Journal of King Saud University - Science 36 (2024) 103541

Contents lists available at ScienceDirect

Journal of King Saud University - Science

journal homepage: www.sciencedirect.com

Full Length Article

Optimizing structure-property models of three general graphical indices for thermodynamic properties of benzenoid hydrocarbons ☆

Suha Wazzan [a], Sakander Hayat [b,*], Wafi Ismail [b]

[a] Department of Mathematics, Science Faculty, King Abdulaziz University, Jeddah 21589, Saudi Arabia
[b] Mathematical Sciences, Faculty of Science, Universiti Brunei Darussalam, Jln Tungku Link, Gadong BE1410, Brunei Darussalam

ARTICLE INFO

Dataset link: https://github.com/Sakander/Predictive_Potential_General_Indices.git

MSC:
05C92
05C90
05C09

Keywords:
Mathematical chemistry
QSPR modeling
Discrete optimization
Multivariate regression analysis
Benzenoid hydrocarbon
Thermodynamic property

ABSTRACT

Cheminformatics is an interdisciplinary field that combines principles of chemistry, computer science, and information technology to process, store, analyze, and interpret chemical data. One area of cheminformatics is quantitative structure–property relationship (QSPR) modeling which is a computational approach that correlates the structural attributes of chemical compounds with their physical, chemical, or biological properties to predict the behavior and characteristics of new or untested compounds. Structure descriptors deliver contemporary mathematical tools required for QSPR modeling. One of a significant class of such descriptors is graph-based descriptors known as graphical descriptors. A degree-based graphical descriptor/invariant of a v-vertex graph $\Omega = (V_\Omega, E_\Omega)$ has a general structure $GD_d = \sum_{ij\in E_\Omega} \pi\left(\deg_{x_i}, \deg_{x_j}\right)$, where π is bivariate symmetric map, and $\deg_{x_i}$ is the degree of vertex $x_i \in V_\Omega$. For $\alpha \in \mathbb{R} \setminus \{0\}$, if $\pi = (\deg_{x_i} \times \deg_{x_j})^\alpha$ (resp. $\pi = (\deg_{x_i} + \deg_{x_j})^\alpha$), then GD_d is called the general product-connectivity PC_α (resp. sum-connectivity SC_α) index of Ω. Moreover, the general Sombor index SO_α has the structure $\pi = (\deg^2_{x_i} \times \deg^2_{x_j})^\alpha$. By choosing the heat capacity ΔH and the entropy E as representatives of thermodynamic properties, we in this paper find optimal value(s) of α which deliver the strongest potential of the predictors $GD_d \in \{PC_\alpha, SC_\alpha, SO_\alpha\}$ for predicting ΔH and E of benzenoid hydrocarbons. In order to achieve this, we employ tools such as discrete optimization and multivariate regression analysis. This, in turn, study completely solves two open problems proposed in the literature.

1. Introduction

Cheminformatics employs quantitative structure–property relationship (QSPR) studies (Katritzky et al., 2001) in order to estimate various thermodynamical and physicochemical characteristics of molecular compounds especially, organic structures. QSPR modeling utilizes contemporary mathematical and computational tools (Basak and Mills, 2001) in order to predict these properties. The historical root of this chemical modeling dates back to the pioneering of Wiener (1947) which provides the notion of a path number (the sum of pairwise distance) in estimation of boiling point of alkanes. Later, researchers named this invariant the Wiener index of graphs. Structure-based molecular descriptors (Gutman and Furtula, 2010) provide the contemporary mathematical tools required for QSPR modeling. Graph-related molecular descriptors also known as graphical invariants or topological indices (Balaban et al., 1983) deliver one of the extensively studied family of descriptors. Graphical invariants take up hydrogen-disregarded chemical structure (also known as a molecular/chemical graph) as input and transform it into a non-zero mathematical real number. These molecular graphs (Gutman and Polansky, 1986) are generated by constructing a correspondence between edges (resp. vertices) and bonds (resp. atoms). In order to effectively estimate a given physicochemical property like heat of formation (Allison and Burgess, 2015) and boiling point, graphical invariants propose a regression equation (Diudea et al., 2001) incorporating underlying chemical information of a compound by characterizing its structure. Wazzan and Ahmed (2024b) employed eccentric neighborhood forgotten indices for prediction of boiling point. Moreover, domination-based (Wazzan and Ahmed, 2024a) (resp. symmetry-adapted domination-based (Wazzan

☆ **Financial Disclosure:** This project was funded by the KAU Endowment (WAQF) at King Abdulaziz University, Jeddah, under grant no. (WAQF: 56-247-2024). The authors Sacknowledge with thanks WAQF and Deanship of Scientific Research (DSR) for technical and financial support. The authors acknowledge Prof. Nuha Wazzan from Chemistry department at King Abdulaziz University for her contribution with the DFT calculations and King Abdulaziz Universityís HighPerformance Computing Centre (Aziz Supercomputer) (http://hpc.kau.edu.sa) for supporting the computation for the work described in this paper.

* Corresponding author.
E-mail addresses: sakander1566@gmail.com, sakander.hayat@ubd.edu.bn (S. Hayat).

https://doi.org/10.1016/j.jksus.2024.103541
Received 21 August 2024; Received in revised form 11 November 2024; Accepted 12 November 2024
Available online 19 November 2024

and Ahmed, 2023)) topological indices were employed for their role in QSPR studies of isomeric octanes.

A graphical invariant could be degree-related (Gutman, 2013) (based on vertices' degrees), distance-based (Xu et al., 2014) (defined on distances), spectral (Consonni and Todeschini, 2008) (basing on eigenvalues of graphical matrices) and counting-related (Hosoya, 1988) polynomial and invariants (obtained by counting certain substructures). New graphical invariants are being introduced (Todeschini and Consonni, 2009) every passing day and sometimes without delivering a significant chemical applicability (Gutman and Furtula, 2010). To encounter the proliferation of these invariants, a firm criterion must be adopted in putting forwarding new descriptors. It is unfortunate that frequently these insignificant molecular descriptors are graphical. Gutman and Tošović (2013) used a mild phrase asserting that not following a firm criterion result in proliferation of these invariants and currently there are a lot more graphical invariant than there should be. These facts deliver a strong motivation for considering new emerging families of graphical descriptors to test their quality in structure–property modeling to put forward efficient descriptors, while singling out inefficient ones.

One of the contemporary research topics in mathematical chemistry nowadays is to consider a family of graphical invariants and conduct a comparative testing for predicting physicochemial/theromodynamical properties. The study was initiated Gutman and Tošović (2013) who considered commonly occurring degree-related graphical invariants for estimating physicochemical characteristics of octanes' isomers and showed that the augmented Zagreb index (AZI) is the only degree-based invariant which qualifies to be considered for QSPR modeling. The study was extended to the thermodynamic properties (by opting the heat capacity ΔH and entropy E as their representatives) of benzenoid hydrocarbons (BHs) by Hayat et al. (2023). Note that they selected the lower 30 initial member of BHs as test molecules of the study. Hayat et al. (2024) further extended the similar study to temperature-based graphical descriptors. For structure–property modeling of lead sulphide, we refer to Lal et al. (2024b). Computational results on graph entropies and degree-based graphical indices are reported in Lal et al. (2024a). Other topics such as vertex-edge resolvability and face index of chemical structure are investigated in Negi and Bhat (2024) and Sharma et al. (2024). Applications of degree-based indices in fuzzy graphs are studied in Islam et al. (2024), Islam and Pal (2021, 2024a) and Islam and Pal (2024b).

In existing studied by Gutman and Tošović (2013) and later by Hayat et al. (2023), the general sum-connectivity SC_α index and the general product-connectivity PC_α were considered only for test values $\alpha \in \{\pm 1, \pm\frac{1}{2}, \pm 2, \pm 3\}$. Since Gutman and Tošović (2013) conducted their comparative testing for physicochemical properties, their results are irrelevant to the current study. However, Hayat et al. (2023) conducted their testing for thermodynamic properties of BHs and concluded that SC_α with $\alpha = -3$ and PC_α with $\alpha = -1, -\frac{1}{2}$ are the best three degree-based invariants for predicting thermodynamic properties of BHs. Thus, they concluded their study by naturally asking the following two questions:

Problem 1.1. Find the optimal value(s) of $\alpha \in \mathbb{R} \setminus \{0\}$ for which the correlation value between $\Delta H, E$ and SC_α for the lower 30 BHs is the strongest.

Problem 1.2. Find the optimal value(s) of $\alpha \in \mathbb{R} \setminus \{0\}$ for which the correlation value between $\Delta H, E$ and PC_α for the lower 30 BHs is the strongest.

This paper intends to employ discrete optimization and multivariate regression analysis to answer both of the above problems. In addition, we also study the above two problems for the general Sombor index.

2. Preliminaries

A graph Ω is a pair (V_Ω, E_Ω) in which V_Ω is the vertex set and $E_\Omega \subseteq \binom{V_\Omega}{2}$ is the edge set. The valency/degree $\deg_x$ of a vertex $x \in V_\Omega$ is defined as $\deg_x = |\{z \in V_\Omega : xz \in E_\Omega\}|$. A degree-based graphical descriptor/invariant of a υ-vertex graph $\Omega = (V_\Omega, E_\Omega)$ has a general structure:

$$GD_d = \sum_{ij \in E_\Omega} \pi\left(\deg_{x_i}, \deg_{x_j}\right), \tag{2.1}$$

where π is bivariate symmetric map (i.e. $\pi(x,y) = \pi(y,x)$), and $\deg_{x_i}$ is the degree of vertex $i \in V_\Omega$.

Having $\pi(\deg_{x_i}, \deg_{x_j}) = \frac{1}{\sqrt{\deg_{x_i} \times \deg_{x_j}}}$, the product-connectivity index was proposed by Randić (1975). It has been known as one of earliest degree-based index. It is defined as:

$$PC(\Omega) = \sum_{ij \in E_\Omega} \frac{1}{\sqrt{\deg_{x_i} \times \deg_{x_j}}}. \tag{2.2}$$

Independent of its connection to the product-connectivity index, Bollobás and Erdös (1998) delivered the generalized version of PC index.

$$PC_\alpha(\Omega) = \sum_{ij \in E_\Omega} \left(\deg_{x_i} \times \deg_{x_j}\right)^\alpha, \tag{2.3}$$

where $\alpha \in \mathbb{R} \setminus \{0\}$. One can observe that $PC_{-\frac{1}{2}}(\Omega) = PC(\Omega)$, for an arbitrary graph Ω.

The additive version of PC index called the sum-connectivity SC index was proposed by Zhou and Trinajstić (2009) in 2009. Mathematically, it has $\pi(\deg_{x_i}, \deg_{x_j}) = \frac{1}{\sqrt{\deg_{x_i} + \deg_{x_j}}}$.

$$SC(\Omega) = \sum_{ij \in E_\Omega} \frac{1}{\sqrt{\deg_{x_i} + \deg_{x_j}}}. \tag{2.4}$$

Diverse applicability of SC index across different disciplines motivated (Zhou and Trinajstić, 2010) to introduce the generalized version of the sum-connectivity index symbolized as SC_α, where $\alpha \in \mathbb{R} \setminus \{0\}$. Thus, we have $GD_d = SC_\alpha$, if $\pi(\deg_{x_i}, \deg_{x_j}) = \left(\deg_{x_i} + \deg_{x_j}\right)^\alpha$.

$$SC_\alpha(\Omega) = \sum_{ij \in E_\Omega} \left(\deg_{x_i} + \deg_{x_j}\right)^\alpha. \tag{2.5}$$

Notice that $SC_{-\frac{1}{2}}(\Omega) = SC(\Omega)$, for an arbitrary graph Ω.

By considering $\pi(\deg_{x_i}, \deg_{x_j}) = \frac{1}{\sqrt{\deg_{x_i}^2 + \deg_{x_j}^2}}$, Gutman (2021) recently put forwarded another degree-based graphical descriptor known as the Sombor SO index.

$$SO(\Omega) = \sum_{ij \in E_\Omega} \frac{1}{\sqrt{\deg_{x_i}^2 + \deg_{x_j}^2}}. \tag{2.6}$$

There has been numerously papers published on the mathematical properties as well as chemical applicability of the Sombor index. This delivered motivation to Phanjoubam et al. (2023) to consider the generalized version of the Sombor index by considering $\pi(\deg_{x_i}, \deg_{x_j}) = \left(\deg_{x_i}^2 + \deg_{x_j}^2\right)^\alpha$ in the standard formula of GD_d.

$$SO_\alpha(\Omega) = \sum_{ij \in E_\Omega} \left(\deg_{x_i}^2 + \deg_{x_j}^2\right)^\alpha, \tag{2.7}$$

where $\alpha \in \mathbb{R} \setminus \{0\}$. One can notice that $SO_{-\frac{1}{2}}(\Omega) = SO(\Omega)$, giving the name "general" to this version of the Sombor index.

Discrete optimization is a branch of optimization in applied mathematics and operations research that deals with finding the best solution from a finite or countable set of possible solutions. Unlike continuous optimization, where variables can take any value within a range, discrete optimization restricts variables to discrete values, often integers or elements from a specific set.

A general discrete optimization problem can be formulated as follows:

$\max/\min \ f(x)$

subject to

$x \in S \subset \mathbb{Z}^n$ or $x \in S \subset \{0,1\}^n$,

where:

- $f(x) : S \to \mathbb{R}$ is the objective function, which we aim to maximize or minimize.
- $x = (x_1, x_2, \ldots, x_n)$ is a vector of decision variables.
- $S \subset \mathbb{Z}^n$ (or, sometimes $x \in S \subset \{0,1\}^n$) represents the feasible set defined by constraints, which limits x to discrete values, such as integers or binary values.

Multivariate regression analysis is a statistical technique used to model the relationship between multiple independent (predictor) variables and multiple dependent (response) variables. Unlike simple or multiple regression, which typically models a single dependent variable, multivariate regression allows for multiple outcomes to be analyzed simultaneously, capturing any correlations among them.

Let:

- $Y \in \mathbb{R}^{n\times m}$: the matrix of dependent variables, where n is the number of observations (samples) and m is the number of dependent variables.
- $X \in \mathbb{R}^{n\times p}$: the matrix of independent variables, where p is the number of independent variables.
- $B \in \mathbb{R}^{p\times m}$: the matrix of regression coefficients, with each column B_j representing the coefficients for the jth dependent variable.
- $E \in \mathbb{R}^{n\times m}$: the matrix of error terms or residuals.

The model for multivariate regression can be written as:

$Y = XB + E$,

where:

- $Y_{i,j}$ represents the ith observation of the jth dependent variable.
- $X_{i,k}$ represents the ith observation of the kth independent variable.
- $B_{k,j}$ represents the effect of the kth independent variable on the jth dependent variable.
- $E_{i,j}$ captures the residuals for each observation and each dependent variable.

3. Materials and methods

Note that a BH generally belongs to the class of benzenoid system (BS). A BS (having BHs as a subclass) is a connected finite graph comprising no cut-vertices having internal faces encompassed by regular hexagon with unit sides. Fig. 1 delivers a benzenoid system L.

The degree sequence in a graph Ω is $(\deg_{x_1}, \deg_{x_2}, \ldots, \deg_{x_\upsilon})$ having vertex sequencing $x_1 \ldots, x_\upsilon$, $x_i \in V_\Omega$. On the perimeter of L in Fig. 1, there exists different paths with degree-sequence (2,3,3,2), (2,3,2), (2,3,3,3,2), and (2,3,3,3,3,2) called bay, fissure, cove and fjord, respectively. Altogether, they are collectively called inlets. Let

$\upsilon_{ab} = |\{xz \in E_\Omega : \deg_x = a, \deg_z = b\}|$.

Assume a BS comprises υ vertices, η hexagons and τ inlets. Cruz et al. (2013) proved:

Lemma 3.1. *Suppose L is a BS with υ vertices, τ inlets and η hexagons. Then,*

$\upsilon_{33} = 3\eta - \tau - 3, \ \upsilon_{23} = 2\tau, \ \upsilon_{22} = \upsilon - 2\eta - \tau + 2.$

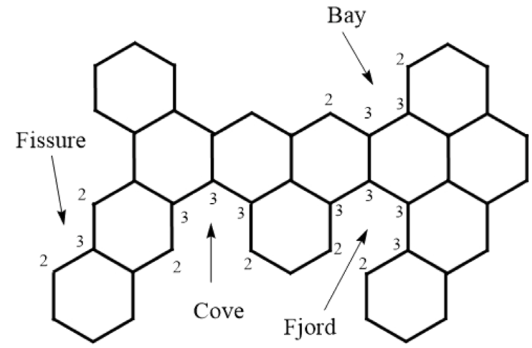

Fig. 1. Instances of a fjord, cove, fissure and a bay in a BS.

Employing Lemma 3.1 on an arbitrary BS comprising υ vertices, τ inlets and η hexagons, one can calculate SC_α, PC_α and SO_α as follows:

$$\begin{aligned} SC_\alpha &= \sum_{ij\in E_\Omega} \left(\deg_{x_i} + \deg_{x_j}\right)^\alpha, \\ &= \upsilon_{33}(3+3)^\alpha + \upsilon_{23}(2+3)^\alpha + \alpha_{22}(2+2)^\alpha, \\ &= 6^\alpha(3\eta - \tau - 3) + 5^\alpha(2\tau) + 4^\alpha(\upsilon - 2\eta - \tau + 2). \end{aligned} \tag{3.8}$$

$$\begin{aligned} PC_\alpha &= \sum_{ij\in E_\Omega} \left(\deg_{x_i} \times \deg_{x_j}\right)^\alpha, \\ &= \upsilon_{33}(3\times 3)^\alpha + \upsilon_{23}(2\times 3)^\alpha + \alpha_{22}(2\times 2)^\alpha, \\ &= 9^\alpha(3\eta - \tau - 3) + 6^\alpha(2\tau) + 4^\alpha(\upsilon - 2\eta - \tau + 2). \end{aligned} \tag{3.9}$$

And, similarly for the general Sombor index, we have:

$$\begin{aligned} SO_\alpha &= \sum_{ij\in E_\Omega} \left(\deg_{x_i}^2 + \deg_{x_j}^2\right)^\alpha, \\ &= \upsilon_{33}(3^2+3^2)^\alpha + \upsilon_{23}(2^2+3^2)^\alpha + \alpha_{22}(2^2+2^2)^\alpha, \\ &= 18^\alpha(3\eta - \tau - 3) + 13^\alpha(2\tau) + 8^\alpha(\upsilon - 2\eta - \tau + 2). \end{aligned} \tag{3.10}$$

In sections that immediately follow, we evaluate SC_α, PC_α and SO_α for the 30 BHs (chosen test molecules) by utilizing Eqs. (3.8), (3.9), and (3.10) respectively.

4. Optimization problem and algorithm

Following Hayat et al. (2023), we consider the heat capacity ΔH and the entropy E to be the representatives of thermodynamic properties of a chemical compound. Moreover, we choose the 30 lower benzenoid hydrocarbons (BHs) as our test molecules. Fig. 2 delivers the lower 30 BHs. Table 1 presents the heat capacity ΔH, the entropy E, the general Randić index R_α, the general sum-connectivity index SCI_α, and the general Sombor index SO_α of 30 lower BHs.

Let $R(\alpha) = R_\alpha(Y, X)$ be the correlation function between $Y \in \{\Delta H, E\}$ and $X \in \{R_\alpha, SCI_\alpha, SO_\alpha\}$. Then, we formulate the following optimization problem:

$$\begin{aligned} \min_\alpha \quad & |R_\alpha(Y, X)| \\ \text{s.t.} \quad & 0 \le |R(\alpha)| \le 1 \\ & \alpha_{\min} < \alpha < \alpha_{\max} \end{aligned} \tag{4.11}$$

Next, we present the pseudo code of corresponding to the above optimization formulation in Algorithm 1.

Note that Algorithm 1 optimizes a correlation function by determining the best value of the parameter α. It constructs a data vector y based on given coefficients and α, then fits a linear model between y and the input data x, calculating the coefficient of determination R^2 as a measure of fit. The objective function is defined to minimize $-\log(1 + R^2)$, aiming to find the optimal α that maximizes correlation. Finally, the algorithm returns the optimal α and the corresponding R^2 value.

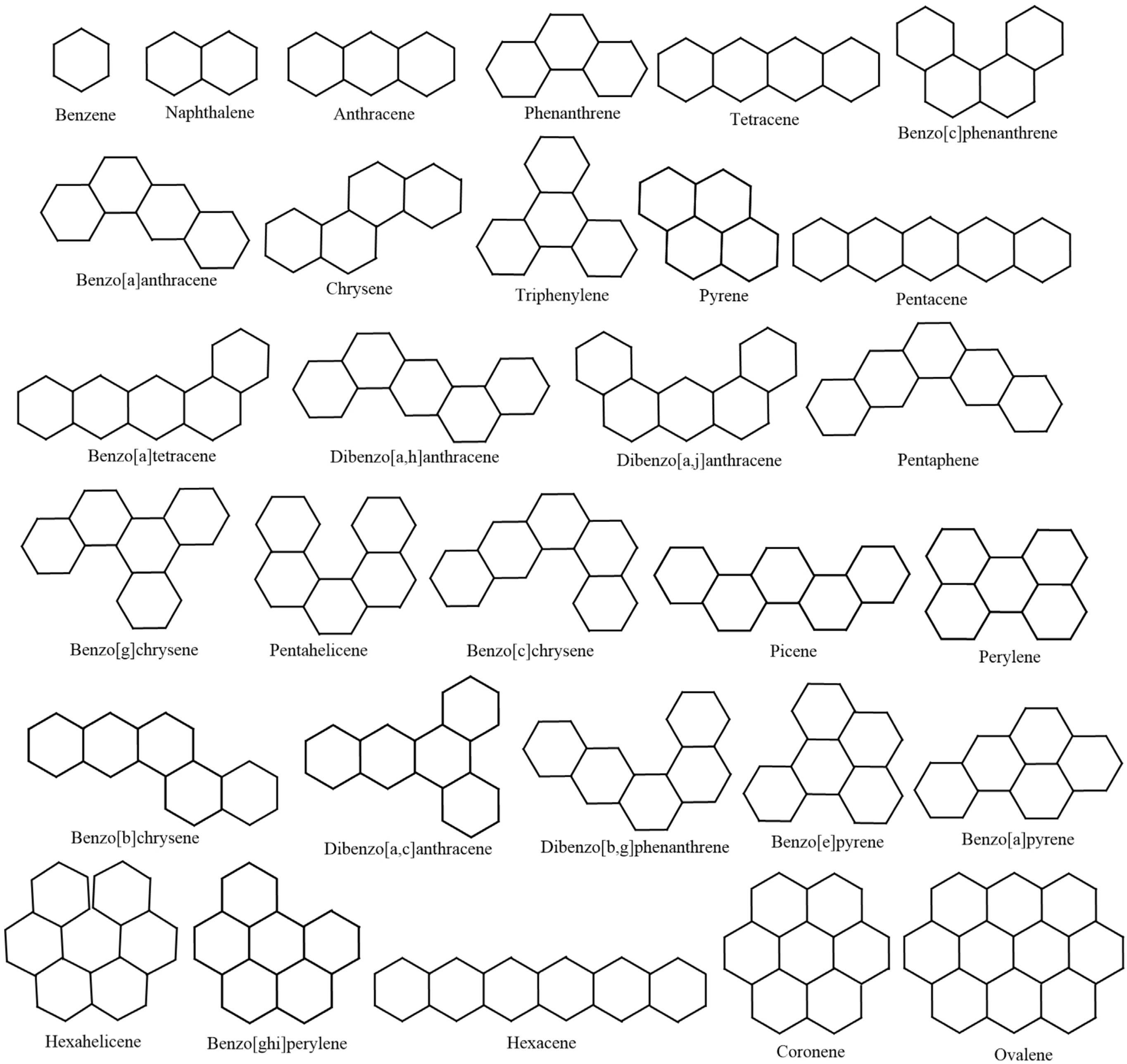

Fig. 2. The 30 lower BHs.

5. Computational results

In this section, a robust linear correlation is established between key molecular attributes — such as the number of atoms, molecular weight, and molecular surface area — and the thermodynamic properties, specifically heat capacity (ΔH) and entropy (E), for the 30 lower benzenoid hydrocarbons. The study demonstrates that as these the molecular features increase, there is a corresponding rise in both ΔH and E, underscoring the predictability of thermodynamic properties based on molecular structure. This foundational insight sets the stage for further exploration of how molecular characteristics influence thermodynamic behavior.

Entropy is the thermodynamic function for predicting the spontaneity of a reaction. Whereas, the *heat capacity of a substance* is defined as the *amount of heat required to raise the temperature of a given quantity of the substance by one degree Celsius.* Several factors can affect the entropy and heat capacities of the substances, including, number of atoms, molecular weight, volume, molecular surface area, boiling point and melting point (Latimer, 1921; Origlia et al., 2001). As the number of atoms in the system increases, regardless of their masses, its entropy and heat capacity values increase. The higher the boiling point and melting point, the larger entropy and heat capacity of the system. In addition, as the volume or the molecular surface area of the compound increases, the entropy and the heat capacity also increase. The entropies (E) and heat capacities (ΔH), the molecular formula (MF), number of atoms (N_{atoms}), molecular weights (MW) and molecular surface area (MSA) of the 30 lower benzenoids are listed in Table 2.

Results obtained show that there are adequate linear correlations between the number of atoms in the molecule (N_{atoms}) versus the heat capacity (Fig. 3(a)) and the entropy (Fig. 3(b)) with R^2 of 0.9994 and 0.9774, respectively. For entropy property, the deviation of the linear correlation is found for $N_{atoms} > 36$. A closer examination of Table 2 reveals that substances with similar molecular formula or number of atoms have almost similar E, and ΔH values. For examples,

Table 1
The molecular structure, heat capacity ΔH, entropy E, general Randić index R_α, general sum-connectivity index SCI_α, and general Sombor index of 30 lower benzenoid hydrocarbons.

Molecule	ΔH	E	R_α	SCI_α	SO_α
Benzene	83.019	269.722	$6\cdot4^\alpha$	$6\cdot4^\alpha$	$6\cdot8^\alpha$
Naphthalene	133.325	334.155	$6\cdot4^\alpha+4\cdot6^\alpha+9^\alpha$	$6\cdot4^\alpha+4\cdot5^\alpha+6^\alpha$	$6\cdot8^\alpha+4\cdot13^\alpha+18^\alpha$
Anthracene	184.194	389.475	$6\cdot4^\alpha+8\cdot6^\alpha+2\cdot9^\alpha$	$6\cdot4^\alpha+8\cdot5^\alpha+2\cdot6^\alpha$	$6\cdot8^\alpha+8\cdot13^\alpha+2\cdot18^\alpha$
Phenanthrene	183.654	395.882	$7\cdot4^\alpha+6\cdot6^\alpha+3\cdot9^\alpha$	$7\cdot4^\alpha+6\cdot5^\alpha+3\cdot6^\alpha$	$7\cdot8^\alpha+6\cdot13^\alpha+3\cdot18^\alpha$
Tetracene	235.165	444.724	$6\cdot4^\alpha+12\cdot6^\alpha+3\cdot9^\alpha$	$6\cdot4^\alpha+12\cdot5^\alpha+3\cdot6^\alpha$	$6\cdot8^\alpha+12\cdot13^\alpha+3\cdot18^\alpha$
Benzo[c]phenanthrene	233.497	447.437	$8\cdot4^\alpha+8\cdot6^\alpha+5\cdot9^\alpha$	$8\cdot4^\alpha+8\cdot5^\alpha+5\cdot6^\alpha$	$8\cdot8^\alpha+8\cdot13^\alpha+5\cdot18^\alpha$
Benzo[a]phenanthrene	234.568	457.958	$7\cdot4^\alpha+10\cdot6^\alpha+4\cdot9^\alpha$	$7\cdot4^\alpha+10\cdot5^\alpha+4\cdot6^\alpha$	$7\cdot8^\alpha+10\cdot13^\alpha+4\cdot18^\alpha$
Chrysene	234.638	455.839	$8\cdot4^\alpha+8\cdot6^\alpha+5\cdot9^\alpha$	$8\cdot4^\alpha+8\cdot5^\alpha+5\cdot6^\alpha$	$8\cdot8^\alpha+8\cdot13^\alpha+5\cdot18^\alpha$
Triphenylene	233.558	450.418	$9\cdot4^\alpha+6\cdot6^\alpha+6\cdot9^\alpha$	$9\cdot4^\alpha+6\cdot5^\alpha+6\cdot6^\alpha$	$9\cdot8^\alpha+6\cdot13^\alpha+6\cdot18^\alpha$
Pyrene	200.815	399.491	$6\cdot4^\alpha+8\cdot6^\alpha+5\cdot9^\alpha$	$6\cdot4^\alpha+8\cdot5^\alpha+5\cdot6^\alpha$	$6\cdot8^\alpha+8\cdot13^\alpha+5\cdot18^\alpha$
Pentacene	286.182	499.831	$6\cdot4^\alpha+16\cdot6^\alpha+4\cdot9^\alpha$	$6\cdot4^\alpha+16\cdot5^\alpha+4\cdot6^\alpha$	$6\cdot8^\alpha+16\cdot13^\alpha+4\cdot18^\alpha$
Benzo[a]tetracene	285.056	513.857	$7\cdot4^\alpha+14\cdot6^\alpha+5\cdot9^\alpha$	$7\cdot4^\alpha+14\cdot5^\alpha+5\cdot6^\alpha$	$7\cdot8^\alpha+14\cdot13^\alpha+5\cdot18^\alpha$
Dibenzo[a,h]anthracene	284.037	508.537	$8\cdot4^\alpha+12\cdot6^\alpha+6\cdot9^\alpha$	$8\cdot4^\alpha+12\cdot5^\alpha+6\cdot6^\alpha$	$8\cdot8^\alpha+12\cdot13^\alpha+6\cdot18^\alpha$
Dibenzo[a,j]anthracene	284.088	507.395	$8\cdot4^\alpha+12\cdot6^\alpha+6\cdot9^\alpha$	$8\cdot4^\alpha+12\cdot5^\alpha+6\cdot6^\alpha$	$8\cdot8^\alpha+12\cdot13^\alpha+6\cdot18^\alpha$
Pentaphene	285.148	506.076	$7\cdot4^\alpha+14\cdot6^\alpha+5\cdot9^\alpha$	$7\cdot4^\alpha+14\cdot5^\alpha+5\cdot6^\alpha$	$7\cdot8^\alpha+14\cdot13^\alpha+5\cdot18^\alpha$
Benzo[g]chrysene	284.595	512.523	$10\cdot4^\alpha+8\cdot6^\alpha+8\cdot9^\alpha$	$10\cdot4^\alpha+8\cdot5^\alpha+8\cdot6^\alpha$	$10\cdot8^\alpha+8\cdot13^\alpha+8\cdot18^\alpha$
Pentahelicene	284.870	500.734	$9\cdot4^\alpha+10\cdot6^\alpha+7\cdot9^\alpha$	$9\cdot4^\alpha+10\cdot5^\alpha+7\cdot6^\alpha$	$9\cdot8^\alpha+10\cdot13^\alpha+7\cdot18^\alpha$
Benzo[c]chrysene	284.503	510.307	$9\cdot4^\alpha+10\cdot6^\alpha+7\cdot9^\alpha$	$9\cdot4^\alpha+10\cdot5^\alpha+7\cdot6^\alpha$	$9\cdot8^\alpha+10\cdot13^\alpha+7\cdot18^\alpha$
Picene	284.785	509.210	$9\cdot4^\alpha+10\cdot6^\alpha+7\cdot9^\alpha$	$9\cdot4^\alpha+10\cdot5^\alpha+7\cdot6^\alpha$	$9\cdot8^\alpha+10\cdot13^\alpha+7\cdot18^\alpha$
Benzo[b]chrysene	284.740	513.879	$8\cdot4^\alpha+12\cdot6^\alpha+6\cdot9^\alpha$	$8\cdot4^\alpha+12\cdot5^\alpha+6\cdot6^\alpha$	$8\cdot8^\alpha+12\cdot13^\alpha+6\cdot18^\alpha$
Dibenzo[a,c]anthracene	284.233	511.770	$9\cdot4^\alpha+10\cdot6^\alpha+7\cdot9^\alpha$	$9\cdot4^\alpha+10\cdot5^\alpha+7\cdot6^\alpha$	$9\cdot8^\alpha+10\cdot13^\alpha+7\cdot18^\alpha$
Dibenzo[b,g]phenanthrene	284.552	509.611	$8\cdot4^\alpha+12\cdot6^\alpha+6\cdot9^\alpha$	$8\cdot4^\alpha+12\cdot5^\alpha+6\cdot6^\alpha$	$8\cdot8^\alpha+12\cdot13^\alpha+6\cdot18^\alpha$
Perylene	251.175	461.545	$8\cdot4^\alpha+8\cdot6^\alpha+8\cdot9^\alpha$	$8\cdot4^\alpha+8\cdot5^\alpha+8\cdot6^\alpha$	$8\cdot8^\alpha+8\cdot13^\alpha+8\cdot18^\alpha$
Benzo[e]pyrene	250.568	463.738	$8\cdot4^\alpha+8\cdot6^\alpha+8\cdot9^\alpha$	$8\cdot4^\alpha+8\cdot5^\alpha+8\cdot6^\alpha$	$8\cdot8^\alpha+8\cdot13^\alpha+8\cdot18^\alpha$
Benzo[a]pyrene	251.973	468.712	$7\cdot4^\alpha+10\cdot6^\alpha+7\cdot9^\alpha$	$7\cdot4^\alpha+10\cdot5^\alpha+7\cdot6^\alpha$	$7\cdot8^\alpha+10\cdot13^\alpha+7\cdot18^\alpha$
Hexahelicene	336.098	555.409	$10\cdot4^\alpha+12\cdot6^\alpha+9\cdot9^\alpha$	$10\cdot4^\alpha+12\cdot5^\alpha+9\cdot6^\alpha$	$10\cdot8^\alpha+12\cdot13^\alpha+9\cdot18^\alpha$
Benzo[ghi]perylene	267.543	472.295	$7\cdot4^\alpha+10\cdot6^\alpha+10\cdot9^\alpha$	$7\cdot4^\alpha+10\cdot5^\alpha+10\cdot6^\alpha$	$7\cdot8^\alpha+10\cdot13^\alpha+10\cdot18^\alpha$
Hexacene	337.204	554.784	$6\cdot4^\alpha+20\cdot6^\alpha+5\cdot9^\alpha$	$6\cdot4^\alpha+20\cdot5^\alpha+5\cdot6^\alpha$	$6\cdot8^\alpha+20\cdot13^\alpha+5\cdot18^\alpha$
Coronene	285.041	468.796	$6\cdot4^\alpha+12\cdot6^\alpha+12\cdot9^\alpha$	$6\cdot4^\alpha+12\cdot5^\alpha+12\cdot6^\alpha$	$6\cdot8^\alpha+12\cdot13^\alpha+12\cdot18^\alpha$
Ovalene	368.518	551.708	$6\cdot4^\alpha+16\cdot6^\alpha+19\cdot9^\alpha$	$6\cdot4^\alpha+16\cdot5^\alpha+19\cdot6^\alpha$	$6\cdot8^\alpha+16\cdot13^\alpha+19\cdot18^\alpha$

Algorithm 1 Optimization of correlation function

```
Input: Coefficients for y ∈ {ΔH, E} values; data x ∈ {R_α, SCI_α, SO_α}
Output: (α̂, R(α̂))

function CalculateY(α)
    Construct data vector y using coefficients and α value
end function

function CalculateRho(α)
    y ← CalculateY(α)
    Fit linear model Mod ← y = β_0 + β_1 x + ε
    R^2 ← coef. of determination from Mod
    return R
end function

function ObjFun(α)
    return −log(1+ CalculateRho(α))
end function

α̂ ← argmax ObjFun(α)
R̂ ← CalculateRho(α̂)
return (α̂, R̂)
```

systems with N_{atoms} = 30, ΔH value (measured in cal/mol.K) lies in the range from 50.938 to 52.44 with an average of 52.125, while E value (measured in cal/mol.K) is ranged from 105.261 – 110.0.37 with an average of 108.055. Additionally, systems with N_{atoms} = 36 have ΔH values range from 63.813 to 64.388 with an average of 64.089, while their E values lie in the range from 115.262 – 123.493 with an average of 120.773.

A closer examination of Table 2 reveals that smaller molecules with lower molecular weights have lower E and ΔH values, while the opposite is true. For example, Table 2 shows that the smallest E of 69.028 and ΔH of 17.151 belong to benzene molecule with molecular weight of 78.048 g/mol. The picture is not similar for larger molecules, Whereas, the maximum E of 84.491 belongs to Ovalene molecule with molecular weight of 398.112 g/mol, while the maximum ΔH value of 133.95 corresponds to Hexacene with molecular weight of 328.128 g/mol. This deviation in the results of obtained for the larger molecule can be clearly viewed by plotting the linear correlation between the MW versus E and ΔH values as shown in Fig. 3. The linear correlation coefficient R^2 of 0.9862 and 0.9298 are belong, correspondingly, to ΔH (Fig. 3(a)) and E (Fig. 3(b)) properties. As can be seen in Fig. 3, the deviation of the linear correlation is clearly observed for larger molecules. In addition, the degree of deviation is greater for E property than for the ΔH one.

Fig. 3 shows good linear correlations between the molecular weight (MW) of the investigated benzenoids and their ΔH and E values. The linear correlations coefficients R^2 of 0.9862 and 0.9298 are belong, correspondingly, to ΔH (Fig. 3(a)) and E (Fig. 3(b)) properties. As can be seen in Fig. 3, the deviation of the linear correlation is clearly observed for larger molecules. In addition, the degree of deviation is greater for E property than for the ΔH one.

Fig. 3 shows good linear correlations between the molecular weight (MW) of the investigated benzenoids and their ΔH and E values.

The entropy and heat capacity of a substance are also correlated with its molecular surface area (MSA); see Fig. 3 for details. Examination of Table 2 illustrates that the smaller the MSA, the smaller E and ΔH values. It is found that benzene molecule with the smallest MSA value of 135.58 $\mathring{A}^2$ has the smallest E and ΔH values of 69.028 and 17.151, respectively. On the other hand, Ovalene with the maximum MSA of 467.46 $\mathring{A}^2$ has the E and ΔH values of 133.467 and 84.491, respectively. Notice that, the maximum E value of 133.95 belongs to Hexacene molecule with MSA of 442.77 $\mathring{A}^2$. An excellent linear correlation is found between the MSA and both E and ΔH properties with R^2 of 0.9744 and 0.9929, respectively. Again, the entropies of the larger systems are deviated from these linear relationships.

Optimized geometries of the thirty investigated aromatic hydrocarbon molecules at the $B3LYP/6\text{−}31G(d)$ level of theory in the gas phase

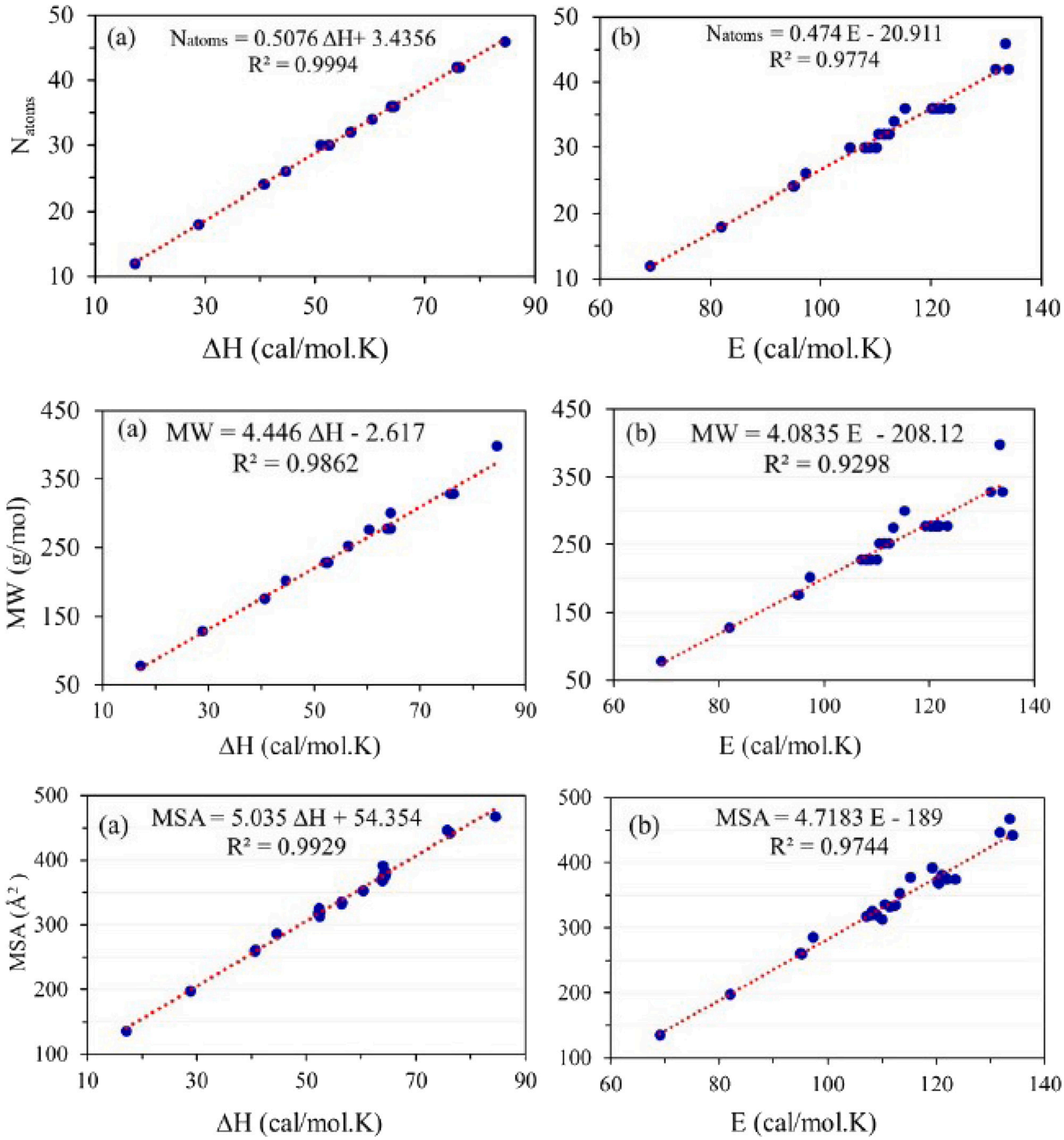

Fig. 3. Correlation curves between N_{atoms}, MW and MSA with the chosen properties.

computed by Gaussian 09 (Frisch, 2009) and visualized by GaussView 05 (Dennington et al., 2007) software packages as implemented in Aziz supercomputer (http://hpc.kau.edu.sa) at King Abdulaziz University's High-Performance Computing Centre.

In general, it can be concluded that the entropies and heat capacities of the 30 lower benzenoids are well correlated with the number of atoms, molecular surface area and molecular weights, however, some deviation from the linear correlation is observed for larger systems.

In conclusion, the analyses in the next two subsections offer complementary approaches to predicting thermodynamic properties. This section establishes strong linear correlations based on physical molecular attributes, such as the number of atoms, molecular weight, and surface area. In contrast, the next two subsections refine this understanding by introducing and optimizing mathematical indices (R_α, SCI_α, and SO_α), which more precisely capture these relationships. Together, these studies enhance our understanding of the influence of molecular structure on thermodynamic properties, effectively bridging the gap between direct physical observations and advanced mathematical modeling through graphical indices.

In this section, we present our computational results by employing Algorithm 1 on different computational platforms such as Octave and R Studio. For results from the next subsection, we employed the computational platform Octave.

Note that, in order to be suited for a linear regression analysis the data is supposed to be tested for normality. There are several tests for normality, including Shapiro–Wilk, Lilliefors, and the tests reported in Jäntschi (2019). Other tests such as Anderson–Darling, Cramér-von Mises, Kolmogorov–Smirnov are reported in Jäntschi (2020). Moreover,

Table 2
Substance, heat capacity, entropy, molecular formula, number of atoms, molecular weights and molecular surface area of the 30 lower benzenoids.

Number	Name	ΔH	E	MF	N_{atoms}	MW	MSA
1	Benzene	17.151	69.028	C_6H_6	12	78.048	135.58
2	Naphthalene	28.816	81.955	$C_{10}H_8$	18	128.064	198.46
3	Anthracene	40.652	94.94	$C_{14}H_{10}$	24	176.064	261.3
4	Phenanthrene	40.561	95.176	$C_{14}H_{10}$	24	176.064	259.85
5	Tetracene	52.516	107.952	$C_{18}H_{12}$	30	228.096	320.22
6	Benzo[c]phenanthrene	50.938	105.261	$C_{18}H_{12}$	30	228.096	318.01
7	Benzo[a]phenanthrene	52.327	108.154	$C_{18}H_{12}$	30	228.096	325.3
8	Chrysene	52.402	108.871	$C_{18}H_{12}$	30	228.096	320.69
9	Triphenylene	52.44	110.037	$C_{18}H_{12}$	30	228.096	313.26
10	Pyrene	44.539	97.259	$C_{16}H_{10}$	26	202.080	286.42
11	Pentacene	64.388	120.96	$C_{22}H_{14}$	36	278.112	381.68
12	Benzo[a]tetracene	64.185	121.281	$C_{22}H_{14}$	36	278.112	379.3
13	Dibenzo[a,h]anthracene	64.079	121.741	$C_{22}H_{14}$	36	278.112	376.8
14	Dibenzo[a,j]anthracene	64.087	121.595	$C_{22}H_{14}$	36	278.112	377.3
15	Pentaphene	64.155	121.385	$C_{22}H_{14}$	36	278.112	379.59
16	Benzo[g]chrysene	63.834	120.473	$C_{22}H_{14}$	36	278.112	368.36
17	Pentahelicene	63.858	120.109	$C_{22}H_{14}$	36	278.112	392.09
18	Benzo[c]chrysene	63.813	120.199	$C_{22}H_{14}$	36	278.112	370.06
19	Picene	64.178	121.984	$C_{22}H_{14}$	36	278.112	374.67
20	Benzo[b]chrysene	64.156	121.463	$C_{22}H_{14}$	36	278.112	377.2
21	Dibenzo[a,c]anthracene	64.21	123.493	$C_{22}H_{14}$	36	278.112	374.76
22	Dibenzo[b,g]phenanthrene	63.858	120.108	$C_{22}H_{14}$	36	278.112	372.71
23	Perylene	56.484	112.373	$C_{20}H_{12}$	32	252.096	334.51
24	Benzo[e]pyrene	56.41	111.423	$C_{20}H_{12}$	32	252.096	332.24
25	Benzo[a]pyrene	56.381	110.484	$C_{20}H_{12}$	32	252.096	336.06
26	Hexahelicene	75.654	131.693	$C_{26}H_{16}$	42	328.128	446.78
27	Benzo[ghi]perylene	60.38	113.229	$C_{22}H_{12}$	34	276.096	353.2
28	Hexacene	76.264	133.95	$C_{26}H_{16}$	42	328.128	442.77
29	Coronene	64.354	115.262	$C_{24}H_{12}$	36	300.096	378.24
30	Ovalene	84.491	133.467	$C_{32}H_{14}$	46	398.112	467.46

it is very important to not have outliers and extreme values, since both may leverage your regression.

5.1. Linear correlation analysis of general graphical indices

The statistical analysis of Figs. 4 and 5 demonstrates the correlation between three general indices (R_α, SCI_α, and SO_α) and the thermodynamic properties of lower benzenoid hydrocarbons, specifically heat capacity (ΔH) and entropy (E). The curves in these figures show how the correlation coefficients vary with the parameter α. These curves have been generated by using the software Octave. Notably, the optimal α values for R_α, SCI_α, and SO_α provide strong correlations with both ΔH and E. For R_α, the optimal value is $\alpha = -1.845$, achieving a correlation coefficient of $\rho = 0.997$ with both ΔH and E. The SCI_α index is optimal at $\alpha = -0.319$, with a corresponding correlation coefficient of $\rho = 0.997$. The SO_α index shows the highest correlation across both properties at $\alpha = -1.067$, with a correlation coefficient of $\rho = 0.998$, indicating its superior predictive potential. These results emphasize the effectiveness of these indices, particularly SO_α, when they are properly optimized for the prediction of thermodynamic properties in benzenoid hydrocarbons.

Figs. 4 and 5 provide a magnified view of the regions around the best values of the parameter α for the general indices R_α, SCI_α, and SO_α when predicting the thermodynamic properties of lower benzenoid hydrocarbons. These figures emphasize the intervals of α where the correlation with the properties is at its peak. Specifically, for predicting heat capacity (ΔH) in Fig. 4, the optimal intervals are approximately $\alpha = [-2, -1.5]$ for R_α, $[-0.5, -0.1]$ for SCI_α, and $[-1.5, -0.5]$ for SO_α. Similarly, for predicting entropy (E) in Fig. 5, the best intervals for these indices are also within these ranges: $R_\alpha = [-2, -1.5]$, $SCI_\alpha = [-0.5, -0.1]$, and $SO_\alpha = [-1.5, -0.5]$.

These intervals represent the ranges of α where each index reaches its highest predictive potential, with the SO_α index particularly standing out due to its consistent performance across both thermodynamic properties. By focusing on these specific intervals, Figs. 4 and 5 underscore the importance of fine-tuning the parameter α to maximize the predictive accuracy of R_α, SCI_α, and SO_α indices for estimating the thermodynamic properties of benzenoid hydrocarbons.

It has been observed that the optimal α intervals for the three general indices, highlighting the regions above the horizontal dashed lines where the correlation coefficient (ρ) is strong. For the general Randić index (R_α), the optimal interval for predicting heat capacity (ΔH) is approximately $[-1.8384, -0.5499]$, while for entropy $[2.5110, 1.3334]$. The general sum-connectivity index (SCI_α) shows strong correlation within the interval $[-3.3914, -1.1480]$ for ΔH and $[-4.4900, -2.7642]$ for E, indicating its effectiveness within these ranges. The general Sombor index (SO_α) demonstrates the broadest and most stable interval of strong correlation, approximately $[-1.5559, -1.4797]$ for both ΔH and E, making it the most versatile and reliable predictor across a wider range of α values compared to the other indices. This analysis underscores the importance of selecting the correct α interval for each index to achieve optimal predictive accuracy for thermodynamic properties.

5.2. Multiple prediction potential of general graphical indices

Note that Section 5 consider the two chosen thermodynamic properties i.e. ΔH and E individually to investigate their prediction potential with $GD_d \in \{SC_\alpha, PC_\alpha, SO_\alpha\}$. This section investigate the same problem with $GD_d \in \{SC_\alpha, PC_\alpha, SO_\alpha\}$ and simultaneously choosing both ΔH and E. In order to perform this study, we employ multivariate regression analysis as we now have more than one independent variables. The multivariate regression analysis has been performed on the statistical environment R Studio.

Let $x_1 = \Delta H$, $x_2 = E$ (resp. $y = PC_\alpha$) the two independent variables (resp. dependent variable). Since there are more than one independent variables are involved, we employ multivariate correlation coefficient to investigate the prediction ability of the general product-connectivity index for predicting ΔH and E. Let $R(\alpha) = \rho(PC_\alpha; \Delta H, E)$ be the multivariate correlation function between PC_α and the two chosen properties ΔH and E. Thus, optimizing $R(\alpha)$ would deliver us the optimal value(s) of α (let us denote that with $\hat{\alpha}$) for which the prediction ability of PC_α and the two test properties ΔH and E is the

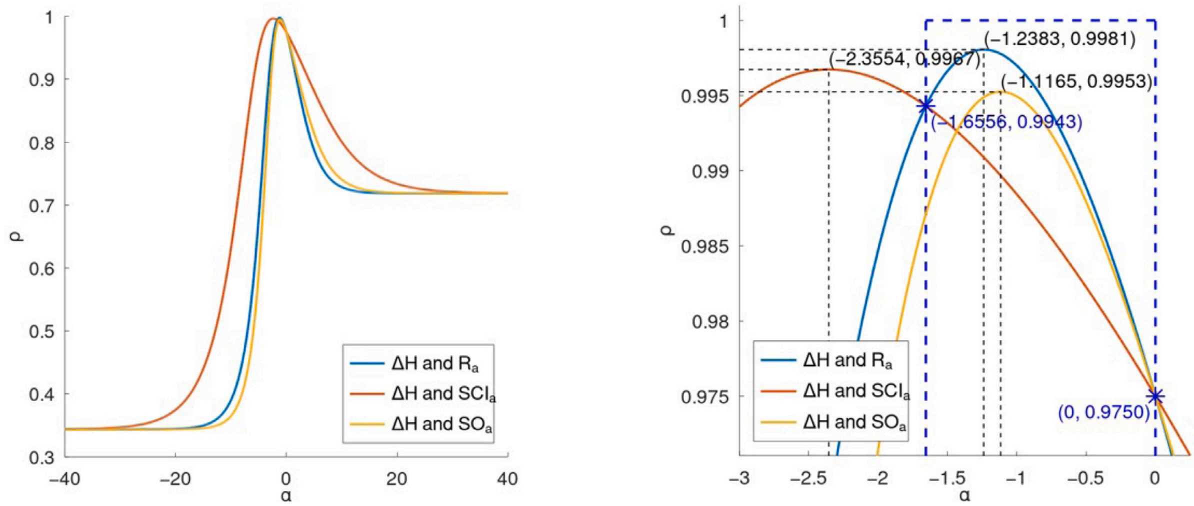

Fig. 4. Far and new views of the correlation curves between general indices and ΔH of lower benzenoids.

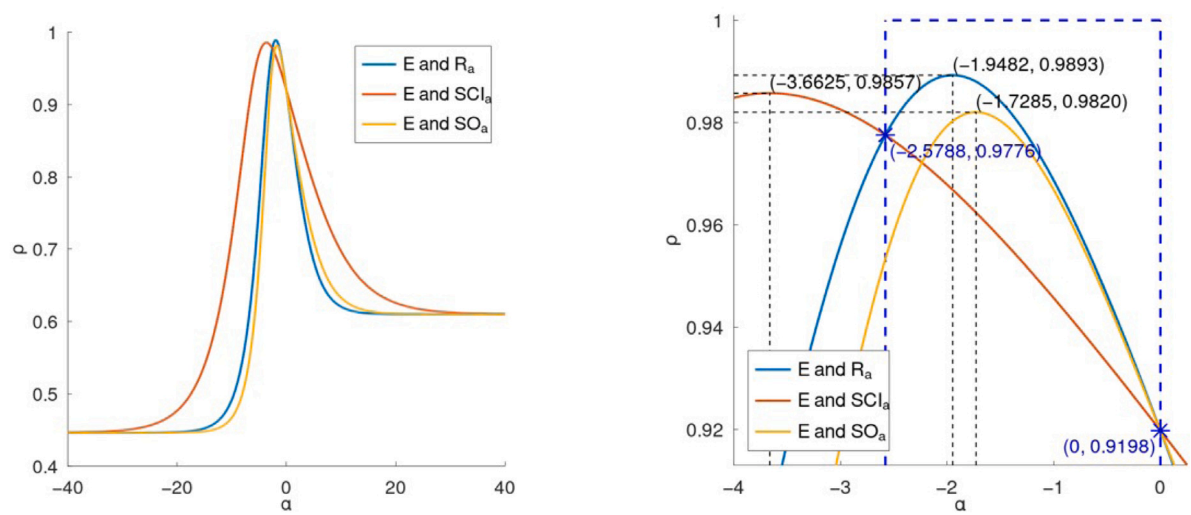

Fig. 5. Far and new views of the correlation curves between general indices and E of lower benzenoids.

strongest. We apply the following same algorithm as we did in Section 5 by replacing the correlation function with the multivariate correlation function.

The main difference between Algorithm 2 and the previous algorithm (Algorithm 1) is the use of multiple independent variables (x_1 and x_2) in the linear model. In Algorithm 1, the linear model is a simple regression with a single predictor variable x, whereas in Algorithm 2, the linear model is a multiple regression with two predictor variables, x_1 and x_2. This modification requires adjusting the calculation of the correlation (specifically R^2), as Algorithm 2 now considers the combined effect of both predictors on y. The objective function and the optimization process remain similar, aiming to find the optimal α that maximizes the correlation in this multivariate context.

A built-in optimizing tool in R Studio language is used by applying Algorithm 2 to generate the required α vs $R(\alpha)$ curves. Fig. 6 depicts such a plot incorporating the bivariate relationship between $R(\alpha)$ and α delivering $\hat{\alpha} = -0.319$ and the corresponding correlation value of $\rho = 0.997$.

Applying the same computational process and Algorithm 2, we obtain Fig. 7 for the general sum-connectivity index SC_α. Multiple correlation curve between $R(\alpha)$ and α show that the optimal value of α is $\alpha_{\max} = -1.845$ and the corresponding correlation value of $\rho = 0.997$.

A similar computational process by employing Algorithm 2 deliver Fig. 8 for the general Sombor index SO_α. Multiple correlation curve

Algorithm 2 Optimization of Multiple Correlation

Input: Coefficients for y values; data x_1, x_2
Output: $(\hat{\alpha}, R(\hat{\alpha}))$

function CALCULATEY(α)
  Construct data vector y using coefficients and α value
end function

function CALCULATEMULTIPLERHO(α)
  $y \leftarrow$ CALCULATEY(α)
  Fit linear model MOD $\leftarrow y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$
  $R^2 \leftarrow$ coef. of determination from MOD
  return R
end function

function OBJFUN(α)
  return $-\log(1+$ CALCULATEMULTIPLERHO$(\alpha))$
end function

$\hat{\alpha} \leftarrow$ argmax OBJFUN(α)
$\hat{R} \leftarrow$ CALCULATEMULTIPLERHO($\hat{\alpha}$)
return $(\hat{\alpha}, \hat{R})$

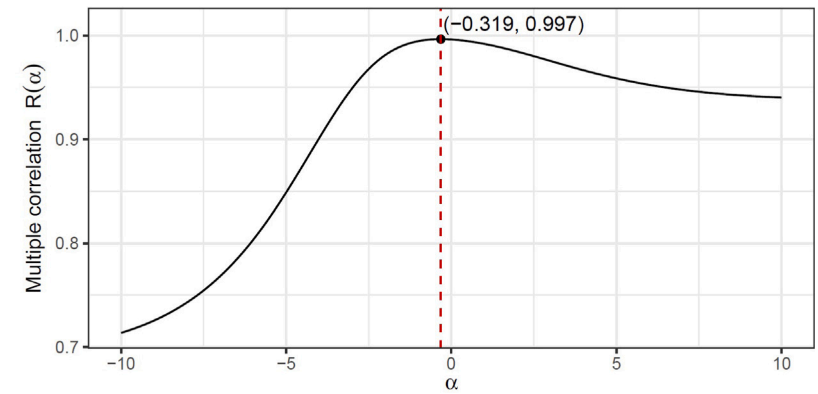

Fig. 6. Multiple correlation curve between $R(\alpha)$ and α delivering $\hat{\alpha} = -0.319$ and the corresponding correlation value of $\rho = 0.997$.

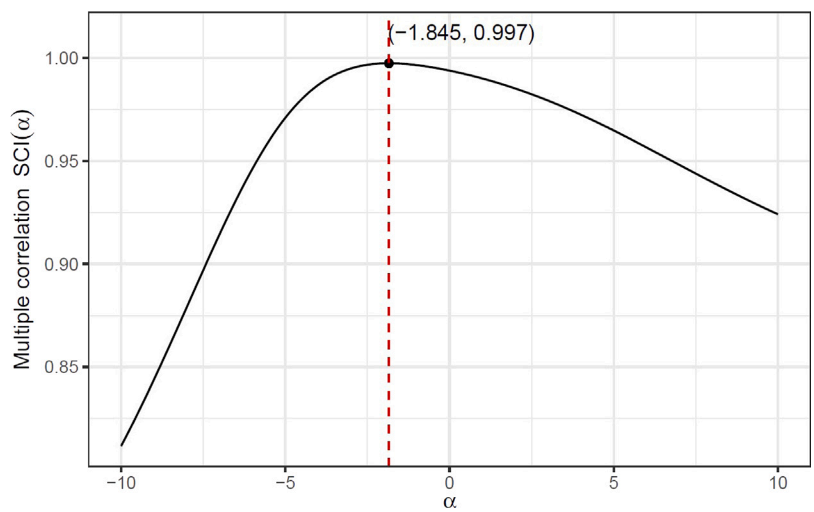

Fig. 7. Multiple correlation curve between $R(\alpha)$ and α delivering $\alpha_{\max} = -1.845$ and the corresponding correlation value of $\rho = 0.997$.

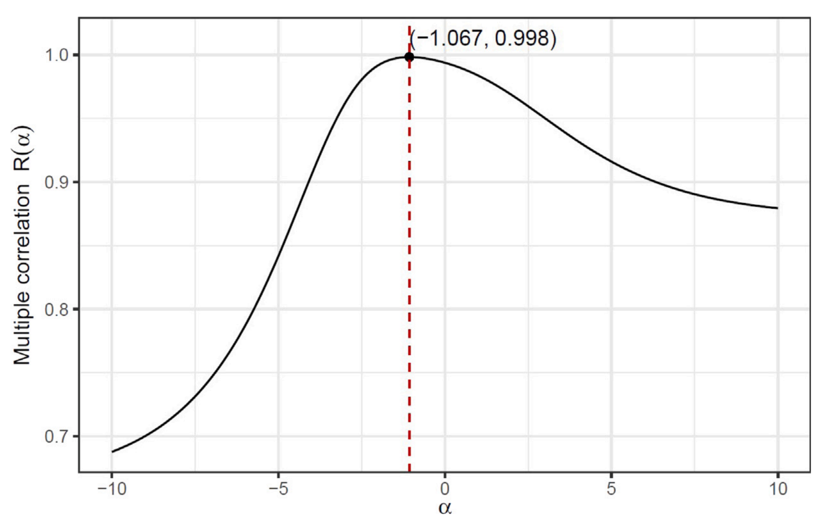

Fig. 8. Multiple correlation curve between $R(\alpha)$ and α delivering $\alpha_{\max} = -1.067$ and the corresponding correlation value of $\rho = 0.998$.

between $R(\alpha)$ and α show that the optimal value of α is $\alpha_{\max} = -1.067$ and the corresponding correlation value of $\rho = 0.998$.

6. Conclusion

Contributions

In this work, we:

- Developed optimal predictive models using three general degree-related indices–general sum/product connectivity and Sombor indices–offering high predictive accuracy for thermodynamic properties of benzenoid hydrocarbons.
- Addressed open problems by determining optimal parameter values of α that maximized correlations between graphical indices and properties such as heat capacity and entropy.

- Validated the effectiveness of each index through discrete optimization and multivariate regression analysis, highlighting the superior performance of the general product-connectivity index over other degree-based indices.

Study implications

As potential study implications, this work:

- Provides a mathematical framework that enhances the use of cheminformatics for predicting thermodynamic properties, supporting the integration of graphical indices in QSPR modeling.

- Establishes that molecular features like the number of atoms, molecular weight, and surface area significantly influence entropy and heat capacity, guiding future molecular property predictions.

- Highlights the general product-connectivity index as particularly effective, suggesting broader applications in chemical graph theory for structure–property modeling.

Limitations

Here we highlight the limitations of this study.

- The study is limited to benzenoid hydrocarbons, restricting generalizability to other classes of chemical compounds.
- Evaluated only a few thermodynamic properties (heat capacity and entropy), which may limit understanding of the indices' predictive potential for other properties such as physicochemical properties.

Future study

Based on the limitations above, here are some possible research directions:

- Extend the application of these indices to a broader range of physicochemical and quantum-theoretic properties.
- Investigate the use of these indices for non-benzenoid and more complex molecular structures to test generalizability.
- Further explore temperature-based graphical indices to enhance predictive capability across various chemical environments and applications.

CRediT authorship contribution statement

Suha Wazzan: Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology. **Sakander Hayat:** Writing – original draft, Methodology, Formal analysis, Conceptualization. **Wafi Ismail:** Writing – original draft, Visualization, Validation, Software, Investigation, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors acknowledge Prof. Nuha Wazzan from Chemistry department at King Abdulaziz University for her contribution with the DFT calculations and King Abdulaziz Universityís HighPerformance Computing Centre (Aziz Supercomputer) (http://hpc.kau.edu.sa) for supporting the computation for the work described in this paper. The authors are grateful to the reviewers and editors for their helpful comments and suggestions which has improved the submitted version of this paper.

Data availability

The data generated related to this study is available in a public repository on GitHub
https://github.com/Sakander/Predictive_Potential_General_Indices.git.

References

Allison, T.C., Burgess, Jr., D.R., 2015. First-principles prediction of enthalpies of formation for polycyclic aromatic hydrocarbons and derivatives. J. Phys. Chem. A. 119, 11329–11365.
Balaban, A.T., Motoc, I., Bonchev, D., Mekenyan, O., 1983. Topological indices for structure–activity corrections. Top. Curr. Chem. 114, 21–55.
Basak, S.C., Mills, D., 2001. Quantitative structure–property relationships (QSPRS) for the estimation of vapor pressure: A hierarchical approach using mathematical structural descriptors. J. Chem. Inf. Comput. Sci. 41 (3), 692–701.
Bollobás, B., Erdös, P., 1998. Graphs of extremal weights. Ars Combin. 50, 225.
Consonni, V., Todeschini, R., 2008. New spectral indices for molecular description. MATCH Commun. Math. Comput. Chem. 60, 3–14.
Cruz, R., Gutman, I., Rada, J., 2013. On benzenoid systems with minimal number of inlets. J. Serb. Chem. Soc. 78 (9), 1351–1357.
Dennington, R.D.I.I., Keith, T., Millam, J., 2007. GaussView, Version 4.1.2. Semichem Inc., Shawnee Mission, KS.
Diudea, M.V., Gutman, I., Jäntschi, L., 2001. Molecular Topology. Nova, Huntington.
Frisch, A., 2009. Gaussian 09 W Reference. Vol. 25, Wallingford, USA, p. 470.
Gutman, I., 2013. Degree-based topological indices. Croat. Chem. Acta. 86, 351–361.
Gutman, I., 2021. Geometric approach to degree-based topological indices: Sombor indices. MATCH Commun. Math. Comput. Chem. 86 (1), 11–16.
Gutman, I., Furtula, B. (Eds.), 2010. Novel Molecular Structure Descriptors – Theory and Applications, Vols. 1 & 2. Univ. Kragujevac, Kragujevac.
Gutman, I., Polansky, O.E., 1986. Mathematical Concepts in Organic Chemistry. Springer-Verlag, New York.
Gutman, I., Tošović, J., 2013. Testing the quality of molecular structure descriptors, vertex-degree-based topological indices. J. Serb. Chem. Soc. 78, 805–810.
Hayat, S., Khan, A., Ali, K., Liu, J.-B., 2024. Structure-property modeling for thermodynamic properties of benzenoid hydrocarbons by temperature-based topological indices. Ain Shams Eng. J. 15 (3), 102586.
Hayat, S., Suhaili, N., Jamil, H., 2023. Statistical significance of valency-based topological descriptors for correlating thermodynamic properties of benzenoid hydrocarbons with applications. Comput. Theor. Chem. 1227, 114259.
Hosoya, H., 1988. On some counting polynomials in chemistry. Discrete Appl. Math. 19, 239–257.
Islam, S.R., Mohsin, B.B., Pal, M., 2024. Hyper-zagreb index in fuzzy environment and its application. Heliyon 10 (16), e36110.
Islam, S.R., Pal, M., 2021. Hyper-Wiener index for fuzzy graph and its application in share market. J. Intell. Fuzzy Systems 41 (1), 2073–2083.
Islam, S.R., Pal, M., 2024a. Multiplicative version of first zagreb index in fuzzy graph and its application in crime analysis. Proc. Natl. Acad. Sci. India A 94 (1), 127–141.
Islam, S.R., Pal, M., 2024b. Neighbourhood and competition graphs under fuzzy incidence graph and its application. Comput. Appl. Math. 43 (7), 411.
Jäntschi, L., 2019. A test detecting the outliers for continuous distributions based on the cumulative distribution function of the data being tested. Symmetry 11 (6), 835.
Jäntschi, L., 2020. Detecting extreme values with order statistics in samples from continuous distributions. Mathematics 8 (2), 216.
Katritzky, A.R., Petrukhin, R., Tatham, D., Basak, S., Benfenati, E., Karelson, M., Maran, U., 2001. Interpretation of quantitative structure- property and- activity relationships. J. Chem. Inf. Comput. Sci. 41 (3), 679–685.
Lal, S., Bhat, V.K., Sharma, S., 2024a. Topological indices and graph entropies for carbon nanotube Y-junctions. J. Math. Chem. 62 (1), 73–108.
Lal, S., Bhat, V.K., Sharma, K., Sharma, S., 2024b. Topological indices of lead sulphide using polynomial technique. Mol. Phys. 122 (3), e2249131.
Latimer, W.M., 1921. The mass effect in th entropy of solids and gases. J. Am. Chem. Soc. 43 (4), 818–826.
Negi, S., Bhat, V.K., 2024. Face index of silicon carbide structures: An alternative approach. Silicon 16, 5865–5876.
Origlia, M., Patterson, B., Woolley, E., 2001. Apparent molar volumes and apparent molar heat capacities of aqueous solutions ofn, N-dimethylformamide and n, N-dimethylacetamide at temperatures from 278.15 to 393.15 K and at the pressure 0.35 MPa. J. Chem. Thermodyn. 33 (8), 917–927.
Phanjoubam, C., Mawiong, S.M., Buhphang, A.M., 2023. On general sombor index of graphs. Asian-Eur. J. Math. 16 (3), 2350052.
Randić, M., 1975. Characterization of molecular branching. J. Am. Chem. Soc. 97, 6609–6615.
Sharma, S.K., Singh, M., Bhat, V.K., 2024. Vertex-edge partition resolvability for certain carbon nanocones. Polycycl. Aromat. Compd. 44 (3), 1745–1759.
Todeschini, R., Consonni, V., 2009. Molecular Descriptors for Chemoinformatics, Vols. 1 & 2. Wiley-VCH, Weinheim, Germany.

Wazzan, S., Ahmed, H., 2023. Symmetry-adapted domination indices: The enhanced domination sigma index and its applications in QSPR studies of octane and its isomers. Symmetry 15 (6), 1202.
Wazzan, S., Ahmed, H., 2024a. Advancing computational insights: Domination topological indices of polysaccharides using special polynomials and QSPR analysis. Contemp. Math. 5 (1), 26–49.
Wazzan, S., Ahmed, H., 2024b. Unveiling novel eccentric neighborhood forgotten indices for graphs and gaph operations: A comprehensive exploration of boiling point prediction. AIMS Math. 9 (1), 1128–1165.
Wiener, H., 1947. Structural determination of the paraffin boiling points. J. Am. Chem. Soc. 69, 17–20.
Xu, K., Liu, M., Das, K.C., Gutman, I., Furtula, B., 2014. A survey on graphs extremal with respect to distance-based topological indices. MATCH Commun. Math. Comput. Chem. 71, 461–508.
Zhou, B., Trinajstić, N., 2009. On a novel connectivity index. J. Math. Chem. 46, 1252–1270.
Zhou, B., Trinajstić, N., 2010. On general sum-connectivity index. J. Math. Chem. 47, 210–218.