide comparable ΔH_f values, indicating consistency across both group contribution methods and hybrid method consisting of group contribution and machine learning. In Table 9, an overview of widely used and recently developed GCMs is provided as a reference for the application to different properties. Despite these advances in GCMs, challenges still remain. Current GCMs face limitations when extrapolating to new chemical spaces, handling extreme conditions, especially when temperature is lower than 298K, or dealing with complex compounds lacking sufficient experimental data, especially for isomers. Consequently, future research should focus on enhancing extrapolation models, integrating quantum mechanical insights, expanding experimental datasets for underrepresented molecules, and combining advanced computational techniques, including machine learning, to build more robust, transferable, and accurate frameworks. Accurate estimation of thermophysical properties is essential for process design, optimisation, and safety analysis in chemical engineering. While high quality experimental data such as those from calorimetry or vapor–liquid equilibrium measurements–can achieve accuracies within ca. 1–2% for many properties [90], the vast chemical design space and the emergence of novel compounds make exhaustive experimentation impractical. GCMs provide a computationally efficient alternative by predicting properties from molecular structure. The accuracy of GCMs is inherently limited by the quality of the experimental or QM data used for parameter fitting. As expectations for predictive models grow, the gap between available experimental accuracy and GCM output must be critically examined. Traditional GCMs often show reduced reliability when extrapolated to complex molecules or multicomponent systems while recent efforts integrated ML with GCM frameworks enables models to learn from data patterns and correct systematic errors [91]. Ultimately, aligning the accuracy of GCM predictions with the fidelity of underlying data is critical for advancing in reliable process modelling and sustainable chemical design. In evaluating future directions for GCM development, it is essential to recognise the inherent trade-off between simplicity and accuracy. Simpler models that rely on fewer parameters and offer greater robustness and easier extrapolation to new chemical spaces albeit sometimes at the expense of fine-grained predictive accuracy. In sharp contrast, more advanced ML-enhanced approaches can achieve higher predictive accuracy but may suffer from reduced extrapolation capabilities and risk overfitting to specific datasets [89]. Therefore, the desired level of accuracy must be carefully considered based on the application context, as different engineering problems demand different degrees of precision in thermophysical property estimation. Balancing model simplicity, computational cost, interpretability, and the required predictive fidelity will remain a central theme in advancing future GCM frameworks. In summary, GCMs continue to evolve as powerful, adaptable tools that bridge the gap between molecular structure and macroscopic properties. With rapid developments, GCMs are expected to play an increasingly important role in advancing green chemistry, process optimisation, and the design of next-generation materials and fuels for more efficient, sustainable, and innovative chemical processes.

Disclosure statement

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

Funding

This work is funded by the Advanced Research Center for Chemical Building Blocks, ARC-CBBC, which is cofunded and cofinanced by The Netherlands Organisation 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] M.S. Rigutto, R. van Veen and L. Huve, in *Studies in Surface Science and Catalysis* (Elsevier, Amsterdam, 2007), Vol. 168, pp. 855–XXVI.

[2] V. Mann, K. Brito, R. Gani and V. Venkatasubramanian, Fluid. Phase. Equilib. **561**,113531(2022). doi:10.1016/j.fluid.2022.113531.

[3] Y. Yang, X. He, G. Wu, Y. Yuan and Z. Liu, ACS Energy Lett. **5**, 2125–2143 (2020).

[4] L. Tao, J.J. Jacobson, L. Zhang, M.A. Jackson, D.B. Hodge, C. Kinchin and M. Wang, Biofuel. Bioprod. Biorefin. **11**, 965–980 (2017).

[5] 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.

[6] J. Gmehling, Fluid. Phase. Equilib. **30**, 119–134 (1986). doi:10.1016/0378-3812(86)80047-4.

[7] R. Gani, Curr. Opin. Chem. Eng. **23**, 184–196 (2019). doi:10.1016/j.coche.2019.04.007.

[8] L. Constantinou and R. Gani, AIChE J. **40**, 1697–1707 (1994). doi:10.1002/aic.v40:10.

[9] D.W. Scott, J. Chem. Phys. **60**, 3144–3165 (1974). doi:10.1063/1.1681500.

[10] S.W. Benson and J.H. Buss, J. Chem. Phys. **29**, 546–572 (1958). doi:10.1063/1.1744539.

[11] R. Gani, B. Nielsen and A. Fredenslund, AIChE J. **37**, 1318–1332 (1991). doi:10.1002/aic.v37:9.

[12] S. Hwang and J. Kang, Int. J. Thermophys. **43**, 9–42 (2022). doi:10.1007/s10765-022-03060-7.

[13] D. Csemány, L. Lisztes, G. Jancsó and B. Hégely, Fluid. Phase. Equilib. **535**,112998(2021).

[14] A.S. Christensen, T. Kubar, Q. Cui and M. Elstner, Chem. Rev. **116**, 5301–5337 (2016). doi:10.1021/acs.chemrev.5b00584.

[15] A.L. Lydersen, Engineering Experiment Station Report 3, College of Engineering, University of Wisconsin, Madison, Wisconsin, 1955.

[16] K.G. Joback and R.C. Reid, Chem. Eng. Commun. **57**, 233–243 (1987). doi:10.1080/00986448708960487.

[17] A. Klamt and F. Eckert, Annu. Rev. Chem. Biomol. Eng. **1**, 101–124 (2002). doi:10.1146/chembioeng.2010.1.issue-1.

[18] E.A. Brignole, S. Bottini and R. Gani, Fluid. Phase. Equilib. **29**, 125–132 (1986). doi:10.1016/0378-3812(86)85016-6.

[19] S. Macchietto, O. Odele and O. Omatsome, Trans. IChemE **68, Part A**, 429–433 (1990).

[20] J. Marrero and R. Gani, Ind. Eng. Chem. Res. **40**, 5256–5267 (2001).

[21] D. Dalmazzone, A. Salmon and S. Guella, Fluid. Phase. Equilib. **242**, 29–42 (2006). doi:10.1016/j.fluid.2005.12.034.

[22] S.I. Sandler and J. Wu, Ind. Eng. Chem. Res. **41**, 1281–1290 (2002).

[23] I.M. Sobol, Math. Comput. Simul. **55**, 271–280 (2001).

[24] A. Rahmanian, A. Mohammadi and D. Richon, Fluid. Phase. Equilib. **414**, 80–88 (2016).

[25] W. Chanachichalermwong, A. Charoensaeng and U. Suriyapraphadilok, J. Surfactants. Deterg. **22**, 907–919 (2019). doi:10.1002/jsde.2019.22.issue-4.

[26] A.R.N. Aouichaoui, F. Fan, S.S. Mansouri, J. Abildskov and G. Sin, J. Chem. Inf. Model. **63**, 725–744 (2023). doi:10.1021/acs.jcim.2c01091.

[27] Y. Chung, J. Pfaendtner and E.J. Maginn, ACS Phys. Chem. Au **2**, 570–582 (2022).

[28] S.R. Rieder, M.P. Oliveira, S. Riniker and P.H. Hünenberger, J. Cheminf. **15**, 10–31 (2023). doi:10.1186/s13321-022-00677-6.

[29] B.E. Poling, J.M. Prausnitz and J.P. O'Connell, *The Properties of Gases and Liquids*, 5th ed. (McGraw-Hill, New York, 2001).

[30] S.R. Rieder, M.P. Oliveira, S. Riniker and P.H. Hünenberger, J. Cheminform. **15**, 10–31 (2023). doi:10.1186/s13321-022-00677-6.

[31] P. Waage and C.M. Guldberg, Forhandlinger: Videnskabs–Selskabet I Christinia **5**, 35–437 (1864).

[32] K.M. Klincewicz and R.C. Reid, AIChE J. **30**, 137–142 (1984). doi:10.1002/aic.v30:1.

[33] D. Ambrose, NPL Reports Chemistry 92, National Physical Laboratory, 1978.

[34] Y. Nannoolal, J. Rarey and D. Ramjugernath, Fluid. Phase. Equilib. **252**, 1–27 (2007). doi:10.1016/j.fluid.2006.11.014.

[35] T. Holderbaum and J. Gmehling, Fluid. Phase. Equilib. **70**, 251–265 (1991). doi:10.1016/0378-3812(91)85038-V.

[36] A.R. Katritzky, M. Kuanar, S. Slavov, C.D. Hall, M. Karelson, I. Kahn and D.A. Dobchev, Chem. Rev. **110**, 5714–5789 (2010). doi:10.1021/cr900238d.

[37] J. Thiele, Justus Liebig's Ann. Chem. **306**, 87–142 (1899). doi:10.1002/jlac.v306:1/2.

[38] M.L. Mavrovouniotis, Ind. Eng. Chem. Res. **29**, 1943–1953 (1990). doi:10.1021/ie00105a029.

[39] E. Stefanis, L. Constantinou and C. Panayiotou, Ind. Eng. Chem. Res. **43**, 6253–6262 (2004). doi:10.1021/ie0497184.

[40] E. Stefanis, L. Constantinou and C. Panayiotou, Int. J. Thermophys. **26**, 1370–1388 (2005). doi:10.1007/s10765-005-8092-7.

[41] E. Stefanis and C. Panayiotou, Int. J. Thermophys. **29**, 568–585 (2008). doi:10.1007/s10765-008-0415-z.

[42] L. Constantinou, R. Gani and J.P. O'Connell, Fluid. Phase. Equilib. **103**, 11–22 (1995). doi:10.1016/0378-3812(94)02593-P.

[43] 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.

[44] A. Tihic, N. von Solms, M. Michelsen, G. Kontogeorgis and L. Constantinou, Ind. Eng. Chem. Res. **47**, 5092–5101 (2008). doi:10.1021/ie0710768.

[45] J. Abildskov, L. Constantinou and R. Gani, Fluid. Phase. Equilib. **118**, 1–12 (1996). doi:10.1016/0378-3812(95)02846-3.

[46] Y. Nannoolal, Ph.D. thesis, University of KwaZulu-Natal, Durban, South Africa, 2006. http://hdl.handle.net/10413/3766.

[47] A. Vishnyakov and J.I. Siepmann, Int. J. Thermophys. **30**, 1220–1230 (2009).

[48] A. Chakraborty and W.H. Green, *Group contribution and machine learning approaches to predict abraham solute parameters, solvation free energy, and solvation enthalpy* (2023).

[49] H. Calis, W. Lüke, I. Drescher and A. Schütze, in *Energy Sources for Transportation: Handbook of Fuels* (Wiley, Newark, NJ, 2021), pp. 161–200.

[50] Y. Nannoolal, J. Rarey and D. Ramjugernath, J. Chem. Eng. Data **64**, 4898–4909 (2019).

[51] A. Groniewsky and B. Hégely, Fluid. Phase. Equilib. **577**,113990(2024).

[52] D.W. Green and R.H. Perry, *Perry's Chemical Engineers' Handbook*, 8th ed. (McGraw-Hill Education, New York, 2007).

[53] A. Fredenslund, J. Gmehling and P. Rasmussen, *Vapor-Liquid Equilibria Using UNIFAC: A Group Contribution Method* (Elsevier Scientific Publishing Company, Amsterdam, Oxford, New York, 1977).

[54] J. Gmehling, B. Kolbe, M. Kleiber and J. Rarey, *Chemical Thermodynamics for Process Simulation*, 2nd ed. (Wiley-VCH, Weinheim, 2020).

[55] R. Gani and E.A. Brignole, Fluid. Phase. Equilib. **13**, 331–340 (1983). doi:10.1016/0378-3812(83)80104-6.

[56] J.D. Seader and E.J. Henley, *Separation Process Principles*, 2nd ed. (John Wiley & Sons, Hoboken, NJ, 2006).

[57] E.N. Fuller, P.D. Schettler and J.C. Giddings, Ind. Eng. Chem. **58**, 18–27 (1966). doi:10.1021/ie50677a007.

[58] R. Rajesh, R. Morales-Rodríguez and R. Gani, Ind. Eng. Chem. Res. **61**, 17469–17493 (2022).

[59] W. Chapman, K. Gubbins, G. Jackson and M. Radosz, Fluid. Phase. Equilib. **52**, 31–38 (1989). doi:10.1016/0378-3812(89)80308-5.

[60] J. Gross and G. Sadowski, Ind. Eng. Chem. Res. **40**, 1244–1260 (2001). doi:10.1021/ie0003887.

[61] V. Papaioannou, T. Lafitte, C.A. no, C.S. Adjiman, G. Jackson, E.A. Müller and A. Galindo, J. Chem. Phys. **140**,054107(2014). doi:10.1063/1.4851455.

[62] A. Fredenslund, R.L. Jones and J.M. Prausnitz, AIChE J. **21**, 1086–1099 (1975). doi:10.1002/aic.v21:6.

[63] D.S. Abrams and J.M. Prausnitz, AIChE J. **21**, 116–128 (1975). doi:10.1002/aic.v21:1.

[64] P. Rehner, G. Bauer and J. Gross, Ind. Eng. Chem. Res. **62**, 5347–5357 (2023). doi:10.1021/acs.iecr.2c04561.

[65] N. Hayer, T. Wendel, S. Mandt, H. Hasse and F. Jirasek, Chem. Eng. J. **504**,158667(2025). doi:10.1016/j.cej.2024.158667.

[66] I. Nagata and J. Koyabu, Thermochim. Acta. **48**, 187–211 (1981). doi:10.1016/0040-6031(81)87034-7.

[67] L. Constantinou and V. Vassiliades, in *Chemical Product Design: Towards a Perspective through Case Studies* (Elsevier, Amsterdam, 2007), p. 34.

[68] A.C. Dimian, C.S. Bildea and A.A. Kiss, *Integrated Design and Simulation of Chemical Processes*, Vol. 35. (Elsevier, Amsterdam, 2014).

[69] R. Gani and E.A. Brignole, Comput. Chem. Eng. **26**, 883–888 (2002). doi:10.1016/S0098-1354(01)00795-5.

[70] J. Schilling, M. Lampe, J. Gross and A. Bardow, Chem. Eng. Sci. **159**, 217–230 (2016). doi:10.1016/j.ces.2016.04.048.

[71] L.E.K. Achenie, R. Gani and V. Venkatasubramanian, *Computer Aided Molecular Design: Theory and Practice*, 1st ed. (Elsevier Science B.V., Amsterdam, The Netherlands, 2003).

[72] P.M. Harper, R. Gani, P. Kolar and T. Ishikawa, Fluid. Phase. Equilib. **158-160**, 337–347 (1999). doi:10.1016/S0378-3812(99)00089-8.

[73] F.E. Pereira, E. Keskes, A. Galindo, G. Jackson and C.S. Adjiman, Comput. Chem. Eng. **35**, 474–491 (2011). doi:10.1016/j.compchemeng.2010.06.016.

[74] J. Gmehling, J. Chem. Thermodyn. **41**, 731–747 (2009). doi:10.1016/j.jct.2008.12.007.

[75] D. Minakata, P. Keli Westerhoff and J. Crittenden, Environ. Sci. Technol. **43**, 6220–6227 (2009). doi:10.1021/es900956c.

[76] M. Saeys, M.F. Reyniers, V.V. Speybroeck, M. Waroquier and G.B. Marin, Chem. Phys. Chem. **7**, 188–199 (2006). doi:10.1002/cphc.v7:1.

[77] R. Gani, Comput. Chem. Eng. **28**, 2441–2457 (2004). doi:10.1016/j.compchemeng.2004.08.010.

[78] J. Gmehling, D. Constantinescu and B. Schmid, Annu. Rev. Chem. Biomol. Eng. **6**, 267–92 (2015). doi:10.1146/chembioeng.2015.6.issue-1.

[79] H.S. Wu and S.I. Sandler, Ind. Eng. Chem. Res. **30**, 881–889 (1991). doi:10.1021/ie00053a010.

[80] Y. Nannoolal, J. Rarey, D. Ramjugernath and W. Cordes, Fluid. Phase. Equilib. **226**, 45–63 (2004). doi:10.1016/j.fluid.2004.09.001.

[81] A. Fredenslund, J. Gmehling, M.L. Michelsen, P. Rasmussen and J.M. Prausnitz, Ind. Eng. Chem. Process Design Devel. **16**, 450–462 (1977). doi:10.1021/i260064a004.

[82] T. Zhou, R. Gani and K. Sundmacher, Engineering **7**, 582–604 (2021).

[83] Y. Liu, Z. Hu, Z. Suo, L. Hu, L. Feng, G. Xiuqing, Y. Liu and J. Zhang, Sci. China Technolog. Sci. **62**, 521–545 (2019).

[84] T. Suzuki, J. Chem. Inf. Comput. Sci. **41**, 1266–1273 (2001). doi:10.1021/ci010295f.

[85] M. Frenkel, R. Chirico, V. Diky, X. Yan, Q. Dong and C. Muzny, J. Chem. Inf. Model. **45**, 816–38 (2005). doi:10.1021/ci050067b.

[86] M. Frenkel, R.D. Chirico, V. Diky, Q. Dong, S. Frenkel, P.A. Franchois and D.L. Embry, J. Chem. Inf. Model. **51**, 1506–1512 (2011). doi:10.1021/ci200096q.

[87] V. Diky, *Presented at the 8th International Research and Practical Conference "Chemical Technology: Science, Econonmy And Production"* (2024).

[88] J. Yi, Q. Lei, W.M. Gifford, J. Liu, J. Yan and B. Zhou, in *Proceedings of the 2019 SIAM International Conference on Data Mining (SDM)* (SIAM, 2019), pp. 324–332.

[89] V. Mann, R. Gani and V. Venkatasubramanian, Fluid. Phase. Equilib. **568**,113734(2023). doi:10.1016/j.fluid.2023.113734.

[90] NIST, Tech. Rep. 590, National Institute of Standards and Technology, 1984.

[91] Y. Chung, F.H. Vermeire, H. Wu, P.J. Walker, M.H. Abraham and W.H. Green, J. Chem. Inf. Model.**62**, 633–644 (2022). doi:10.1021/acs.jcim.1c01103.

[92] Y. He, Y. Feng, L. Qiu and D. Tang, Energy **306**,132490 (2024). doi:10.1016/j.energy.2024.132490.

[93] J. Xue, X. Feng, Q. Jia, Q. Wang and F. Yan, AIChE J. **71**, 18823–18826 (2025).

[94] X. Hou, L. Yu, C. He and K. Wu, AIChE J. **68**, 17408–17421 (2021).

[95] P. Simamora and S.H. Yalkowsky, Ind. Eng. Chem. Res. **33**, 1405–1409 (1994). doi:10.1021/ie00029a042.

[96] G.M. Wilson and L.V. Jasperson, in *AIChE Annual Meeting* (New Orleans, LA, 1996), p. 21.

[97] T. Lan, Y. Wang, R. Ali, H. Liu, X. Liu and M. He, Fuel Proces. Technol. **251**,108918(2024).

[98] Y. Liu, Y. Wang, Z. Wang, X. Wang and Y. Duan, J. Chem. Eng. Chin. Univ. **36**, 985–993 (2022). doi:10.1007/s40242-020-0289-6.

[99] Q. Li, J. Ren, Y. Liu and Y. Zhou, Int. J. Refrig. **143**, 28–36 (2022). doi:10.1016/j.ijrefrig.2022.06.024.

[100] R. Gardas, Ind. Eng. Chem. Res. **47**, 5751–5757 (2008). doi:10.1021/ie800330v.

[101] T.D. Pietro, L. Cesari and F. Mutelet, Fluid. Phase. Equilib. **572**,113854(2023). doi:10.1016/j.fluid.2023.113854.

[102] J. Han, M. Li, N. Tian, C. Liu, Y. Zhang, Z. Ji and X. Sun, Fluid. Phase. Equilib. **589**,113729(2023).

[103] R. Haghbakhsh, S. Raeissi and A. Duarte, Sci. Rep. **11**, 6294 –6312 (2021). doi:10.1038/s41598-021-85824-z.

[104] A. Ahmadi, R. Haghbakhsh, S. Raeissi and V. Hemmati, Fluid. Phase. Equilib. **401**, 64–74 (2015). doi:10.1016/j.fluid.2015.05.021.

[105] J. Albert and K. Müller, J. Korean Ind. Eng. Chem. **53**, 17522–17526 (2014).

[106] V. Villazón-León, A. Bonilla-Petriciolet, J. Tapia and G. Luna-Barcenas, Fluid. Phase. Equilib. **576**,113938(2023). doi:10.1016/j.fluid.2023.113938.

[107] Z. Liu, L. Shang, K. Huang, Z. Yue, A. Han, D. Wang and H. Zhang, Environ. Sci. Technol. **59**, 6830–6840 (2024).