Extension of the Constantinou and Gani group contribution method with the Tochigi method through an automatic conversion procedure

Axel Groniewsky [a,*], László Hégely [b]

[a] Department of Energy Engineering, Faculty of Mechanical Engineering, Budapest University of Technology and Economics, Műegyetem rkp. 3 H-1111 Budapest, Hungary
[b] Department of Building Services and Process Engineering, Faculty of Mechanical Engineering, Budapest University of Technology and Economics, Műegyetem rkp. 3 H-1111 Budapest, Hungary

ARTICLE INFO

ABSTRACT

Group contribution methods (GCMs) are widely employed across various disciplines to estimate compound properties when experimental data is lacking in the literature. While several methods exist, none are comprehensive, as they exhibit gaps either in functional groups or the predicted properties or simply do not offer the required accuracy. As a result, different methods may be necessary to evaluate distinct properties of the same compound. Typically, switching between these models is performed manually due to variations in the sets of functional groups employed. However, with the advancements in computational power and numerical optimum search, there is a growing need to automate the conversion between different group contribution methods. This study presents a procedure to extend the property estimation of the Constantinou and Gani method with vapor pressure by supplementing it with the Tochigi method using an automated group conversion algorithm. The difficulties of automatic conversion procedures, resulting from the differences in the group sets and the shortcomings of the GCMs, are also highlighted. It is also demonstrated that the accuracy of the acentric factor estimation can only be refined to a limited extent by incorporating the Tochigi method, which is, however, indispensable for the several groups where the Constantinou and Gani group contribution values are missing.

1. Introduction

Group contribution methods (GCMs) are useful in estimating compound properties, particularly when the literature has limited experimental data. These methods work based on the fundamental assumption that a compound's properties are derived from the attributes of its functional groups. By evaluating the magnitude of a group's influence and its prevalence within the compound, GCMs can provide approximate values for various properties exhibited by the compound [1,2].

Utilizing the methodology alongside an optimum search process can be highly advantageous, especially in computer-aided molecular design (CAMD), when the objective is not merely to choose a compound from a predetermined list but to design a novel compound. However, it is essential to note that material selection typically involves more than one property to consider. More often than not, it entails multi-criteria optimum search challenges, encompassing not just thermodynamic and transport properties but also environmental and various other factors.

Song and Song [3] combined a simulated annealing algorithm with group contribution methods to design solvents with desirable physicochemical and environmental properties using 26 pre-selected functional groups to investigate feasible molecular structures. Lazzus [4] estimated the liquid density of imidazolium-based ionic liquids and predicted the flash point temperature of organic compounds [5] using a combined method that included a multilayer neural network with particle swarm optimization and a simple group contribution method. Zhao and Deng [6] simultaneously optimized the parameters of the working fluid and the cycle to extract maximum work from an organic Rankine cycle (ORC) using five simple functional groups. White et al. [7] coupled CAMD of the organic Rankine cycle working fluid based on the statistical associating fluid theory (SAFT)-γ Mie equation of state with thermodynamic modeling and optimization. As SAFT-γ Mie is only suitable for determining thermodynamic properties, dynamic viscosity, thermal conductivity, and surface tension were predicted with a group-contribution method employing hydrocarbon-type functional groups only. Wende et al. [8] applied multi-criteria screening of organic ethanolamines for efficient CO_2 capture based on 31 amine absorbents.

Wen Su et al. [9] developed a GCM based artificial neural network (GCM-ANN) to predict the ORC performance from key properties of working fluids. However, the pool of working fluids was limited to 54 compounds and pre-determined before the ANN was trained.

Chemmangattuvalappil et al. [10] applied CAMD of optimal sustainable solvent for liquid-liquid extraction, using only seven potential building blocks. Bardow et al. [11] presented a framework for integrating the design of fluids and processes and demonstrated the applicability by designing the working fluids for organic Rankine cycles and solvents for CO_2 capture. The fluid was designed based on a group-contribution method for the PC-SAFT parameters (GPC-SAFT). They also used a one-stage approach of the GPC-SAFT method in a case study [12] for a subcritical ORC process. The approach efficiently identified the optimal working fluid and the corresponding optimal process parameters employing continuous process variables and integer variables representing the molecular structure of the working fluid. Likewise, the CAMD method based on PC-SAFT for working fluid design of a high-temperature organic Rankine cycle was applied by Wang et al. [13]. Wang et al. [14] improved the cycle performance of a CO_2 transcritical power cycle employing CO_2-based mixtures instead of pure CO_2 and used GPC-SAFT to avoid the empirical screening of CO_2 mixtures.

Gebreslassie and Diwekar [15] used efficient ant colony optimization (EACO) for solvent selection where the solvent-solute distribution coefficient is maximized subject to structural feasibility, property, and process constraints, using computer-aided molecular design. Shahmohammadi et al. [16] applied CAMD for the optimal design of novel precursor materials for atomic layer deposition.

The comprehensive literature review, while not all-encompassing, leads to the conclusion that the subject matter garners significant interest and finds application across a broad spectrum of areas. It is important to acknowledge the variety of common methodologies employed; however, the diversity of these approaches precludes the identification of any one method as definitively superior. This variability underscores the reality that no single model is flawless in all aspects, suggesting the potential necessity for developing automated algorithms to facilitate seamless transitions between varying methodologies.

When GCM is used for material design, the solution within the multidimensional search space is usually represented by a vector containing the number of functional groups within the applied GCM. Consequently, employing various GCMs in conjunction is uncommon because most methods have different functional groups, and the automated conversion between different methods can be challenging. However, using only one GCM can be restrictive or insufficient in the following cases:

- For one or more properties estimated by the GCM, contribution values of certain groups are missing, restricting the molecular search space. Another GCM can be applied to calculate these properties either for every molecule or only for those containing such functional groups.
- Properties estimated only by another GCM are also needed. These properties might be required as input values of the first GCM (e.g. normal boiling point as input for the calculation of vapor pressure), although this is less common.

In order to apply CAMD in such cases, an automatic conversion method between the functional groups is required. Indeed, one of the challenges identified by the review of Mann et al. [17] is whether a conversion method between representation systems (e.g. different sets of functional groups) can be developed.

The Constantinou and Gani model [18], favored for its straightforward approach, versatility, reasonable accuracy, and adaptability, stands out among GCMs. Regardless, as it cannot predict the vapor pressure of pure materials, its applicability is severely limited in mechanical, chemical, and process engineering (i.e., distillation, refrigeration, and process flow analysis). Despite the method's ability to estimate ten properties along with the acentric factor, its efficacy is hampered by the limited number of functional groups with available contribution values for its calculation. While 67 first- and 48 second-order functional groups can be used to estimate the critical pressure of a compound, only 47 first- and 27 second-order groups are applicable to predict acentric factors. Moreover, due to the missing contributions of certain groups like CF_2, the model falls short in estimating the acentric factor for several refrigerants such as R218, R245ca, R365mfc, RE245cb2 and RE347mcc.

The number of GCMs for the prediction of vapor pressure is rather limited. Since vapor pressure is a function of temperature, such methods must provide temperature-dependent group contributions. These GCMs differ from each other in the functional groups used, the range of vapor pressures for which they were fitted and whether they require critical and/or boiling point temperatures for the estimation. The differences in the sets of functional groups mean that the methods may not be applicable for the same molecules, which makes their comparison difficult [19].

A number of methods require the knowledge of the critical and/or boiling point temperatures of the molecules. In a CAMD framework, these parameters must be estimated by another GCM, which raises the need for the automatic conversion between their respective functional groups. Li et al. [20] proposed a method that can describe halogen- (except I) and sulfur-containing molecules, as well. The contributions of the 117 groups are only used, however, to estimate the critical temperature and pressure. The knowledge of the boiling point, not predicted by this method, is also required. Other model parameters do not depend on the presence of the functional groups, only on the family (acid, alcohol, phenol, other) of the compound. The method of Myrdal and Yalkowsky [21] needs the boiling and the melting temperature apart from structural information, which are the number of nonring, nonterminal sp^3 and sp^2 atoms, the number of rings, OH, COOH and NH_2 groups and the rotational symmetry number (σ) of the molecule. The latter parameter is the number of indistinguishable orientations of the molecule. The determination of σ requires the knowledge of the complete molecular structure, not only the number of the different groups, as it is common in other GCMs. Thus, an automatic conversion would require the generation and evaluation of all possible molecules for a set of groups selected. The method of Nannoolal et al. [22] and its extension by Moller et al. [23] require the boiling point. They use a large number of first-order functional groups and also take the interaction between them into account by incorporating second-order groups and group interaction terms. In the case of Tochigi et al. [24], providing the boiling point is optional but increases the estimation accuracy. This method uses different contribution values according to the class of molecule but does not include second-order groups. It is also possible to predict vapor pressures using the UNIFAC method but, apart from the normal boiling point, this also requires virial coefficients [25,26].

Other models only require structural information. Macknick and Prausnitz [27] proposed a GCM only for hydrocarbons, which was later extended by Edwards and Prausnitz [28] for heavy compounds containing nitrogen or sulfur. The method of Tu [29] can be applied more generally but does not describe molecules containing F, Br or I. Moreover, in the case of benzene derivatives, the alkyl substituents and their relative position on the aromatic ring must be known. Coutsikos et al. [30] proposed a GCM for the vapor pressure of solid organic compounds. It can be extrapolated to liquid state provided that the temperature is not very far from the range of the fitting data. Otherwise, by approaching the boiling point, errors can reach 100 %. The method of Capouet and Müller [31], developed for complex organic molecules, requires the vapor pressure of the hydrocarbon backbone, that is, in the case of CAMD, another GCM. The SIMPOL.1 method [32] used a total of 30 functional groups, including second-order ones, but it lacks halogen- or sulfur-containing groups. The method EVAPORATION [33], developed for complex, low-volatility compounds, does not include these groups either. Wang et al. [34] proposed a method using only five functional

groups to estimate the vapor pressure of fatty acid esters. It must also be noted that the equation of Lee-Kesler (not a GCM) can also predict vapor pressures given the values of critical temperature and pressure and the acentric factor [2]. However, the availability of the acentric factor or its estimation by another GCM can be an issue.

By comparing GCM predictions to experimental data, O'Meara et al. [35] stated that EVAPORATION performed best where it was applicable; otherwise, Nannoolal et al. [22] had the lowest mean bias error. According to Compernolle et al. [33], the method of Nannoolal et al. [22] can sometimes largely overestimate the vapor pressure, while that of Myrdal and Yalkowsky [21] and SIMPOL.1 can suffer from underestimation. They found their own method (EVAPORATION) to be the most accurate, which is not surprising given that it was fitted to the experimental data used for comparison.

In the present work, the method of Tochigi et al. [24] is chosen because it can estimate the vapor pressure of a wide range of compounds (including halogen- and sulfur-containing ones), while it does not use higher-order groups or require the exact knowledge of the molecular structure, which would make an automatic conversion procedure considerably more complicated. Moreover, it uses a simpler set of groups than Li et al. [20].

The primary objective of this study is to add vapor pressure prediction to the Constantinou and Gani method [18] by incorporating the Tochigi method [24]. The two methods are connected through an automatic conversion procedure, which converts Constantinou and Gani groups to the ones used in the Tochigi method, enabling their application in the search for numerical optima. Since the acentric factor is determined from the saturated vapor pressure at $0.7 \cdot T_{crit}$, this method can also be used to refine the prediction of the acentric factor. Nevertheless, it is also demonstrated that the corrections of the shortcomings of the O'Connell-extended Constantinou and Gani method [36] are minor; the accuracy of the acentric factor can only be refined to a limited extent by incorporating the Tochigi method. However, applying the Tochigi method for the several (31 first- and 16 second-order) functional groups lacking contribution values for the acentric factor in the O'Connell-extended Constantinou and Gani method can be recommended.

2. Methodology

First, the group contribution methods of Constantinou and Gani [17] (including its extension by Constantinou et al. [19]) and Tochigi et al. [18] are presented briefly. Subsequently, the automatic conversion algorithm is described with a discussion of the shortcomings of the two GCMs leading to difficulties in their automated use.

2.1. Constantinou and Gani method

The well-known method of Constantinou and Gani [17] was chosen because it has an extensive set of functional groups, can partially distinguish isomers by using second-order groups, and it can be used to predict a considerable number of properties: the critical properties (temperature, pressure, molar volume), the normal boiling point, the melting point, the enthalpy of formation, the Gibbs energy of formation and the standard enthalpy of vaporization at 298 K. Constantinou et al. [19] added two more predictable properties, the acentric factor and the liquid molar volume at 298 K, using the same set of functional groups.

In contrast to the classical method of Joback and Reid [1], Constantinou and Gani [17] used two sets of groups. A molecule is fragmented into a number of non-overlapping first-order groups that completely cover it. In this sense, all the groups of Joback and Reid [1] are first-order ones. Second-order groups are built from first-order ones; they can partially overlap and do not have to cover the entire molecule, in fact, a molecule does not have to contain second-order groups. The aim of using second-order groups is to improve property prediction by partially describing proximity effects of first-order groups and

isomerism. Further improvement can be reached by introducing third-order groups, as in the GCM of Marrero and Gani [37], but at the expense of even higher complexity.

In the Constantinou and Gani method, 78 first- and 43 second-order groups are used, given in Tables A1 and A2 of the Supplementary material with an identification number. First-order groups will be referred to here mostly by name, but occasionally the identification number will also be used to avoid long lists of text. Second-order groups will always be referred to by number. In this work, for each first-order group with a valency higher than one, l_{min} and l_{max} length values were defined, which, if the group is in a ring, express the minimum and maximum number of atoms that are actually part of the ring. For many groups, $l_{min}=l_{max}=l$, e.g. for CH_2 $l=1$, that is, this group always increases the ring length (membership) by one. If $l_{min} \neq l_{max}$, an l value can be freely selected between the bounds. For example, if in a ring, only one carbon atom of the group C=C is included, $l=1$, while if the whole group in the ring, $l=2$. l can only take more than two values by the heterocyclic groups C_5H_3N and C_4H_2S.

Although the set of first-order groups consists of 78 elements, and that of the second-order groups 43, it shall be noted that the actively available number of functional groups of the original [18] and the extended Constantinou and Gani method [36], both using the same set of groups, varies depending on the predicted variable. While 75+41 functional groups can be used to estimate the normal boiling point of a compound, only 51+26 functional groups are available for the acentric factor that often hinders the prediction of the compound under investigation. Hence, incorporating Tochigi [24] into the Constantinou and Gani method can be a valuable tool for estimating the correlation between vapor pressure and temperature in pure substances but also for predicting the acentric factor in cases where functional groups are missing. The elements of both sets ([18,36] and [24]) are found in the Supplementary material.

2.2. Tochigi method

Tochigi et al. [24] proposed a method for predicting Antoine constants using a group contribution method to determine vapor pressures. They also describe GCMs to calculate the critical properties, which can also be used if the critical parameters are lacking. The 1817 compounds treated in the work were grouped into six classes:

Class 1: saturated, unsaturated and aromatic hydrocarbons;
Class 2: alcohols, ethers, sulfur-containing compounds;
Class 3: aldehydes, ketones, acids, esters;
Class 4: amines, nitriles, nitro compounds;
Class 5: complex group compounds;
Class 6: halogen-containing compounds.

Depending on the class of the compound, the same functional groups (given in Table A3 of the Supplementary material) can have different contribution values. These groups are also given here an identification number, which will only be used to list the groups of the molecules taken from the RefProp database.

Tochigi et al. determined the vapor pressure using the semi-empirical Antoine equation:

$$log_{10} P_{sat@mm\ Hg} = AntA - \frac{AntB}{T + AntC} \qquad (1)$$

The same equation is used in the Vapor-Liquid Equilibrium Data Collection published as part of the DECHEMA Chemistry Data Series [38]. Vapor pressure P_{sat} is calculated in mmHg while temperature T in °C. AntA, AntB, and AntC are the Antoine constants.

Tochigi et al. [24] used two approaches to estimate the vapor pressure. In the first one, all the three Antoine constants are calculated from group contributions, while in the second one, AntA is determined from the measured normal boiling point of the compounds and the AntB and
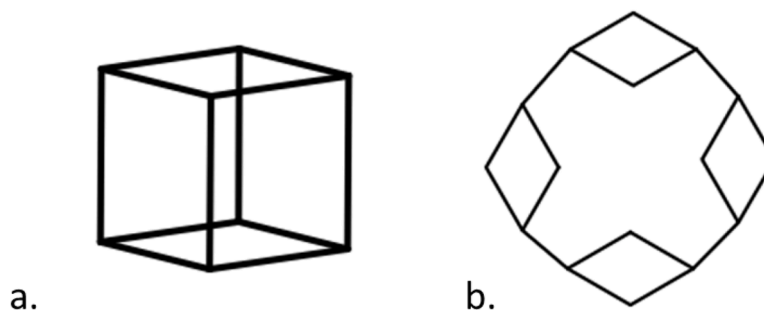
Fig. 1. Example molecular structures for different ring definitions: a. cubane, b. pentacyclo[9.1.1.1^{2,4}.1^{5,7}.1^{8,10}]hexadecane.

AntC values.

2.3. Acentric factor

Since Tochigi et al. [18] applied experimental normal boiling points to improve the prediction of vapor pressure, the reported results might be reliable foundation for acentric factor calculation. The acentric factor (ω) was introduced by Kenneth Pitzer [39], providing valuable insights into fluid behavior by measuring the non-sphericity (acentricity) of the molecules [40]. It has since become a standard for characterizing the phase of individual pure components alongside parameters like molecular weight, critical parameters, or critical compressibility. The acentric factor is defined as:

$$\omega = -1.00 - log_{10}(P_{sat@T} / P_{crit}) \tag{2}$$

where P_{crit} is the critical pressure and P_{sat} is the vapor pressure at temperature T where $T/T_{crit}=0.7$ and T_{crit} is the critical temperature in K.

2.4. Automatic conversion procedure

First, a discussion of the shortcomings of the GCMs studied concerning their automated use is given, which lead to several difficulties when designing an automatic conversion procedure. Subsequently, the procedure is described in several subsections. Some details of the procedure are described in the Supplementary material.

2.4.1. Shortcomings of the Constantinou and Gani and the Tochigi methods and the difficulties of automatic conversion

The most important differences between the sets of functional groups of the Constantinou and Gani and the Tochigi methods are:

- The Tochigi method assigns different contribution values and partially different functional groups depending on the class of the compound. However, this can be easily treated by the automatic conversion.
- Constantinou and Gani (CG) groups connect to others only by single bonds: double and triple bonds are present only inside groups. (An apparent exception is the set of aromatic groups, but they must connect to each other.) Tochigi (T) groups can connect to others by single and multiple bonds.
- In contrast to the Tochigi method, Constatinou and Gani do not differentiate between groups in open chains and non-aromatic rings.
- Constantinou and Gani build up aromatic rings from groups consisting of a single aromatic carbon and potentially another molecular fragment, which might or might not have another free valency (e.g., the groups ACCH$_2$ and ACCH). Each aromatic group must be connected to exactly two other aromatic groups. This approach makes it, in theory, possible to describe any type of non-heterocyclic aromatic ring. The method also contains some heteroaromatic rings as separate groups (C$_5$H$_4$N, C$_5$H$_3$N: pyridine; C$_4$H$_3$S, C$_4$H$_2$S: thiophene), but

in lack of other information, it is assumed that other heteroaromatic rings cannot be described. In contrast, the Tochigi method applies seven aromatic groups, each consisting of one (Ph: phenyl, o-Ph, m-Ph, p-Ph: phenylenes, Bz: benzoyl) or two (Naf1: 1-naphthyl, Naf2: 2-naphthyl) non-heterocyclic aromatic rings. Heteroaromatic rings can be constructed from normal (cyclic) groups.

Before discussing the difficulties of automatic conversion, the shortcomings of the GCMs with respect to their automated use should be addressed. The main weakness of the Constantinou and Gani method is their use of 3, 4, 5, 6 and 7 membered rings as second-order groups without defining what they mean by ring. At first glance, identifying the number and size of the rings in a molecule appears to be a trivial and intuitive task, which explains the obliviousness of the authors. However, in the case of more complex molecules, the number of rings depends on their definition. The fundamental problem is that a universal definition of "chemically relevant" rings does not exist since the relevance is context-dependent.

A number of papers are dedicated to the question of (automatic) ring perception in molecules, represented by graphs. Particularly insightful is the work of Berger et al. [41], who reviewed different ring definitions, their advantages and disadvantages, as well as the failures and errors of different ring perception algorithms. Calculating all the cycles in a graph is straightforward, but most of these cycles would not be relevant since they can be constructed from smaller cycles. In fact, the cycles of a graph form a vector space with the symmetrical difference of the cycles as vector addition. Therefore, any other cycle can be obtained as the symmetrical difference of two or more elements of the cycle basis. There are almost always multiple possible cycle bases. One of the most popular concepts is the smallest set of smallest rings (SSSR) by Plotkin [42], which corresponds to a minimum cycle basis, that is, a cycle basis where the sum of the sizes of its elements is minimal. The dimension of the cycle basis, thus, the number of elements in SSSR is given by the cyclomatic number μ (also called circuit rank, first Betti number, nullity, or Frèrejacque number) and can be easily calculated:

$$\mu = \frac{\sum_{i=1}^{n} d_i}{2} - n + 1 \tag{3}$$

where d_i is the degree of vertex (first-order group) i in the molecular graph and n is the number of vertices. The first term is the number of edges (bonds) in the graph. Note that knowing the actual molecular structure is not necessary to calculate μ, only the number and degree of the first-order groups. The problems with SSSR are that it is non-unique and does not always contain all the rings that could be considered chemically relevant. Take the classical example of the hydrocarbon cubane (Fig. 1a). The value of μ is 5, and five four-membered rings indeed contain all the vertices and edges, but the molecule consists of 6 indistinguishable rings. Six possible SSSRs exist, each leaving out one of the rings. In this case, the SSSRs do not differ from each other, but if one of the CH groups is changed to another type of group, different sets are obtained.

Table 1
Minimal ring systems up to $\mu=3$.

μ	b_4	b_3	n	Structure	Molecule name
1	0	0	3		cyclopropane
2	0	2	4		bicyclobutane
	1	0	5		spiropentane
3	0	4	4		tetrahedrane
	1	2	5		tricyclo[2.1.0.0^{1,3}]pentane
	2	0	5		[1.1.1]propellane

Many other ring sets can be defined instead of the SSSR (Berger et al., [41]), but only a few are both unique and contain a minimum cycle basis. One example is the set of relevant cycles. A cycle in a graph is relevant if it cannot be obtained as the symmetrical difference of other cycles. However, the set of relevant cycles (and other similar sets) can contain an excessive number of cycles (rings). In the example of pentacyclo[9.1.1.1^{2,4}.1^{5,7}.1^{8,10}]hexadecane (Fig. 1b), relevant cycles include, apart from the four-membered rings, all the 16 possible twelve-membered cycles, as well (Goetzke and Klein, [43]). Nevertheless, from a chemical point of view, only the four-membered rings and any single one of the twelve-membered rings should be taken into account. The concept of unique ring families (URF), proposed by Kolodzik et al. [44], promises to combine the essential characteristics of ring definitions: uniqueness, being chemically relevant, and efficient to compute. Therefore, interpreting the rings of the Constantinou and Gani method as URFs would be a promising approach, independently of the original intent of the authors. However, a fundamental problem here is that the number of URFs (n_{URFs}) can only be determined by detecting them in the molecular graph. Thus, knowing only the number and type of the groups does not offer enough information, and only a lower and upper bound can be computed:

$$\mu \leq n_{URFs} \leq \frac{\sum_{i=1}^{n} b_i}{2}\left(\sum_{i=1}^{n} b_i + n\right) \qquad (4)$$

where b_i is the number of vertices with degree i.

The solution proposed here is to use μ as the number of rings but to exclude molecules containing ring systems where the SSSR might not cover all chemically relevant rings. A ring system means a single ring or multiple rings sharing at least one common atom (Patterson, [45]). Table 1 shows the minimal ring systems up to $\mu=3$. In this paper, a minimal ring system is defined as a ring system that can be realized with a minimum number of groups (vertices) for given values of μ, b_4, and b_3. The names of the molecules consisting only of the given minimal ring systems are also given.

Even at $\mu=3$, tetrahedrane presents the same problems as cubane: only three out of the four of the intuitively perceived rings are predicted by the value of μ. Therefore, the decision was taken to limit the number of rings in ring systems to two, allowing only monocyclic, bicyclic and spirosystems. Note that the number of rings in the molecule is not restricted.

By the Constantinou-Gani method, it is also not completely clear whether a 3-membered ring refers to a ring consisting of three groups (each of which might participate with more than one atom in the ring) or three atoms. Here, the latter interpretation is used even if it makes the conversion more complicated since it is more reasonable from a chemical point of view.

The main problem with the method of the Tochigi is the complete lack of explanation of the groups used. Many of them are straightforward, but in some cases, multiple interpretations would be possible. For example, the group O(Hal) could be interpreted as an ether group (O) connected either to at least one halogen-containing group (but not a halogen atom) or to a halogen atom. The second case, resulting in acyl halides, seems less likely, but Tochigi et al. [18] offer no guidance. Given the existence of the HalOH group, which certainly does not correspond to the hypohalous acids, it is assumed that the first interpretation of O(Hal) (and similar groups with the (Hal) suffix) is correct. The lack of explanations could also lead to the belief that the method cannot describe heteroaromatic rings, but one of the molecules for which the model's accuracy is presented is 2-methylpyridine.

Another issue is that more than one Tochigi group could be chosen for the same molecular fragment. For example, in the molecule 2-chloro-1-benzofuran-3-one (Fig. 2a), the O atom could either be O(Hal-Cyc) since it is halogen-adjacent and inside a ring or O(Ph) since it is connected to a benzene ring. Similarly, the ketone group could either be C=O(Ph) or C=O(Hal). Tochigi et al. [18] offer no guidance here, likely because the authors did not envisage such situations.

In other cases, a molecule can be constructed from multiple sets of groups. Acetic benzoic anhydride (Fig. 2b) can be constructed as CH$_3$+COO+Bz (benzoyl) or as CH$_3$+COOCO+Ph (phenyl). Both dichlorofluoromethane (R21, CHCl$_2$F) and chlorodifluoromethane (R22, CHClF$_2$) have three possible constructions depending on whether H, Cl, or F is considered as a group apart from the rest of the molecule. In the ambiguous situations highlighted above, one must make arbitrary choices due to the lack of guidance from the authors.

An ideal conversion procedure between the two methods would be a total but not necessarily injective nor surjective function between the two sets of functional groups, mapping each group in the domain (here, the set of Constantinou-Gani groups *CG*) to a group in the codomain (the
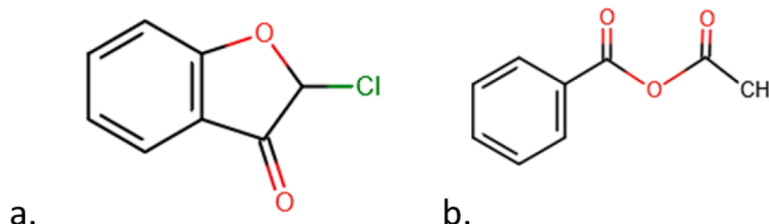
Fig. 2. Example molecules to illustrate the ambiguity of the Tochigi method: a. 2-chloro-1-benzofuran-3-one, b. acetic benzoic anhydride.
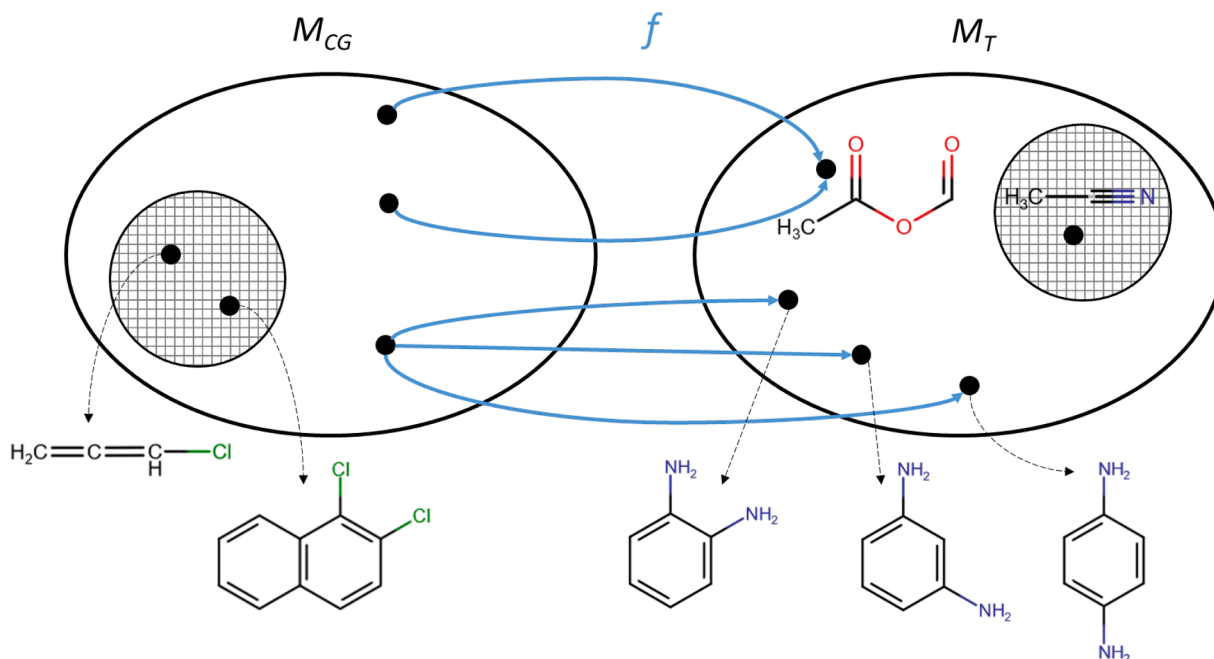
Fig. 3. Illustration of the conversion function $f: M_{CG} \rightharpoonup M_T$ with example molecules. The shaded areas represent the set of elements not included in the function's domain and codomain, respectively.

set of Tochigi groups T_i for the given i class of the molecule), but permitting groups in the codomain to be the image of two or no groups in the domain. However, such a function does not exist in this case for several reasons. Most importantly, the existence (or lack) of second-order groups and the relative position of the first-order groups (e.g., being in a cycle, halogen-adjacent) can change the image of the group. In the case of phenylene groups, Tochigi differentiates between the orto, meta, and para isomers, but such distinction cannot be made in the Constantinou-Gani method. Therefore, a unique function is not possible. Moreover, not every Constantinou-Gani group has an image: the group $CH_2{=}C{=}CH$ is mapped to $CH_2{=}$, $=C=$ and $CH{=}$, but the group $=C=$ exists only in class 1. Additionally, several groups of the same type must occasionally be converted into a single image group; for example, five ACH and one AC are converted to a single Ph group.

The conversion procedure can be thus rather considered to a be a partial, multi-valued, non-injective and non-surjective function between the sets of possible molecules built from first- and second-order Constantinou-Gani groups (M_{CG}) and the sets of possible molecules built from Tochigi groups (M_T): $f: M_{CG} \rightharpoonup M_T$ (Fig. 3). Each element of both sets are sets themselves, with the functional groups in the molecule as elements. f is multi-valued because the image is defined as the set of possible group sets that could correspond to the Constantinou-Gani groups. To take the example of diaminobenzene (Fig. 3), which has three isomers, the image is:

$$f(\{ACH, ACH, ACH, ACH, ACNH_2, ACNH_2\})$$
$$= \{\{o-Ph, NH_2, NH_2\}, \{m-Ph, NH_2, NH_2\}, \{p-Ph, NH_2, NH_2\}\} \quad (5)$$

In such cases, the user can freely select one of the elements of the image.

The function is partial since there are molecules that can be described by Constantinou and Gani but not by Tochigi, e.g. chloroallene or dichloronaphthalenes (Fig. 3). It is non-injective: some elements of M_T are not images of any element of M_{CG}, such as acetonitrile (CH_3CN), that cannot be described with Constantinou-Gani groups. It is also non-surjective: some elements in M_T are the images of multiple elements in M_{CG}, for example, the acetic formic anhydride:

$$f(\{CH_3COO, CHO\}) = \{H(C{=}O), COOCO, CH_3\} \quad (6)$$

$$f(\{HCOO, CH_3CO\}) = \{H(C{=}O), COOCO, CH_3\} \quad (7)$$

An automated conversion procedure, which approximates the true partial function f, is presented below. The need for an approximation arises as it would be virtually impossible to determine a set of exact but completely general conversion rules due to the very large number of possible interactions between groups without generating all the possible molecular graphs for a given element in M_{CG}. Nevertheless, the most important interactions are taken into account in the procedure proposed.

2.4.2. Overview of the conversion procedure and treatment of aromatic rings

Fig. 4 shows a flowchart of the algorithm. First, the Tochigi class of the molecule must be determined based on the groups present (Table 2, Rules C1-C8). Rule C4 is specific for acetophenone, where the corresponding Tochigi groups are Bz and CH_3. Once the class is known, the value of μ is calculated. If $\mu{=}0$, the molecule contains no rings, and the rules for acyclic molecules are used. If at least one ring is present, the number of aromatic Constantinou-Gani groups n_{Ar} is determined. Note that from now on, the paper refers only to non-heteroaromatic rings by the term aromatic since Constantinuou and Gani treats the very few heteroaromatic rings they allow as separate groups. Heteroaromatic groups are also not included in μ, as it is unnecessary for the conversion. If $n_{Ar}{=}0$, the molecule is cyclic but not aromatic and is treated accordingly. If the molecule is aromatic, it is checked whether it is benzene or naphthalene since these compounds are treated separately by Rules A1 and A2 (Table 3). Using Naf1 instead of Naf2 gives a slightly better agreement with vapor pressure values calculated with measured Antoine constants.

For other aromatic molecules, the number of aromatic rings must be determined. Aromatic rings are divided into two categories: main and condensed rings. Phenyl and phenylene groups consist only of a main ring, while naphthyl groups consist of a main and a condensed ring. Their numbers $n_{main} \geq 1$ and $n_{cond} \geq 0$ must satisfy the following relationships:
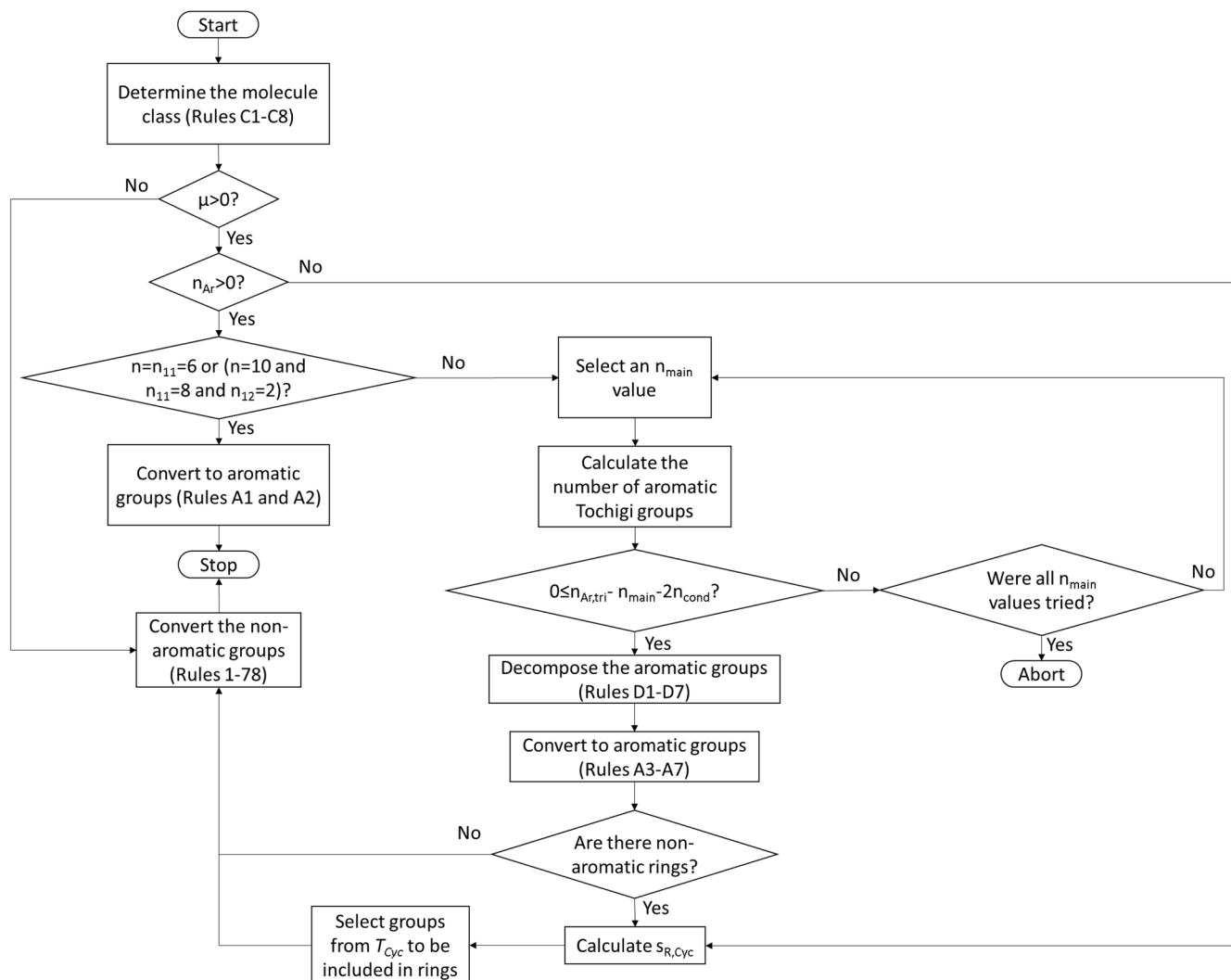
$$6n_{main} + 4n_{cond} = n_{Ar} \quad (8)$$

Fig. 4. Flowchart of the automatic conversion algorithm.

Table 2
Rules for determining the Tochigi class of molecules.

Rule number	Rule description	Class	Remark
C1	Contains only groups 1–15, 53 and 54	1	
C2	Contains only groups 1–17, 24–26, 50, 53–54 and 72–78 and is not class 1	2	
C3	Contains only groups 1–17,19, 24–26, 50, 53–54 and 72–78, must contain from the groups 11–15 and 17; is not class 1		
C4	$CH_3CO+5*ACH+AC$		acetophenone
C5	Contains only groups 1–15, 18–23, 39, 53–54 and 61 and is not class 1	3	
C6	Contains only groups 1–15, 28–38, 47–49 and 53–54 and is not class 1	4	
C7	Contains only groups 1–26, 28–39, 47–49, 53–54, 57, 61 and 66–73 and is not class 1–4	5	
C8	Contains at least one from groups 27, 40–46, 51–52, 55–56, 58–60 and 62–65 and does not contain groups 50 and 74–76	6	

$$n_{main} + n_{cond} \leq \mu \tag{9}$$

$$n_{cond} \leq n_{main} \tag{10}$$

$$2n_{cond} \leq n_{AC} \tag{11}$$

$$3n_{cond} \leq n_{Ar,tri} \tag{12}$$

where $n_{Ar,tri}$ is the number of trivalent aromatic groups:

$$n_{Ar,tri} = n_{AC} + n_{ACCH_2} + n_{ACCH} \tag{13}$$

Inequalities (11) and (12) express that each naphthyl group contains

at least two AC groups and another trivalent group, which connects it to the rest of the molecule.

The above system (8)-(12) may have multiple solutions, in which case an n_{main} value can be freely chosen. After this, the numbers of aromatic Tochigi groups are determined. These groups are Ph (phenyl), o-Ph, m-Ph, p-Ph (phenylenes), Naf1, Naf2 (naphthyls), PhOH (hydroxyphenyl), and Bz (benzoyl). The number of the different naphthyl groups can be freely chosen, while their total number is determined by n_{cond}:

$$n_{Naf1} + n_{Naf2} = n_{cond} \tag{14}$$

The total number of phenyl-like (non-naphthyl) groups is denoted by $n_{Ph,t}$:

Table 3

Rules for the conversion of benzene (A1) and naphthalene (A2).

Rule number	Constantinou-Gani groups	Tochigi groups	Remark
A1	6*ACH	Ph+H	benzene
A2	8*ACH+2*AC	Naf1+H	naphthalene

Table 4

Decomposition rules for aromatic Constantinou-Gani groups.

Rule number	original group	new groups
D1	$ACCH_3$	$AC+CH_3$
D2	$ACCH_2$	$AC+CH_2$
D3	ACCH	AC+CH
D4	$ACNH_2$	$AC+NH_2$
D5	ACCl	AC+Cl
D6	$ACNO_2$	$AC+NO_2(Ph)$
D7	ACF	AC+F

Table 5

Rules for the conversion of aromatic groups.

Rule number	Constantinou-Gani groups	Tochigi groups
A3	5*ACH+3*ACH	Naf1 or Naf2
A4	4*ACH+AC+ACOH	PhOH
A5	5*ACH+Bzgen	Bz
A6	4*ACH+2*AC	o-Ph or m-Ph or p-Ph
A7	5*ACH+AC	Ph

$$n_{Ph,t} = n_{Ph} + n_{o-Ph} + n_{m-Ph} + n_{p-Ph} + n_{PhOH} + n_{Bz} \qquad (15)$$

$n_{Ph,t}$ and n_{PhOh} can be straightforwardly calculated:

$$n_{Ph,t} = n_{main} - n_{cond} \qquad (16)$$

$$n_{PhOH} = \begin{cases} 0 \ in \ class \ 2 \\ n_{AcOH} \ in \ class \ 5 \ and \ 6 \end{cases} \qquad (17)$$

n_{Bz} is unambiguously determined in class 2 (as the only group containing a C=O fragment), but only an upper bound exists in the other possible classes:

$$n_{Bz} = n_{CH3CO} + n_{CH2CO} \ in \ class \ 2 \qquad (18)$$

$$n_{Bz} \leq \min(n_{AC} - 2n_{cond} - n_{PhOH}, n_{Bzgen}) \ in \ class \ 3, \ 4 \ and \ 6 \qquad (19)$$

where n_{Bzgen} is the number of Constantinou-Gani groups containing a C=O fragment, which can lead to the formation of a Bz group when they are adjacent to a phenyl or phenylene group. The Bzgen groups are: CH_3CO, CH_2CO in all classes and $CONH_2$, $CONHCH_3$, $CONHCH_2$, $CON(CH_3)_2$, $CONCH_3CH_2$, $CON(CH_2)_2$ in class 6 only.

Additionally, the following equation can be written for the number of trivalent groups:

$$n_{Ph} + 2(n_{o-Ph} + n_{m-Ph} + n_{p-Ph}) + n_{PhOH} + n_{Bz} = n_{Ar,tri} - 3n_{cond} \qquad (20)$$

Subtracting eq. (15) from (20) and expressing $n_{Ph,t}$ from eq. (16):

$$n_{o-Ph} + n_{m-Ph} + n_{p-Ph} = n_{Ar,tri} - n_{main} - 2n_{cond} \qquad (21)$$

From eq. (21) it follows that

$$0 \leq n_{Ar,tri} - n_{main} - 2n_{cond} \qquad (22)$$

If inequality (22) is violated, a new n_{main} value must be chosen, otherwise any solution of the system (14), (17)-(19), (21) can be used. If no n_{main} yields a solution, the conversion is not possible and the algorithm is aborted.

In the next step, the Constantinou-Gani groups other than ACH and ACOH are decomposed into AC groups and the corresponding non-aromatic groups (Rules D1-D7, Table 4). Rule D6 is special because

Table 6

Rules on the influence of second-order groups on the selection of cyclic groups.

Rule number	Second-order Constantinou-Gani group	First-order Constantinou-Gani group(s) in ring	Remark
R1	Alicyclic side chain $(C_{Cyclic}C_m)$	CH C	
R2	$C_{Cyclic}{=}O$	CH_2CO CH_2COO COO $CONHCH_2$ $CONH(CH_2)$	
R3	$CH_{m,cyclic}$-OH	CH C	
R4	$CH_{m,cyclic}$-NH_p-$CH_{n,cyclic}$	$2*CH_2+NH$ $2*CH+NH$ $2*C+NH$ $CH_2+CH+NH$ CH_2+C+NH $CH+C+NH$ $2*CH_2+N$ $2*CH+N$ $2*C+N$ CH_2+CH+N CH_2+C+N $CH+C+N$	$\mu{=}1$
R5	$CH_{m,cyclic}$-NH_p-$CH_{n,cyclic}$	$2*CH_2+NH$ $2*CH+NH$ $2*C+NH$ $CH_2+CH+NH$ CH_2+C+NH $CH+C+NH$ $2*CH_2+N$ $2*CH+N$ $2*C+N$ CH_2+CH+N CH_2+C+N $CH+C+N$ $2*CH$ $2*C$ $CH+C$	$\mu{\geq}2$
R6	$CH_{m,cyclic}$-S-$CH_{n,cyclic}$	$2*CH_2+S$ $2*CH+S$ $2*C+S$ CH_2+CH+S CH_2+C+S $CH+C+S$	$\mu{=}1$
R7	$CH_{m,cyclic}$-S-$CH_{n,cyclic}$	$2*CH$ $2*C$ $CH+C$	$\mu{\geq}2$

the result contains a Constantinou-Gani and a Tochigi group ($NO_2(Ph)$). After the decomposition, conversion to the Tochigi groups according to the numbers determined above takes place following Rules A3-A7 (Table 5). The rules should be applied in the order presented, specifically Rules A6 and A7 must not precede the others. The non-aromatic groups of the molecule will be converted later (except for NO_2).

2.4.3. Treatment of non-aromatic rings

The next step of the algorithm is to determine whether non-aromatic rings are also present. If the inequality (8) becomes equality, all rings are aromatic and non-cyclic parts are converted, which will be discussed later. If non-aromatic rings are also present, the potential interactions must be taken into account: if a non-aromatic and an aromatic ring share atoms, it influences the size of the former. Before describing this procedure, the treatment of molecules containing only non-aromatic rings should be discussed.

Not only the number of rings μ is known, but also the sizes of the rings having 7 or less members as these are second-order groups in the method of Constantinou and Gani. The size of a ring R_i is the sum of the lengths of the groups constituting it:

$$s_{R_i} = \sum_{j \in R_i} l_j \qquad (23)$$

During the conversion procedure, the task is to decide whether those groups where the Tochigi method distinguishes an acyclic and a cyclic variant are actually parts of rings or not. In this paper, the set of such groups is denoted as T_{Cyc}, and its elements are identified in Table A3. T_N is defined as the set of all other group g that are not elements of T_{Cyc} but can be part of a ring because they have a degree, that is, free valency, of at least 2:

$$T_N = \{\forall g \in T | g \notin T_{Cyc} \wedge d_g \geq 2\} \tag{24}$$

The method of Tochigi treats elements of T_N independently of their position.

The total length of groups selected from T_{Cyc} is $s_{R,Cyc}$. In the case of multiple rings, it is not necessary to consider how many such groups are in each ring, only the total value is needed. The value of $s_{R,Cyc}$ can be usually chosen between a lower and an upper bound. Once this is done, a number of groups must be selected from T_{Cyc} so that the sum of their length equals $s_{R,Cyc}$. If no such combination of groups exists, a new value of $s_{R,Cyc}$ is to be chosen. If certain second-order groups exist, some first-order ones must be selected (Rules R1-R7, Table 6) to be in ring(s). When multiple rows are shown for a rule, any of those can be freely chosen. By the second-order groups of Rules R4-R7, the entire groups must be part of a ring if μ=1. However, a second possibility exists if there are multiple rings: the hydrocarbon groups are part of two separate rings, while the amine or S group is acyclic. Subsequently, the conversion of the first-order groups to Tochigi groups is performed, which will be discussed later.

If the molecule contains a single ring (μ=1) with a size $s_R \leq 7$, the following inequalities are valid:

$$\max\left(s_R - \sum_{j \in T_N} l_j, 0\right) \leq s_{R,Cyc} \leq \min\left(s_R, \sum_{j \in T_{Cyc}} l_j\right) \tag{25}$$

If the molecule does not contain enough groups from T_N to form the ring, $s_{R,Cyc}$ has a non-zero lower bond. The upper bound is either simply the ring size or the number of available groups from T_{Cyc}. If $s_R > 7$, the actual ring size is not known and has no upper limit. In this case, the bounds of $s_{R,Cyc}$ change:

$$\max\left(8 - \sum_{j \in T_N} l_j, 0\right) \leq s_{R,Cyc} \leq \sum_{j \in T_{Cyc}} l_j \tag{26}$$

The change in the lower bound expresses that constructing a ring of length 8 is enough.

In the case of two rings (μ=2), the molecule can be bicyclic, when the two rings share at least two groups, a spirocompound, when the rings only share a single group, or it might consist of two rings sharing no group, in which case they will be called independent rings.

In the case of independent rings, s_r is simply replaced by the sum of the ring sizes $s_{R_1} + s_{R_2}$ in the bounds:

$$\max\left(s_{R_1} + s_{R_2} - \sum_{j \in T_N} l_j, 0\right) \leq s_{R,Cyc} \leq \min\left(s_{R_1} + s_{R_2}, \sum_{j \in T_{Cyc}} l_j\right) \tag{27}$$

If a ring has more than 7 members, its s_R value must be changed to 8 in the lower bound and removed from the upper bound.

For a bicyclic ring system, the longer the bridge (the groups shared by both rings), the fewer groups are needed to reach the same ring sizes. The length of the bridge l_b (including the two bridgeheads) in two rings with a total ring length $s_{\Sigma R}$ must be between the following limits:

$$2 \leq l_b \leq s_{\Sigma R} - 2 \tag{28}$$

At the upper bound, only one atom of each ring is not part of the bridge. The total ring length $s_{\Sigma R}$ is:

$$s_{\Sigma R} = s_{R_1} + s_{R_2} - l_b \tag{29}$$

The lower and upper bound of l_b thus correspond to a maximal and

minimal value of $s_{\Sigma R}$, respectively:

$$s_{\Sigma R,min} = \frac{s_{R_1} + s_{R_2}}{2} + 1 \tag{30}$$

$$s_{\Sigma R,max} = s_{R_1} + s_{R_2} - 2 \tag{31}$$

Consequently:

$$\max\left(\frac{s_{R_1} + s_{R_2}}{2} + 1 - \sum_{j \in T_N} l_j, 0\right) \leq s_{R,Cyc} \leq \min\left(s_{R_1} + s_{R_2} - 2, \sum_{j \in T_{Cyc}} l_j\right) \tag{32}$$

Both bounds of $s_{R,Cyc}$ are always lower for a bicyclic ring system than for independent rings since fewer groups are necessary for the same ring sizes. On the other hand, spiroring systems have an $s_{\Sigma R}$ value that is between those of bicyclic and independent rings for both bounds. Therefore, spirosystems must only be considered if other configurations are impossible due to the lack of suitable groups. As seen from Table 1, this is the case when there are less than two groups that can have a degree of three in the ring system:

$$b_3 + b_4 < 2 \tag{33}$$

In such a case, the following relationships are valid:

$$\max\left(s_{R_1} + s_{R_2} - 1 - \sum_{j \in T_N} l_j, 0\right) \leq s_{R,Cyc} \leq \min\left(s_{R_1} + s_{R_2} - 1, \sum_{j \in T_{Cyc}} l_j\right) \tag{34}$$

Otherwise, as seen earlier, bicycles determine the lower bound and independent rings the upper bound:

$$\max\left(\frac{s_{R_1} + s_{R_2}}{2} + 1 - \sum_{j \in T_N} l_j, 0\right) \leq s_{R,Cyc} \leq \min\left(s_{R_1} + s_{R_2}, \sum_{j \in T_{Cyc}} l_j\right) \tag{35}$$

With an unrestricted number of rings, the upper bound is determined by assuming that all rings are independent from each other:

$$s_{R,max} = \sum_{i=1}^{\mu} s_{R_i} \tag{36}$$

In order to determine $s_{R,min}$, the s_{R_i} values are ordered in a decreasing order:

$$s_{R_1} \geq s_{R_2} \geq \cdots \geq s_{R_\mu} \tag{36a}$$

The lowest value (the largest rings with the same number of groups) is obtained by pairing the rings consecutively to form bicyclic systems:

$$s_{R,min} = \sum_{j=1}^{j_f} \left(\frac{s_{R_{2j-1}} + s_{R_{2j}}}{2} + 1\right) + (\mu \bmod 2)s_{R_\mu} \tag{37}$$

where j_f is the number or ring pairs:

$$j_f = \frac{\mu - \mu \bmod 2}{2} \tag{38}$$

The last term in eq. (37) refers to the smallest ring when the number of rings is odd. Eq. (37) can be simplified to obtain:

$$s_{R,min} = \frac{\sum_{i=1}^{\mu - \mu \bmod 2} s_{R_i}}{2} + \frac{\mu - \mu \bmod 2}{2} + (\mu \bmod 2)s_{R_\mu} \tag{39}$$

The bounds of $s_{R,Cyc}$ are therefore:

Table 7

Minimal and maximal l_{Ar} values.

Condition	$l_{Ar,min}$	$l_{Ar,max}$
$n_{ACCH_2} + n_{ACCH} + n_{Bz} = 0$		4
$n_{ACCH_2} + n_{ACCH} + n_{Bz} = 1$		5
$n_{ACCH_2} + n_{ACCH} + n_{Bz} = 2$		6
$n_{AC} = 0$	4	
$n_{AC} = 1$	3	
$n_{AC} = 2$	2	

$$\max\left(\frac{\sum_{i=1}^{\mu-\mu \bmod 2} s_{R_i}}{2} + \frac{\mu - \mu \bmod 2}{2} + (\mu \bmod 2)s_{R_\mu} - \sum_{j \in T_N} l_j, 0\right) \le s_{R,Cyc}$$

$$\le \min\left(\sum_{i=1}^{\mu} s_{R_i}, \sum_{j \in T_{Cyc}} l_j\right) \tag{40}$$

As in the case of $\mu=2$, there might not be enough groups to form all the bicyclic systems and some of them must be spirosystems. This happens if:

$$c = b_3 + b_4 < \mu - \mu \bmod 2 \tag{41}$$

In such a case, only c-c mod 2 pairs can be bicyclic. To obtain the lowest $s_{R,min}$ value, these rings must be the largest ones. $s_{R,min}$ is then:

$$s_{R,min} = \sum_{j=1}^{\frac{c-c \bmod 2}{2}} \left(\frac{s_{R_{2j-1}} + s_{R_{2j}}}{2} + 1\right)$$
$$+ \sum_{j = \frac{c-c \bmod 2}{2} + 1}^{\frac{\mu - \mu \bmod 2}{2}} \left(\frac{s_{R_{2j-1}} + s_{R_{2j}}}{2} - 1\right) + (\mu \bmod 2)s_{R_\mu} \tag{42}$$

$$s_{R,min} = \frac{\sum_{i=1}^{c-c \bmod 2} s_{R_i}}{2} + \sum_{i=c-c \bmod 2+1}^{\mu-\mu \bmod 2} s_{R_i} + c - c \bmod 2 - \frac{\mu - \mu \bmod 2}{2}$$
$$+ (\mu \bmod 2)s_{R_\mu} \tag{43}$$

To determine the $s_{R,max}$ value, the c-c mod 2 largest ring pairs are considered to be independent:

$$s_{R,max} = \sum_{j=1}^{\frac{c-c \bmod 2}{2}} \left(s_{R_{2j-1}} + s_{R_{2j}}\right)$$
$$+ \sum_{j = \frac{c-c \bmod 2}{2} + 1}^{\frac{\mu - \mu \bmod 2}{2}} \left(\frac{s_{R_{2j-1}} + s_{R_{2j}}}{2} - 1\right) + (\mu \bmod 2)s_{R_\mu} \tag{44}$$

By simplifying eq. (44):

$$s_{R,max} = \sum_{i=1}^{\mu} s_{R_i} - \frac{1}{2}(\mu - \mu \bmod 2 - c + c \bmod 2) \tag{45}$$

The bounds are therefore:

After the group from T_{Cyc} are selected, the conversion can take place following Rules 1–78, which will be discussed later.

After discussing the treatment of non-aromatic cyclic molecules, the mixed systems containing both aromatic and non-aromatic rings can be addressed. Due to the limitations of Tochigi's method, an aromatic ring can only share atoms with a single non-aromatic ring. In the vein of limiting ring systems to a maximum of two rings, the existence of ring systems containing an aromatic and more than two non-aromatic rings is also prohibited. Naphytl groups cannot be parts of ring systems.

If aromatic and non-aromatic rings do not share atoms (that is, they are independent of each other), they can be treated according to the procedures presented above. However, the presence of mixed ring systems modifies the bounds of $s_{R,Cyc}$ since a lower number of groups is required for the same size of the non-aromatic ring. Table 7 contains the minimal ($l_{Ar,min}$) and maximal ($l_{Ar,max}$) lengths that the aromatic ring can provide. For example, if two AC groups are present, $l_{Ar,min}=2$ since if they are adjacent, only these two groups are part of the non-aromatic ring. On the other hand, if there are no AC groups, there must be two groups of length two (ACCH$_2$, ACCH or Bz) and even if they are adjacent, the rings will share at least four members. In a molecule containing a single mixed ring system, the bounds of $s_{R,Cyc}$ are:

$$\max\left(s_R - l_{Ar,max} - \sum_{j \in T_N} l_j, 0\right) \le s_{R,Cyc} \le \min\left(s_R - l_{Ar,min}, \sum_{j \in T_{Cyc}} l_j\right) \tag{47}$$

In a general case, the number and type of the possible combinations of the different non-aromatic and aromatic groups become too numerous to be easily described by general rules. Instead, the following procedure is proposed. The upper bound of $s_{R,Cyc}$ can be determined by assuming that all rings are independent from each other. Writing the inequalities with a general value $s_{R,min}$:

$$\max\left(s_{R,min} - \sum_{j \in T_N} l_j, 0\right) \le s_{R,Cyc} \le \min\left(\sum_{i=1}^{\mu} s_{R_i}, \sum_{j \in T_{Cyc}} l_j\right) \tag{48}$$

$s_{R,min}$ can be obtained by generating all the possible pairs of rings and summing their individual $s_{R,min}$ values. The true $s_{R,min}$ is the lowest among all the results. If the number of non-aromatic rings is i and the number of eligible aromatic rings is j where:

$$j = n_{o-Ph} + n_{m-Ph} + n_{p-Ph} + n_{Bz} \tag{49}$$

thus, there are the following number of possible ring pairs:

$$i\left(j + \frac{i-1}{2}\right) \tag{50}$$

A last question related to rings is that of anhydrides. If there are $n_{CAnhgen}$ number of CH$_2$COO and COO groups selected to be part of rings and:

$$n_{CH2COO,CG} + n_{COO,CG} - n_{CAnhgen} < n_{II26} \tag{51}$$

where $n_{CH2COO,CG}$ and $n_{COO,CG}$ are the numbers of the respective

$$\max\left(\frac{\sum_{i=1}^{c-c \bmod 2} s_{R_i}}{2} + \sum_{i=c-c \bmod 2+1}^{\mu-\mu \bmod 2} s_{R_i} + c - c \bmod 2 - \frac{\mu - \mu \bmod 2}{2}\right.$$
$$\left. + (\mu \bmod 2)s_{R_\mu} - \sum_{j \in T_N} l_j, 0\right) \le s_{R,Cyc} \le \min\left(\sum_{i=1}^{\mu} s_{R_i} - \frac{1}{2}(\mu - \mu \bmod 2 - c + c \bmod 2), \sum_{j \in T_{Cyc}} l_j\right) \tag{46}$$

Table 8
Prediction accuracies of the different methods used.

		AARD$_{all}$, %	AAE$_{all}$	Outliers, pcs	AARD$_{<99.99}$, %	AAE$_{<99.99}$
C&G	P$_{crit}$, bar	7.80	2.88	0	7.80	2.88
	T$_{crit}$, K	8.64	35.6	0	8.64	35.6
	T$_{crit}$, °C	34.3	35.6	9	17.1	26.9
Tochigi based on RefProp data	P$_{sat,@0.7T}$, bar	49.1	0.97	4	44.9	0.96
	P$_{sat,@0.7T,boil}$, bar	10.3	0.30	2	7.52	0.19
	ω, 1	155	0.30	31	40.9	0.14
	ω$_{boil}$, 1	22.7	0.04	4	12.0	0.03
Tochigi based on C&G results	P$_{sat,@0.7T}$, bar	128	2.31	18	57.3	1.22
	P$_{sat,@0.7T,boil}$, bar	12,502	516	9	29.1	0.59
	ω, 1	369	0.82	49	40.5	0.15
	ω$_{boil}$, 1	275	0.54	19	29.4	0.09
C&G	ω$_{C\&G}$, 1	10.5	0.04	0	10.5	0.04

Constinantinou-Gani groups in the molecule and n_{II26} is the number of second-order anhydride groups (CO-O-CO), the total number of cyclic anhydride groups (n_{CAnht}) must be:

$$n_{II26} - \left(n_{CH2COO,CG} + n_{COO,CG} - n_{CAnhgen}\right) \leq n_{CAnht} \leq n_{II26} \qquad (52)$$

and at least $n_{II26} - \left(n_{CH2COO,CG} + n_{COO,CG} - n_{CAnhgen}\right)$ CH$_2$CO groups must be additionally included in rings.

2.4.4. Group conversion rules

The algorithm's final step is converting the Constantinou-Gani groups to the Tochigi ones, except for the already converted aromatic groups. Before discussing the general conversion rules, a number of special cases are distinguished (Rules S1-S, given in Table A4 of the Supplementary material), with each case being a single molecule; that is, the molecule consists only of the groups listed in Table A4. If such a special case is encountered, the general rules are omitted, and the conversion is done. In the case of aromatic molecules, the already converted aromatic Tochigi groups are shown instead of their constituent Constantinou-Gani groups.

Table A5 in the Supplementary material shows the general conversion Rules 1–78. The numbering corresponds to the identification numbers of the Constantinou-Gani groups. There is an unambiguous rule for certain groups, while for several others, the conversion only depends on whether the group is in a ring. However, several groups have multiple rules depending on their position relative to other groups, which is sometimes, but most often not, described by a second-order group. The conversion rules of such groups are discussed in detail in Section 2 of the Supplementary material.

As mentioned already, some of the groups have a more complicated set of conversion rules. In many such cases, multiple options are available due to the presence of either other first-order groups or second-order groups that can influence the conversion. The number of Constantinou-Gani groups converted by each option has an upper limit, typically including the available number of influencing other groups or the known number of relevant second-order groups. One of the options is always an exception, by which the remaining instances of the group are converted. In some other cases, the multiplicity of options exists because a group can be completely or partially in a ring or because the group can be partially in a ring in two different ways. Such more complicated rules are applied for the following groups: "F (except as above)" (denoted simply F here, Rules 65a-d), OH (Rules 16a-e), CH$_3$COO (Rules 21a-d), CH$_2$COO (Rules 22a-i), HCOO (Rules 23a-d), COO (Rules 61b-h), CH$_3$CO (Rules 18a-c), CH$_2$CO (Rules 19b-d), CH$_3$O (Rules 24a-c), CH$_2$O (Rules 25b-e), CH-O (Rules 26b-e), CHNH (Rules 32b-c), CH$_2$N (Rules 34b-c), CON(CH$_2$)$_2$ (Rules 71b-c), CH$_2$S and CHS (Rules 75a-d, 76a-f, 75+76a-c), COOH (Rules 39a-c), C$_2$H$_5$O$_2$ (Rule 72), and C$_2$H$_4$O$_2$ (Rules 73a-d). The rules are explained in detail in Section 2 of the Supplementary material. As illustrated in the Supplementary material, the results depend on the order of (some of) the rules; thus, the order described there should be respected.
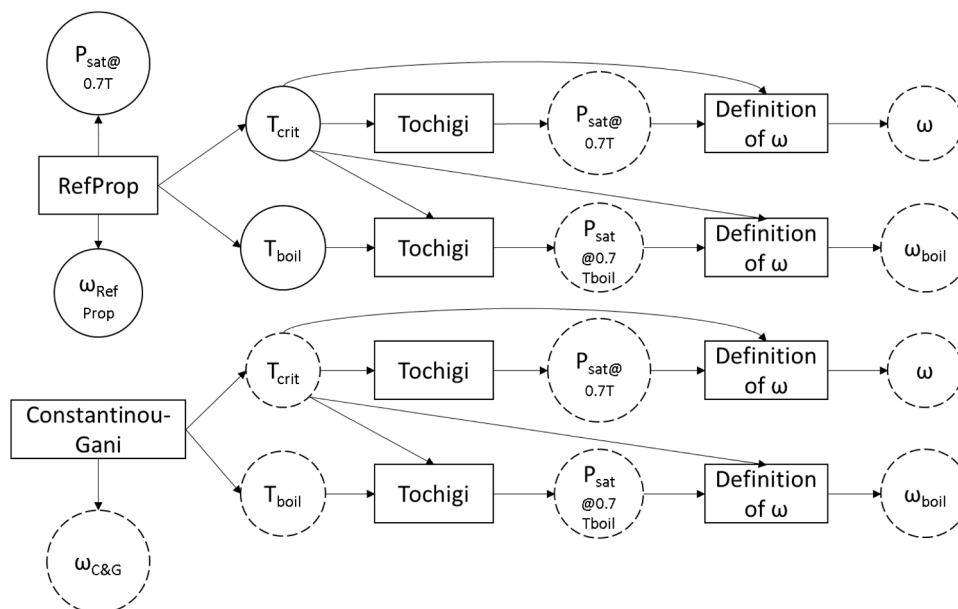
**Fig. 5.** Schema for obtaining the predicted (dashed circles) and measured (continuous circles) parameters.
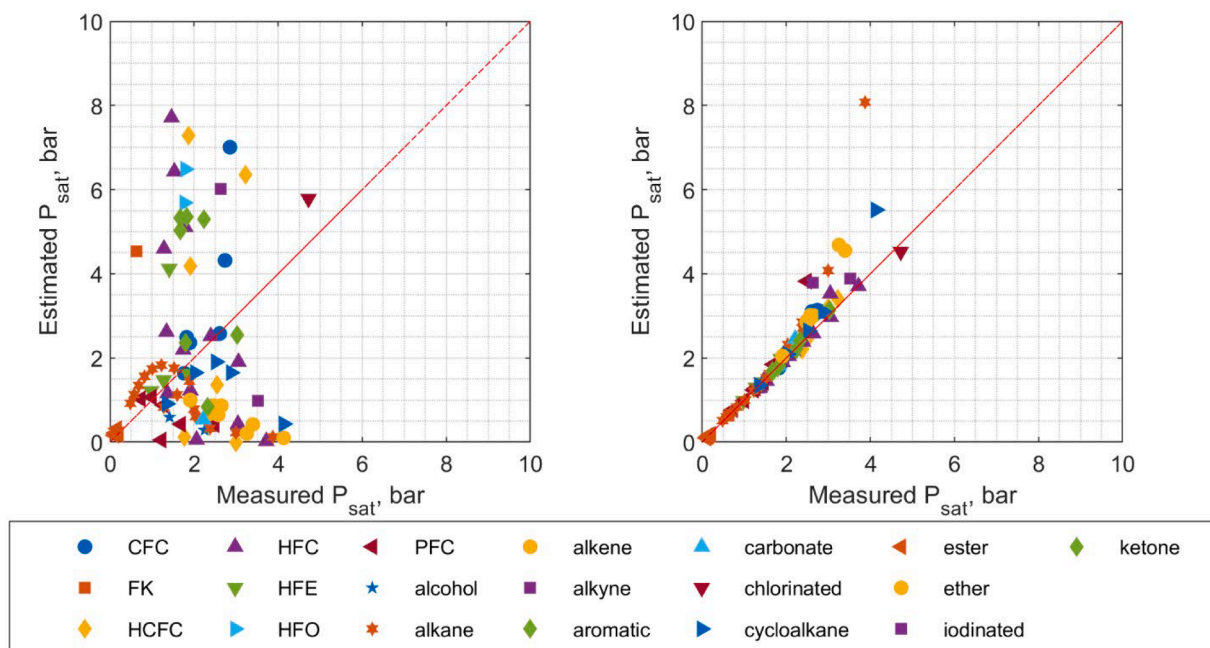
Fig. 6. Tochigi estimate for saturation pressure vs. measured saturation pressure at $0.7 \cdot T_{crit}$. Left: estimation is based on the Constantinou and Gani [18] method; right: estimation is based on the critical values and normal boiling point taken from the RefProp database. (CFC: chlorofluorocarbons, HCFC: hydro-chlorofluorocarbons, HFC: hydrofluorocarbons, HFE: hydrofluoroethers, HFO: hydrofluoroolefins, FK: fluorinated ketones, PFC: perfluorocarbons).
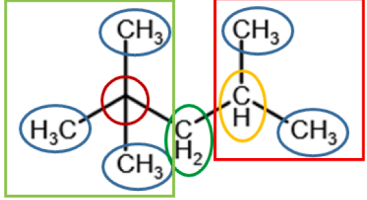
3. Results

The results of reclassifying materials from Constantinou and Gani's first- and second-order functional groups to Tochigi's groups are demonstrated in Table A7 in the Supplementary material. Of the 88 pure fluids from RefProp (version 9.1) developed and maintained by NIST [46], 58 materials required only first-order groups, while 30 compounds necessitated second-order groups, as well, for more precise estimation.
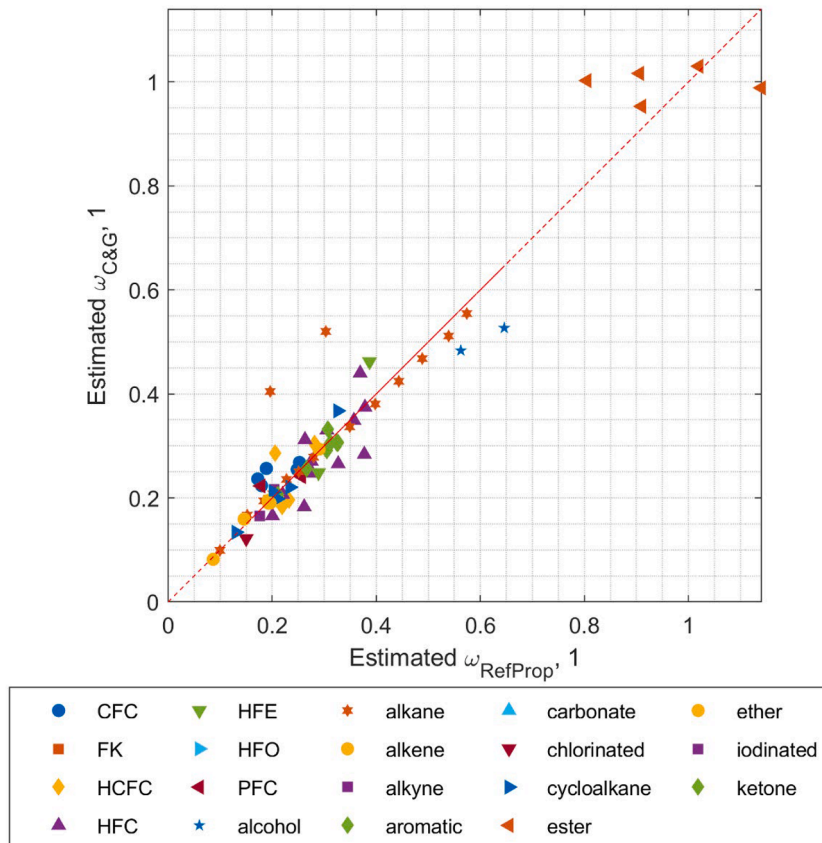
Fig. 7. O'Connell-extended Constantinou-Gani estimate for the acentric factor vs. acentric factor from the NIST database. (CFC: chlorofluorocarbons, HCFC: hydrochlorofluorocarbons, HFC: hydrofluorocarbons, HFE: hydrofluoroethers, HFO: hydrofluoroolefins, FK: fluorinated ketones, PFC: perfluorocarbons).

Table 9

Detailed results for isooctane.

ID	Name	Type	Structural formula
28	isooctane (2,2,4-trimethylpentane)	alkane (acyclic)	

Constantinou-Gani first-order groups		5xCH$_3$ (group no. 1), CH$_2$ (2), CH (3), C (4)			
Constantinou-Gani second-order groups		(CH$_3$)$_2$CH (1), (CH$_3$)$_3$C (1)			
Conversion rules applied	C1	Class 1: saturated hydrocarbons, unsaturated hydrocarbons, and aromatic hydrocarbons			
	1	CH$_3$→CH$_3$			
	2a	CH$_2$→CH$_2$ (acyclic)			
	3a	CH→CH (acyclic)			
	4a	C→C (acyclic)			
Tochigi groups		5xCH$_3$ (group no. 1), CH$_2$ (2), CH (3), C (4), Const. (23)			
Properties from RefProp	P$_{crit}$, bar	25.72	T$_{crit}$, °C	270.9	
	T$_{boil}$, °C	99.21			
	P$_{sat@0.7T}$, bar	1.279	ω$_{RefProp}$	0.303	
Properties from Constantinou-Gani	P$_{crit}$, bar	25.61	AARD, %	0.422	
	T$_{crit}$, °C	267.2	AARD, %	3.41	
	T$_{boil}$, °C	96.26	AARD, %	3.07	
	ω$_{C\&G}$	0.5201	AARD, %	71.7	
Properties from Tochigi based on RefProp data	P$_{sat@0.7T}$, bar	0.8956	AARD, %	30.0	
	ω	0.4582	AARD, %	51.2	
	P$_{sat@0.7Tboil}$, bar	1.2954	AARD, %	1.30	
	ω$_{boil}$	0.2979	AARD, %	1.69	
Properties from Tochigi based on C&G results	P$_{sat@0.7T}$, bar	0.8323	AARD, %	34.9	
	ω	0.4900	AARD, %	61.7	
	P$_{sat@0.7Tboil}$, bar	1.3159	AARD, %	2.91	
	ω$_{boil}$	0.2910	AARD, %	3.94	
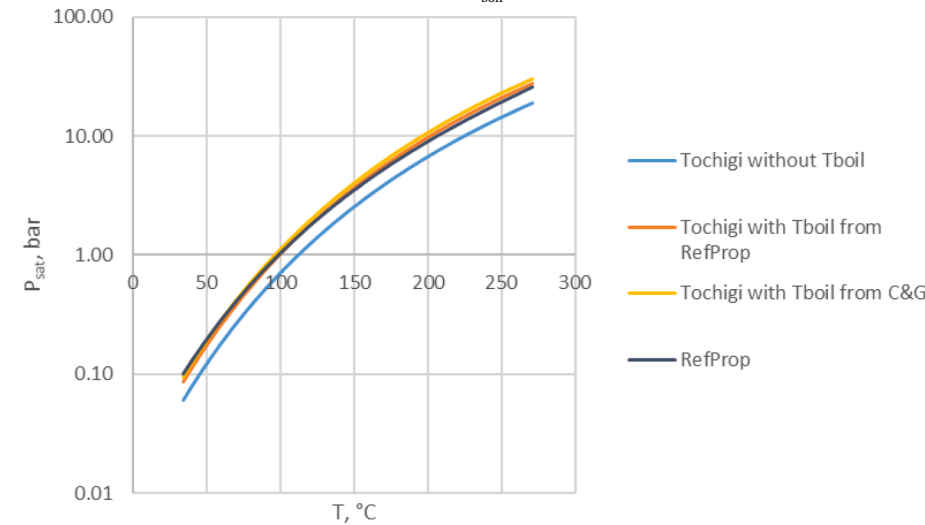
P$_{sat}$ estimated by different methods as a function of temperature

The table's odd columns display the types of functional groups from Constantinou and Gani [18] and Tochigi et al. [24], while the even columns show the corresponding quantities required. It must be noted that by using Constantinou-Gani groups, three molecules cannot be correctly described. By HCFO-1233zd-E ((1E)-1-chloro-3,3,3-tri-fluoroprop-1-ene) and R40 (chloromethane), a hydrogen atom is missing from the groups, by dodecafluoro-2-methylpentan-3-one, there are two superfluous hydrogens.

To showcase the algorithm's applicability, vapor pressure at 0.7·T$_{crit}$ (P$_{sat@0.7T}$) were determined and analyzed for the compounds. The prediction accuracy of GCM models in the literature is usually presented by four statistical parameters, namely the average absolute relative deviation (AARD), the average percent deviation (bias), the average absolute error (AAE), and the root mean square error (RMS), which are defined as follows:

$$AARD = \frac{1}{N} \sum_{i=1}^{N} 100 \cdot \left| \frac{y_{exp,i} - y_{cal,i}}{y_{exp,i}} \right| \tag{53}$$
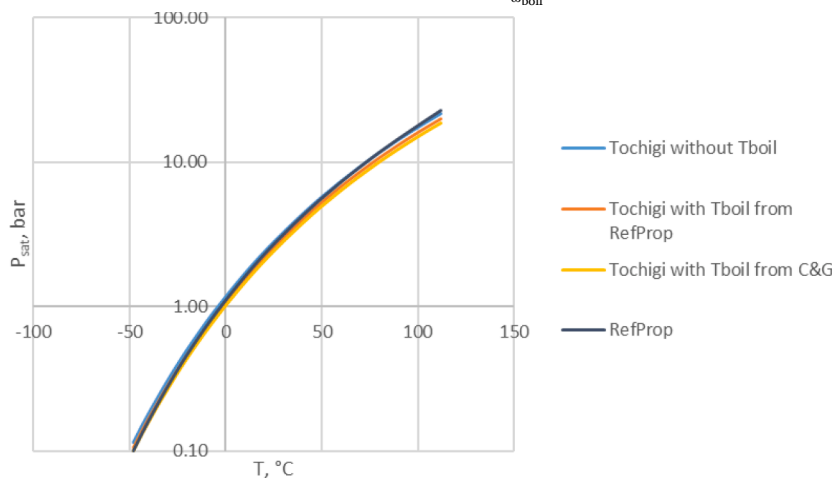
$$bias = \frac{1}{N} \sum_{i=1}^{N} 100 \cdot \frac{y_{exp,i} - y_{cal,i}}{y_{exp,i}} \tag{54}$$

$$AAE = \frac{1}{N} \sum_{i=1}^{N} \left| y_{exp,i} - y_{cal,i} \right| \tag{55}$$

Table 10
Detailed results for C_4F_{10} (perfluorobutane).

ID	Name			Type		Structural formula	
43	C_4F_{10} (perfluorobutane)			PFC (acyclic)			

Constantinou-Gani first-order groups		$2xCF_3$ (group no. 58), $2xCF_2$ (59)			
Constantinou-Gani second-order groups		–			
Conversion rules applied		C8	Class 6: halogen-containing compounds		
		58	$CF_3 \rightarrow CF_3$		
		59a	$CF_2 \rightarrow CF_2$ (acyclic)		
Tochigi groups		$5xCF_3$ (group no. 182), CF_2 (188), Const. (234)			
Properties from RefProp		P_{crit}, bar	23.22	T_{crit}, °C	113.2
		T_{boil}, °C	-2.027		
		$P_{sat@0.7T}$, bar	0.9946	$\omega_{RefProp}$	0.372
Properties from Constantinou-Gani		P_{crit}, bar	23.12	AARD, %	0.471
		T_{crit}, °C	113.1	AARD, %	0.080
		T_{boil}, °C	-0.384	(AARD, %	81.1)
		$\omega_{C\&G}$	–		
Properties from Tochigi based on RefProp data		$P_{sat@0.7T}$, bar	1.0693	AARD, %	7.50
		ω	0.3369	AARD, %	9.45
		$P_{sat@0.7Tboil}$, bar	0.9861	AARD, %	0.861
		ω_{boil}	0.3720	AARD, %	0.0073
Properties from Tochigi based on C&G results		$P_{sat@0.7T}$, bar	1.0669	AARD, %	7.26
		ω	0.3378	AARD, %	9.19
		$P_{sat@0.7Tboil}$, bar	0.9232	AARD, %	7.18
		ω_{boil}	0.4006	AARD, %	7.70

P_{sat} estimated by different methods as a function of temperature

P_{sat} estimated by different methods as a function of temperature

$$RMS = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(y_{exp,i} - y_{cal,i} \right)^2} \qquad (56)$$

where y_{exp} is the experimental value of a material property studied, taken from the RefProp database and y_{calc} is its predicted value. As the formulas show, AAE and RMS measure the average difference between predicted and actual values slightly differently. RMS is commonly used when large errors should be penalized more, while AAE is preferred when a robust predictor shall be applied that treats all errors equally and is less affected by outliers. Furthermore, bias is less helpful in this context as it returns the average of values with different signs. Therefore, Table 8 evaluates the results with AARD and AAE while accounting for the outliers, defined as values exceeding 100 % error.

As the acentric factor is also part of the investigation, both data from the RefProp database and the Constantinou and Gani methods were employed to obtain the critical temperature as illustrated in Fig. 5. While the AARD and AAE values calculated in Kelvin for the critical temperature and in bar for the pressure were relatively low and did not exceed the 100% error rate (outliers) in any of the cases, the AARD results were

less satisfactory when the temperature was expressed in degrees Celsius. As the constants of the semi-empirical Antoine equation were fitted to temperatures set in °C, this reduced the estimate's accuracy.

The saturated vapor pressure at $0.7 \cdot T_{crit}$ and the acentric factor were determined in four different ways (Fig. 5). The saturation pressure was calculated with the Antoine constants defined with group contribution only ($P_{sat@0.7T}$ and ω) and with AntA being determined from the normal boiling point of the compounds ($P_{sat@0.7T,boil}$ and ω_{boil}). Also, the critical parameters and the boiling point temperature were either taken from the RefProp database or estimated by Constantinou and Gani [18].

The results show that the most accurate estimation is obtained when the Antoine equation is calculated based on the AntA determined from the normal boiling point, and the material properties are taken from the NIST RefProp Database [47]. Accuracy is significantly reduced if the properties are estimated using the Constantinou and Gani [18] method. In this case, the group contribution estimate of AntA gives more precise outcomes (Fig. 6). Although the $AARD_{all}$ and AAE_{all} values of the vapor pressure obtained by using the boiling points are very high, this is, in fact, caused by the inability of the Tochigi method to estimate the vapor pressure of ethylene, the lightest material studied here. In this case,

removing ethylene (where P_{sat} was estimated as 45,274 bar) decreases $AARD_{all}$ to 49.1 %, which is already much lower than the one obtained without using the boiling point. Note that the Tochigi fails for ethylene even if the T_{boil} is provided from the RefProp database.

The calculations were also evaluated on the basis of compound classes (Tables A8 and A9 in the Supplementary material). Although the comparison cannot be considered representative, as the sample sizes of the classes are non-identical and also low, the results show that hydrofluoroethers (HFEs), esters, aromatics, and chlorofluorocarbons (CFCs) are predicted more accurately than the rest. While alcohols also appear to be accurate, it's worth noting that the sample only included two compounds from this category (Fig. 6). Interestingly, the results predicted by the Tochigi method based on Constantinou-Gani estimations can be slightly worse for some compound classes if the normal boiling point is used. This can happen if the error of the estimation of the boiling point by the Constantinou-Gani method is very high.

While the results calculated with normal boiling point-based AntA constants are promising, especially when the required material properties are obtained from the RefProp database, overall, it can be concluded that the Constantinou et al. method [36] offers more precise acentric factors than the other procedures studied based on the Tochigi method, assuming all the functional groups necessary for describing the compound are available (Fig. 7).

As an illustration, detailed results of the Constantinou-Gani and the Tochigi methods are given for two compounds, isooctane and C_4F_{10}, in Tables 9 and 10, respectively. Also included are the data from RefProp, the functional groups of both methods and the conversion rules applied. A plot of P_{sat} estimated by the different methods as a function of temperature between $P_{sat}=0.1$ bar to close to T_{crit} is shown, as well.

Isooctane (Table 9) was chosen to illustrate a case where both GCMs are applicable for the estimation of ω, but the Tochigi method outperforms the Constantinou-Gani one. The Constantinou-Gani and Tochigi groups (shown by ellipses in the formula) are identical in this case; thus, the conversion is straightforward. Constantinou-Gani also uses two second-order groups (shown by rectangles), but this does not affect the procedure. The prediction error of Constantinou-Gani is low. The exception is ω, where the Tochigi method provides better results even without taking T_{boil} into account. The prediction errors of the Tochigi method are slightly better if T_{crit} and T_{boil} are taken from RefProp instead of calculated by the Constantinou-Gani method, but, in both cases, they are greatly improved by taking T_{boil} into account.

C_4F_{10} (Table 10) was chosen to show a case where the Tochigi method gives a very good estimation of both P_{sat} and ω, while the O'Connell-extended Constantinou-Gani method cannot predict ω due to the missing contribution value of the CF_2 group. Once again, the identical Constantinou-Gani and Tochigi groups make the conversion simple. The Tochigi method provides good estimations even without taking T_{boil} into account; doing so leads to considerable improvements if RefProp data is used but only slight ones if T_{crit} and T_{boil} are calculated by the Constantinou-Gani method.

4. Conclusion

An extension of the well-known group contribution method of Constantinou and Gani with the method of Tochigi is proposed to enable the prediction of the vapor pressure, as well. The O'Connell-extended Constantinou and Gani method can also predict the molecules' acentric factor, but the contribution values are missing for several functional groups. In such cases, the vapor pressure calculated with the Tochigi method can be used to obtain the acentric factor. An automatic conversion algorithm is proposed, which converts the Constantinou-Gani groups to Tochigi ones. The difficulties of creating such an algorithm are likewise discussed, which arise from the differences of the two sets of groups, the undefined ring concept of Constantinou-Gani, and the lack of clear group definitions and guidelines of Tochigi.

The investigation has demonstrated that the Tochigi method yields

higher precision in estimating saturated vapor pressures when the normal boiling point and critical parameters are available from the RefProp database. In contrast, relying on the Constantinou-Gani method provides less precise results. More favorable outcomes are attainable in this case when the estimation process is grounded solely on critical parameters. The acentric factor, determined from the saturation pressure, shows a lower accuracy level than those computed using the extended Constantinou-Gani method; however, the Tochigi method is useful when the contributions for the necessary Constantinou-Gani functional groups are unavailable.

The complexity of the proposed conversion algorithm highlights that group contribution methods were generally not formulated with their automation in mind and require manual processing with often arbitrary decisions by more complex molecules. The complexity would be further magnified should more than two group contribution methods be used since, in such a case, the choices made during the conversions between different pairs of methods must also be consistent with each other. Therefore, with the spreading use of numerical optimization and machine learning, easy-to-automate group contribution methods are needed.

CRediT authorship contribution statement

Axel Groniewsky: Writing – review & editing, Writing – original draft, Conceptualization. **László Hégely:** Writing – review & editing, Writing – original draft, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:
Axel Groniewsky reports financial support was provided by Hungarian Academy of Sciences. Axel Groniewsky reports financial support was provided by Ministry for Innovation and Technology Hungary. Laszlo Hegely reports financial support was provided by Ministry of Culture and Innovation of Hungary. Axel Groniewsky reports a relationship with Hungarian Academy of Sciences that includes: funding grants. Axel Groniewsky reports a relationship with Ministry for Innovation and Technology Hungary that includes: funding grants. Laszlo Hegely reports a relationship with Ministry of Culture and Innovation of Hungary that includes: funding grants. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.fluid.2024.114148.

References

[1] K.G. JOBACK, R.C. REID, Estimation of pure-component properties from group-contributions, Chem. Eng. Commun. 57 (1–6) (Jul. 1987) 233–243, https://doi.org/10.1080/00986448708960487.

[2] B.E. Poling, J.M. Prausnitz, J.P. O'Connell, The properties of gases and liquids, McGRAW-HILL, 2001, 5th editio.

[3] J. Song, H.H. Song, Computer-aided molecular design of environmentally friendly solvents for separation processes, Chem. Eng. Technol. 31 (2) (Feb. 2008) 177–187, https://doi.org/10.1002/ceat.200700233.

[4] J.A. Lazzús, Estimation of density as a function of temperature and pressure for imidazolium-based ionic liquids using a multilayer net with particle swarm optimization, Int. J. Thermophys. 30 (3) (2009) 883–909, https://doi.org/10.1007/s10765-009-0591-5.

[5] J.A. Lazzús, Prediction of flash point temperature of organic compounds using a hybrid method of group contribution + Neural network + Particle swarm optimization, Chinese J. Chem. Eng. 18 (5) (2010) 817–823, https://doi.org/10.1016/S1004-9541(09)60133-6.

[6] W. Su, L. Zhao, S. Deng, Simultaneous working fluids design and cycle optimization for Organic Rankine cycle using group contribution model, Appl. Energy 202 (2017) 618–627, https://doi.org/10.1016/j.apenergy.2017.03.133.

[7] M.T. White, O.A. Oyewunmi, M.A. Chatzopoulou, A.M. Pantaleo, A.J. Haslam, C. N. Markides, Computer-aided working-fluid design, thermodynamic optimisation and thermoeconomic assessment of ORC systems for waste-heat recovery, Energy 161 (2018) 1181–1198, https://doi.org/10.1016/j.energy.2018.07.098.

[8] B. Liu, Y. Yu, H. Liu, Z. Cui, W. Tian, Multi-criteria screening of organic ethanolamines for efficient CO2 capture based on group contribution method, Processes 11 (5) (2023), https://doi.org/10.3390/pr11051524.

[9] Y. Peng, W. Su, N. Zhou, L. Zhao, How to evaluate the performance of sub-critical organic rankine cycle from key properties of working fluids by group contribution methods? Energy Convers. Manag. 221 (2020) 113204 https://doi.org/10.1016/j.enconman.2020.113204.

[10] J.Y. Ten, Z.H. Liew, X.Y. Oh, M.H. Hassim, N. Chemmangattuvalappil, Computer-aided molecular design of optimal sustainable solvent for liquid-liquid extraction, Process Integr. Optim. Sustain. 5 (2) (2021) 269–284, https://doi.org/10.1007/s41660-021-00166-7.

[11] M. Lampe, M. Stavrou, J. Schilling, E. Sauer, J. Gross, A. Bardow, Computer-aided molecular design in the continuous-molecular targeting framework using group-contribution PC-SAFT, Comput. Chem. Eng. 81 (Oct. 2015) 278–287, https://doi.org/10.1016/j.compchemeng.2015.04.008.

[12] J. Schilling, M. Lampe, J. Gross, A. Bardow, One-stage approach for the integrated design of ORC processes and working fluid using PC-SAFT, in: Z. Kravanja, M.B.T. C.A.C.E. Bogataj (Eds.), 26 European symposium on computer aided process engineering 38, Elsevier, 2016, pp. 1335–1340.

[13] E. Wang, J. Mao, B. Zhang, Y. Wang, On the CAMD method based on PC-SAFT for working fluid design of a high-temperature organic Rankine cycle, Energy 263 (2023) 125935, https://doi.org/10.1016/j.energy.2022.125935.

[14] J. Wang, et al., Computer-aided molecular design of CO2-based mixture working fluid harvesting engine waste heat, Appl. Therm. Eng. 227 (2023) 120481, https://doi.org/10.1016/j.applthermaleng.2023.120481.

[15] B.H. Gebreslassie, U.M. Diwekar, Efficient ant colony optimization (EACO) for solvent selection using computer aided molecular design, in: K.V. Gernaey, J. K. Huusom, R.B.T.C.A.C.E. Gani (Eds.), 12 International symposium on process systems engineering and 25 European symposium on computer aided process engineering 37, Elsevier, 2015, pp. 743–748.

[16] M. Shahmohammadi, R. Mukherjee, C.G. Takoudis, U.M. Diwekar, Optimal design of novel precursor materials for the atomic layer deposition using computer-aided molecular design, Chem. Eng. Sci. 234 (2021) 116416, https://doi.org/10.1016/j.ces.2020.116416.

[17] V. Mann, R. Gani, V. Venkatasubramanian, Group contribution-based property modeling for chemical product design: A perspective in the AI era, Fluid Phase Equilib 568 (2023) 113734, https://doi.org/10.1016/j.fluid.2023.113734.

[18] L. Constantinou, R. Gani, New group contribution method for estimating properties of pure compounds, AIChE J 40 (10) (Oct. 1994) 1697–1710, https://doi.org/10.1002/aic.690401011.

[19] M. Bilde, et al., Saturation vapor pressures and transition enthalpies of low-volatility organic molecules of atmospheric relevance: from dicarboxylic acids to complex mixtures, Chem. Rev. 115 (10) (May 2015) 4115–4156, https://doi.org/10.1021/cr5005502.

[20] P. Li, P.S. Ma, S.Z. Yi, Z.G. Zhao, L.Z. Cong, A new corresponding-states group-contribution method (CSGC) for estimating vapor pressures of pure compounds, Fluid Phase Equilib 101 (1994) 101–119, https://doi.org/10.1016/0378-3812(94)02607-6.

[21] P.B. Myrdal, S.H. Yalkowsky, Estimating pure component vapor pressures of complex organic molecules, Ind. Eng. Chem. Res. 36 (6) (Jun. 1997) 2494–2499, https://doi.org/10.1021/ie950242l.

[22] Y. Nannoolal, J. Rarey, D. Ramjugernath, Estimation of pure component properties: Part 3. Estimation of the vapor pressure of non-electrolyte organic compounds via group contributions and group interactions, Fluid Phase Equilib 269 (1) (2008) 117–133, https://doi.org/10.1016/j.fluid.2008.04.020.

[23] B. Moller, J. Rarey, D. Ramjugernath, Estimation of the vapour pressure of non-electrolyte organic compounds via group contributions and group interactions, J. Mol. Liq. 143 (1) (2008) 52–63, https://doi.org/10.1016/j.molliq.2008.04.020.

[24] K. Tochigi, M. Yamagishi, S. Ando, H. Matsuda, K. Kurihara, Prediction of Antoine constants using a group contribution method, Fluid Phase Equilib 297 (2) (2010) 200–204, https://doi.org/10.1016/j.fluid.2010.05.011.

[25] T. Jensen, A. Fredenslund, P. Rasmussen, Pure-component vapor pressures using UNIFAC group contribution, Ind. Eng. Chem. Fundam. 20 (3) (Aug. 1981) 239–246, https://doi.org/10.1021/i100003a010.

[26] O. Ben Yair, A. Fredenslund, Extension of the UNIFAC group-contribution method for the prediction of pure-component vapor pressures, Ind. Eng. Chem. Process Des. Dev. 22 (3) (Jul. 1983) 433–436, https://doi.org/10.1021/i200022a015.

[27] A.B. Macknick, J.M. Prausnitz, Vapor pressures of heavy liquid hydrocarbons by a group-contribution method, Ind. Eng. Chem. Fundam. 18 (4) (Nov. 1979) 348–351, https://doi.org/10.1021/i160072a010.

[28] D.R. Edwards, J.M. Prausnitz, Estimation of vapor pressures of heavy liquid hydrocarbons containing nitrogen or sulfur by a group-contribution method, Ind. Eng. Chem. Fundam. 20 (3) (Aug. 1981) 280–283, https://doi.org/10.1021/i100003a016.

[29] C.H. Tu, Group-contribution method for the estimation of vapor pressures, Fluid Phase Equilib 99 (1994) 105–120, https://doi.org/10.1016/0378-3812(94)80025-1.

[30] P. Coutsikos, E. Voutsas, K. Magoulas, D.P. Tassios, Prediction of vapor pressures of solid organic compounds with a group-contribution method, Fluid Phase Equilib 207 (1) (2003) 263–281, https://doi.org/10.1016/S0378-3812(03)00029-3.

[31] M. Capouet, J.F. Müller, A group contribution method for estimating the vapour pressures of α-pinene oxidation products, Atmos. Chem. Phys. 6 (6) (2006) 1455–1467, https://doi.org/10.5194/acp-6-1455-2006.

[32] J.F. Pankow, W.E. Asher, SIMPOL.1: a simple group contribution method for predicting vapor pressures and enthalpies of vaporization of multifunctional organic compounds, Atmos. Chem. Phys. 8 (10) (2008) 2773–2796, https://doi.org/10.5194/acp-8-2773-2008.

[33] S. Compernolle, K. Ceulemans, J.F. Müller, EVAPORATION: a new vapour pressure estimation methodfor organic molecules including non-additivity and intramolecular interactions, Atmos. Chem. Phys. 11 (18) (2011) 9431–9450, https://doi.org/10.5194/acp-11-9431-2011.

[34] T.Y. Wang, X.Z. Meng, M. Jia, X.C. Song, Predicting the vapor pressure of fatty acid esters in biodiesel by group contribution method, Fuel Process. Technol. 131 (2015) 223–229, https://doi.org/10.1016/j.fuproc.2014.11.030.

[35] S. O'Meara, A.M. Booth, M.H. Barley, D. Topping, G. McFiggans, An assessment of vapour pressure estimation methods, Phys. Chem. Chem. Phys. 16 (36) (2014) 19453–19469, https://doi.org/10.1039/C4CP00857J.

[36] L. Constantinou, R. Gani, J.P. O'Connell, Estimation of the acentric factor and the liquid molar volume at 298 K using a new group contribution method, Fluid Phase Equilib 103 (1) (1995) 11–22, https://doi.org/10.1016/0378-3812(94)02593-P.

[37] J. Marrero, R. Gani, Group-contribution based estimation of pure component properties, Fluid Phase Equilib 183-184 (2001) 183–208, https://doi.org/10.1016/S0378-3812(01)00431-9.

[38] J. Gmehling, U. Onken, W. Arlt, Vapor-liquid equilibrium data collection, I–XV, John Wiley & Sons, Ltd, 2014.

[39] F.B. Delforouz, M.R. Khorsand Movaghar, S. Shariaty, New empirical correlations for predicting Minimum Miscibility Pressure (MMP) during CO2 injection; implementing the group method of data handling (GMDH) algorithm and Pitzer's acentric factor, Oil Gas Sci. Technol. – Rev. IFP Energies Nouv 74 (2019), https://doi.org/10.2516/ogst/2019035 [Online]. Available:.

[40] N. Shamsundar, J.H. Lienhard, Saturation and metastable properties of the van der waals fluid, Can. J. Chem. Eng. 61 (6) (Dec. 1983) 876–880, https://doi.org/10.1002/cjce.5450610617.

[41] F. Berger, C. Flamm, P.M. Gleiss, J. Leydold, P.F. Stadler, Counterexamples in chemical ring perception, J. Chem. Inf. Comput. Sci. 44 (2) (Mar. 2004) 323–331, https://doi.org/10.1021/ci030405d.

[42] M. Plotkin, Mathematical basis of ring-finding algorithms in CIDS, J. Chem. Doc. 11 (1) (Feb. 1971) 60–63, https://doi.org/10.1021/c160040a013.

[43] K. Goetzke, H.J. Klein, Properties and efficient algorithmic determination of different classes of rings in finite and infinite polyhedral networks, J. Non. Cryst. Solids 127 (2) (1991) 215–220, https://doi.org/10.1016/0022-3093(91)90145-V.

[44] A. Kolodzik, S. Urbaczek, M. Rarey, Unique ring families: a chemically meaningful description of molecular ring topologies, J. Chem. Inf. Model. 52 (8) (Aug. 2012) 2013–2021, https://doi.org/10.1021/ci200629w.

[45] A.M. Patterson, Proposed international rules for numbering organic ring systems, J. Am. Chem. Soc. 47 (2) (Feb. 1925) 543–561, https://doi.org/10.1021/ja01679a040.

[46] E.W. Lemmon, M.O. McLinden, D.G. Friend, NIST chemistry webbook, NIST standard reference database, 2017.

[47] NIST, RefProp 9.1, 2021 [Online]. Available: https://www.nist.gov/srd/refprop.