CREATIVE MATH. & INF.
19 (2010), No. 2, 135 - 139

Online version available at http://creative-mathematics.ubm.ro/
Print Edition: ISSN 1584 - 286X Online Edition: ISSN 1843 - 441X

Modelling normal boiling points of alkanes by linear regression using the SD index

ZOIŢA BERINDE

ABSTRACT.

Multiple linear regression methods based on the SD index, introduced previously by the author are used to develop quantitative structure-property models for estimating the normal boiling points of alkanes. The performance of the SD index, in this application, is comparable to the performance of the Wiener number and Harary index.

1. INTRODUCTION

In general, there are causal relationships between the molecular structures of organic compounds and their chemical, physical and biological properties [11], [1], [17], [12], [14], [10]. Many of these relationships have been investigated using graph-theoretical descriptors of molecular structures, known as topological indices [13], [15]. The alkanes represent a special attractive class of compounds as a starting point for the application of molecular mass and extent of branching, and because the alkanes are nonpolar, a number of complexities that arise with more polar compounds are avoided [9].

2. THE WEIGHTED ELECTRONIC CONNECTIVITY MATRIX AND THE TOPOLOGICAL INDEX SD

For a molecule having N atoms, whose graphs is $G = (V(G)), E(G))$, the weighted electronic connectivity matrix $CEP(G)$ is a symmetric and quadratic $N \times N$ matrix given by:

$$CEP(G) = \{[CEP]_{ij}\,;\, i,j \in V(G)\}, \tag{2.1}$$

where $V(G)$ is the set of vertices of the molecular graph G, $E(G)$ is the set of edges, and

$$CEP_{ij} = w.e.d._{(i,j)}, \text{ if } i \neq j \text{ and } (i,j) \in E(G) \text{ and } CEP_{ij} = 0, \text{ otherwise,} \tag{2.2}$$

where $w.e.d._{(i,j)}$ denotes the weighted electronic distance between the atoms (vertices) i and j. The concept of weighted electronic distance, w.e.d., was defined in [7] by the following formula:

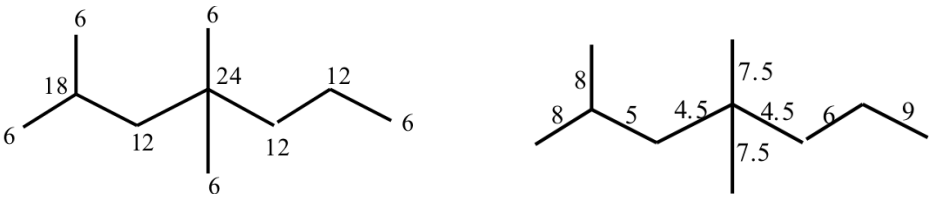
$$w.e.d.(i,j) = \frac{1}{b_r} \cdot \frac{Z_i^{'} + Z_j^{'}}{v_i \cdot v_j}, \tag{2.3}$$

where:
$-b_r$ is the bond weight with values: 1, for single bond, 2 for double bond, 3 for triple bond and 1.5 for aromatic bond, like in Barysz and al. [2];
$-v_i, v_j$ denote the degree of vertices i, and j, respectively;
$-Z_i'$ denotes the formal degree of vertex i, and it is defined by

$$Z_i' = Z_i \cdot v_i, \tag{2.4}$$

where Z_i is the order number of atom i in Mendeleev's periodic system, that is, the number of all electrons in the atom i.

In the graph G.1, representing 2,4,4-trimethylheptane, we illustrated the notion of formal degree, Z', and the weighted electronic distance, d.e.p. We have chosen this structure because it contains all four kinds of carbons: primary, secondary, tertiary and quaternary.



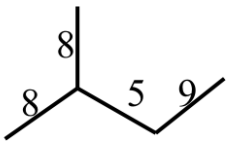
Received: 25.03.2010; In revised form: 20.06.2010; Accepted: 15.08.2010.
2000 *Mathematics Subject Classification.* 92E10, 05C12.
Key words and phrases. *Molecular graph, weighted electronic distance, weighted electronic connectivity matrix, SD topological index, Wiener number, Harary index, QSPR study, boiling points, multiple linear regression.*

Figure 1. The graphs G.1{Z'}, G.2{w.e.d.}

Similar formulas which use the atomic number and the multiplicity order of bonds have been considered in an other context by Barysz and collaborators [2], where it is detected only the presence of multiple bonds and heteroatoms, but their model is not able to reproduce the information related to the neighborhood of bonds, as in the case of weighted electronic connectivity.

The formal degree Z_i' of the vertex i is in fact a local invariant on vertex (LOVI), which replaces the classical degree of the vertex, while the weighted electronic distance, $d.e.p.(i,j)$ represents a local invariant on edge (LOEI).

In Figure 1 the corresponding CEP matrix for the hydrogen-deleted graph (G.2) of 2-methyl-butane is given , while, in Figure 2 the corresponding CEP matrix for the hydrogen-deleted graph G.3 of 2-methyl-butane, is illustrated.



$$CEP(G.3) = \begin{bmatrix} 0 & 8 & 0 & 0 & 0 \\ 8 & 0 & 5 & 0 & 8 \\ 0 & 5 & 0 & 9 & 0 \\ 0 & 0 & 9 & 0 & 0 \\ 0 & 8 & 0 & 0 & 0 \end{bmatrix} \begin{matrix} \text{SEP}_i \\ 8 \\ 21 \\ 14 \\ 9 \\ 8 \end{matrix}$$

Figure 1. The CEP matrix for 2-methyl-butane

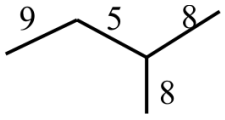
The sum of all entries on the i-th row in CEP (G2) is denoted by SEP_i (and is written at the matrix right hand side):

$$SEP_i = \sum_{j=1}^{n} [CEP]_{ij}\,, \quad i = 1, 2, \ldots, n \tag{2.5}$$

The electronic connectivity matrix can function as a basis for the construction of several new topological indices. The SD index is given by:

$$SD = \sum_{i=1}^{n} \frac{SEP_i}{z_i}. \tag{2.6}$$

The calculation technique of SD index is illustrated in Figure 2 for the hydrogen-suppressed graph G.2 of 2-methylbutane.



$$SD\,(G.2) = \frac{9}{1} + \frac{14}{2} + \frac{21}{3} + 2 \cdot \frac{8}{1} = 39.$$

Figure 2. The calculation of the SD index for 2-methylbutane.

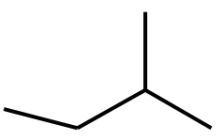
3. A NEW PARAMETER: D_C

We now propose a new parameter, denoted by D_C, which is calculated using by the formula:

$$D_C = N_C - N_P, \tag{3.7}$$

where N_C represents the number of carbon atoms in the molecular graph and N_P represents the number of carbon atoms primary in the molecular graph.

The calculation technique of parameter D_C is illustrated in Figure 3 for the hydrogen-suppressed graph G.2 of 2-methylbutane.



$$D_C = 5 - 3 = 2.$$

Figure 3. The calculation technique of parameter DC for the hydrogen-suppressed graph of 2-methylbutane, G.2.

S_D and D_C could be used together in order to improve the QSPR models for several representative physical properties on several data sets of alkanes by using two-variable linear regression analysis.

4. Results and discussions

Values of SD index were calculated for a set of 44 alkanes, including hexanes, heptanes, octanes and nonanes and are presented in Table 1. Values for the boiling points (bp) were taken from [9].

Regression equation and statistical parameters for a monovariable correlation of boiling points, by using the SD index, are shown below:

$$bp = -106.388 + 3.742\,SD;\ \ N = 44;\ \ r = 0.982; s = 4.52; F = 1128 \tag{4.8}$$

were r represents the correlation coefficient, s the standard deviation and F is the Fisher ratio.

For the same set of 44 alkanes as in [16], the simple linear regression with the SD index leads to a correlation ($r = 0.982$ and $s = 4.52$ °C) which is better than the simple model obtained by Harary [16] ($r = 0.948$ and $s = 7.58$ °C) and much better than the one obtained by using the Wienner number [16], when $r = 0.942$ and $s = 7.99$ °C.

Obviously, a single DC index cannot give a simple accurate correlation. It is important to stress that the statistical parameters of the QSPR equations could be improved by using multidimensional correlations involving also other topological descriptors. In this work we propose a combined use of the SD index and parameter DC.

The regression equation and its statistical parameters for bivariable correlation of boiling points, the SD index and DC of the 44 alkanes considered in the study are depicted below:

$$bp = -104.886 + 3.442\,SD + 4.158\,D_C;\ \ N = 44;\ \ r = 0.991; s = 3.2; F = 1143. \tag{4.9}$$

The values of bp predicted by this equation are shown in Table 1.

Statistics indicate that equation (4.9) represents indeed a very good model for calculating boiling points.

Table 1. Observed and calculated boiling points bp of alkanes correlated with the SD index

N	Alkanes	D_C	SD	bp Observed	bp Calculated	Residuals
1	6	4	45.00	68.74	66.63	2.11
2	2M5	3	45.00	60.27	62.48	-2.21
3	3M5	3	46.00	63.28	65.92	-2.64
4	23MM4	2	45.33	57.98	59.46	-1.48
5	2M6	4	51.00	90.05	87.29	2.76
6	3M6	4	52.00	91.85	90.73	1.12
7	3E5	4	53.00	93.47	94.17	-0.7
8	24MM5	3	51.00	80.50	83.13	-2.63
9	22MM5	3	51.00	79.19	83.13	-3.94
10	23MM5	3	52.33	89.78	87.71	2.07
11	33MM5	3	52.50	86.06	88.29	-2.23
12	223MMM4	2	51.50	80.88	80.69	0.19
13	2M7	5	57.00	117.64	112.10	5.54
14	3M7	5	58.00	118.92	115.54	3.38
15	4M7	5	58.00	117.71	115.54	2.17
16	25MM6	4	57.00	109.10	107.94	1.16
17	3E6	5	58.00	118.53	115.54	2.99
18	24MM6	4	57.00	109.42	107.94	1.48
19	22MM6	4	58.33	106.84	112.52	-5.68
20	23MM6	4	59.00	115.60	114.83	0.77
21	34MM6	4	59.33	117.72	115.96	1.76
22	33MM6	4	58.50	111.97	113.11	-1.14
23	2M3E5	4	59.33	115.65	115.96	-0.31
24	224MMM5	3	57.00	99.24	103.78	-4.54
25	234MMM5	3	58.67	113.47	109.53	3.94
26	223MMM5	3	58.50	109.84	108.94	0.9
27	233MMM5	3	59.00	114.76	110.67	4.09
28	2233MMMM4	2	57.75	106.47	102.21	4.26
29	22MM7	5	63.00	132.69	132.75	-0.06
30	24MM7	5	64.00	133.50	136.19	-2.69
31	25MM7	5	64.00	136.00	136.19	-0.19
32	26MM7	5	63.00	135.21	132.75	2.46
33	34MM7	9	65.33	140.60	140.77	-0.17
34	23ME6	5	65.33	138.00	140.77	-2.77
35	24ME6	5	65.00	133.80	139.63	-5.83
36	223MMM6	4	64.50	133.60	133.75	-0.15
37	224MMM6	4	64.00	126.54	132.03	-5.49
38	225MMM6	4	63.00	124.08	128.59	-4.51
39	233MMM6	4	65.00	137.68	135.48	2.2
40	244MMM6	4	64.5	130.65	133.75	-3.1
41	233MME5	4	66.50	142.00	140.64	1.36
42	234MEM5	4	65.67	136.73	137.781	-1.051
43	2233MMMM5	3	65.25	140.27	132.18	8.09
44	2244MMMM5	3	63.00	122.28	124.43	-2.15

5. Conclusions

Mihalić and Trinajstić [15] proposed that the quality of this type of approach can be conveniently measured by the correlation coefficient and the standard deviation, and therefore, a good QSPR model must have $r > 0.9$, while s could depend on the property under study.

According to this statement, equation (4.9) represents a very good model to describe the boiling points of alkanes.

As a result of the studies carried out in [3]-[8], we can state that the replacement of the common topological distance with the weighted topological distance within the process of representing the chemical structure leads to a viable QSPR model, with significant correlation qualities.

REFERENCES

[1] Balaban, A. T., *Chemical Applications of Graph Theory*, Academic Press, New York, 1976
[2] Barysz M., Jashari G., Lall R. S., Srivastava V. K. and Trinajstić, N., *Chemical Applications of Topology and Graph Theory* (King R. B., ed.), Elsevier, Amsterdam, 1983, 222-230
[3] Berinde, Z., *On the caracterization of the compactness an branching using the topological indices ZEP and RZ*, U. P. B. Sci. Bull., Series B, **63** (2001), No. 4, 15–21
[4] Berinde, Z., *Consideraţii privind modelarea matricială a compuşilor halogenaţi*, Rev. Chim. (Bucuresti), **52** (2001), No. 12, 788–792
[5] Berinde, Z., *On a local invariant for modeling heteroatoms*, Proceed of PAMM 135, Baia Mare - Borşa, 4-7 october 2001, 47-54
[6] Berinde, Z., *Un nou indice topologic de tip Randić*, Rev. Chim. (Bucureşti), **53** (2002), No. 1, 812-816
[7] Berinde, Z., *Aplicaţii ale topologiei moleculare în studiul proprietăţilor fizico-chimice ale compuşilor organici*, Cub Press 22, 2001
[8] Berinde, Z., *Vertex- and edge-weighted molecular graphs for amines*, Revue Roumaine de Chimie, **51** (2006), No. 11, 1131-1135
[9] Diane, E., Needham, I-Chien Wei and Seybold, P. G., *Molecular modeling of the physical properties of alkanes*, J. Am. Chem. Soc. **110** (1988), No. 13, 4186-4194
[10] Diudea, M. and Ivanciuc, O., *Topologie moleculară*, Comprex, Cluj Napoca, 1995
[11] Kier L. B. and Hall L. H., *Molecular Connectivity in Chemistry and Drug Research*, Academic Press, New York, 1976
[12] Kier L. B. and Hall L. H., *Molecular Connectivity in Structure, Activity Analisys*, Wiley, New York, 1986
[13] Hansen, P., J. and Jurs, P.C., *Chemical applications of graph theory. Part I. Fundamentals and topological indices*, J. Chem. Ed. **65** (1988), No. 7, 574-580
[14] Medeleanu, M., *Corelări structură-proprietăţi prin metode topologice*, PhD Thesis, Bucureşti, 1997
[15] Mihalić, Z. and Trinajstić, N., *A graph-theoretical approach to structure-property relationships*, J. Chem. Educ. **69** (1992), No. 9, 701-712
[16] Plavsić, D., Nikolić, S. and Trinajstić, N., *On the Harary index for the characterization of chemical graphs*, J. of Math. Chem. **12** (1993), No. 1, 235-250
[17] Trinajstić, N., *Chemical Graph Theory*, CRC Press, Inc. Boca Raton, Florida, 1983

NORTH UNIVERSITY OF BAIA MARE
DEPARTMENT OF CHEMISTRY
VICTORIEI 76
430122 BAIA MARE, ROMANIA
E-mail address: zoitaberinde@ubm.ro
E-mail address: zoita_berinde@yahoo.com