

## MODELING BOILING POINTS OF ALKANE DERIVATIVES

ESMAT MOHAMMADINASAB<sup>a</sup>

**ABSTRACT.** Relationship between the boiling points of a series of alkanes and some topological indices and geometrical descriptors is investigated using the multi linear regression MLR method. The results revealed that Wiener, Randić and volume descriptors play a more important role in the description of boiling points of alkanes, in comparison to the other molecular descriptors.

**Keywords:** *Graph theory, MLR method, Alkane Derivatives.*

### INTRODUCTION

Alkanes, with general formula  $C_nH_{2n+2}$ , are organic compounds that only contain carbon and hydrogen atoms and only carbon-carbon single bonds. Alkanes react very poorly to ionic or other polar chemical species. The boiling point of alkanes, as a physico-chemical parameter, is of great importance in chemical engineering and chemical reactions [1].

Chemical graph theory is a branch of mathematics which combines graph theory and chemistry and has been extensively applied to predict the physico-chemical and biological properties of organic compounds through the quantitative structure-activity/property relationship (QSAR/QSPR). These are mathematical models which relate the physico-chemical properties of a set of molecules to structural indices or other types of molecular descriptors [2-10]. It is worth mentioning that many properties of chemical compounds are closely related to topological indices of their molecular structures; therefore, the correlational studies, enabling the prediction of molecular properties are of practical importance.

Hence, the necessity of finding formulas for calculating some topological indices and then performing QSAR/QSPR studies.

---

<sup>a</sup> *Department of Chemistry, Faculty of Sciences, Islamic Azad University, Arak Branch, Arak, Iran*

\* *Corresponding author: e-mohammadinasab@iau-arak.ac.ir*

## TOPOLOGICAL INDICES

### WIENER INDEX

Harold Wiener in 1947, introduced one of the first molecular descriptors, of topological nature, for acyclic saturated hydrocarbons. Wiener index of the graph  $G$  is the half-sum of all entries in the distance matrix  $D$  [11]:

$$W = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n D_{ij} \quad (1)$$

where the entries  $D_{ij}$  denote the (topological) distance between any two vertices  $i$  and  $j$  in the graph  $G$ .

### HYPER-WIENER INDEX

A related distance-based topological index is called the Hyper-Wiener index,  $WW(G)$ ; it is defined as:

$$WW(G) = \frac{1}{2} \sum d_{ij} + \frac{1}{2} \sum (d_{ij})^2 \quad (2)$$

where  $d_{ij}$  denotes the topological distance between the vertices  $i$  and  $j$  in the graph  $G$  and the summation runs over all (unordered) pairs of vertices of  $G$  [12,13].

### BALABAN INDEX

Balaban index,  $J=J(G)$ , of a graph on  $n$  node and  $m$  edges, is calculated by using the sum of all distances,  $D_i$ , from  $i$  to all the other vertices of  $G$ ; it was defined in 1982 by the Romanian chemist, A. T. Balaban, as follows [14]:

$$J = \frac{m}{\mu + 1} \sum_{i=1}^n \sum_{j=1}^n [(D_i)(D_j)]^{-0.5} \quad (3)$$

where  $\mu = m - n + 1$  is the cyclomatic number.

### RANDIC INDEX

In 1975, the Croatia scientist, Milan Randić introduced the first connectivity index [15], named nowadays the "Randić index"; it was defined

as the sum of all the bonds contribution, by means of  $d_i$  and  $d_j$  (being the degrees of the vertices/atoms "i", "j"):

$$\chi = \sum_{(i,j) \in D(G)} (d_i d_j)^{-1/2} \quad (4)$$

### HARARY INDEX

The Harary index,  $H(G)$ , was introduced independently in 1993 by Plavsic *et al.* [16] and Ivanciuc *et al.* [17]. In fact, this index was introduced by Ciubotariu and the QSAR Group of Timisoara (Romania) since 1987 [18,19]. The Harary index is defined as:

$$H = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (1/D_{i,j}) \quad (5)$$

where  $1/D_{i,j}$  represents the reciprocal distances in the graph.

### GEOMETRIC INDICES

Geometric analysis provides characteristic values related to the geometrical structure of a molecule such as minimal and maximal z length, minimal and maximal projection area, force field energies or van der Waals volume. The Dreiding energy related to the 3D structure of a molecule was determined using the Dreiding force field.

A quantitative structure–property relationship (QSPR) study was herein performed for prediction of boiling points of 66 aliphatic alkane derivatives using topological and geometric indices.

### RESULTS AND DISCUSSION

The experimental data of boiling point of alkanes and their derivatives are shown in Table 1.

The values of related molecular indices of 66 different types of alkanes and their derivatives were calculated using formulas (1 to 5); the values of geometry descriptors of all the mentioned compounds were extracted from book and web book [20].

The relationship between the boiling point of molecules used in the search and 11 different types of geometric and topological indices was investigated using the Excel software (equations 6-16).

**Table 1.** The values of experimental boiling point of the set of 66 alkanes

No	Compound	T <sub>bp</sub> (K)	No	Compound	T <sub>bp</sub> (K)
1	propane	231.05	34	2,3,3-trimethylpentane	387.85
2	n-butane	272.65	35	2,3,4-trimethylpentane	386.80
3	2-methylpropane	261.45	36	2-methyl-3-ethylpentane	388.80
4	n-pentane	309.25	37	3-methyl-3-ethylpentane	391.50
5	2-methylbutane	300.95	38	2,2,3,3-tetramethylbutane	379.60
6	2,2-dimethylpropane	282.65	39	n-nonane	423.90
7	n-hexane	342.15	40	2-methyloctane	415.95
8	2-methylpentane	333.45	41	3-methyloctane	416.45
9	3-methylpentane	336.45	42	4-methyloctane	415.00
10	2,2-dimethylbutane	322.85	43	2,2-dimethylheptane	405.85
11	2,3-dimethylbutane	331.15	44	2,3-dimethylheptane	413.65
12	n-heptane	371.55	45	2,4-dimethylheptane	406.65
13	2-methylhexane	363.15	46	2,5-dimethylheptane	408.00
14	3-methylhexane	365.15	47	2,6-dimethylheptane	408.35
15	2,2-dimethylpentane	352.35	48	3,3-dimethylheptane	410.45
16	2,3-dimethylpentane	362.95	49	3,4-dimethylheptane	413.25
17	2,4-dimethylpentane	353.65	50	3,5-dimethylheptane	409.15
18	3,3-dimethylpentane	359.25	51	4,4-dimethylheptane	408.35
19	2,3,3-trimethylpentane	354.05	52	3-ethylheptane	416.15
20	3-ethylpentane	366.65	53	2,2,4-trimethylhexane	399.65
21	n-octane	398.85	54	2,2,5-trimethylhexane	397.15
22	2-methylheptane	390.75	55	2,3,3-trimethylhexane	410.90
23	3-methylheptane	391.15	56	2,3,4-trimethylhexane	412.15
24	4-methylheptane	390.85	57	2,3,5-trimethylhexane	404.45
25	2,2-dimethylhexane	379.95	58	2,2,3,3-tetramethylpentane	413.42
26	2,3-dimethylhexane	388.75	59	2,2,3,4-tetramethylpentane	406.15
27	2,4-dimethylhexane	382.55	60	2,3,3,4-tetramethylpentane	414.65
28	2,5-dimethylhexane	382.15	61	2,2,4,4-tetramethylpentane	395.85
29	3,3-dimethylhexane	385.15	62	2,4-dimethyl-3-ethylpentane	409.88
30	3,4-dimethylhexane	390.85	63	3,3-diethylpentane	419.35
31	3-ethylhexane	391.50	64	n-decane	447.27
32	2,2,3-trimethylpentane	387.85	65	3-methylnonane	440.95
33	2,2,4-trimethylpentane	372.40	66	4-methylnonane	438.85

The following equations indicate the relationship between T<sub>bp</sub> and the values of calculated molecular indices.

$$\text{Bp} = 0.36 \text{ WW} + 322 \quad R^2 = 0.68 \quad (6)$$

$$\text{Bp} = 1.15 \text{ W} + 296.47 \quad R^2 = 0.84 \quad (7)$$

$$\text{Bp} = 9.92 \text{ H} + 236.82 \quad R^2 = 0.92 \quad (8)$$

MODELING BOILING POINTS OF ALKANE DERIVATIVES

$$Bp=47.88 J +231.94 \quad R^2 =0.29 \quad (9)$$

$$Bp=58.10 X +172.05 \quad R^2 =0.97 \quad (10)$$

$$Bp=6.41P +277.31 \quad R^2 =0.55 \quad (11)$$

$$Bp=1.64 V +142.90 \quad R^2 =0.96 \quad (12)$$

$$Bp=7.00 \text{ Min P A} +173.19 \quad R^2 =0.50 \quad (13)$$

$$Bp=16.69 \text{ Min Z L} +214.1 \quad R^2 =0.49 \quad (14)$$

$$Bp=4.86 \text{ Max P A} +149.02 \quad R^2 =0.85 \quad (15)$$

$$Bp=25.99 \text{ Max Z L} +224.5 \quad R^2 =0.13 \quad (16)$$

According to eqs. (6 to 16), and the square correlation coefficients one can see that there are better correlations between  $T_{bp}$  with: Randic > Volume > Harary > Max P A > Wiener of this class of alkanes, respectively. A poor correlation was obtained between  $T_{bp}$  and Max Z L, Min Z L, Min P A, Platt, HyperWiener and the Balaban index of mentioned alkanes.

In the next step, SPSS software, multiple linear regression (MLR) method (eq. 17) and backward procedure were used.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \epsilon \quad (17)$$

Whether or not the regression model explains a statistically significant proportion of data was ascertained through the ANOVA Table of output based on the MLR model in terms of the relationship between  $T_{bp}$  and effective molecular indices.

Therefore, different models were examined and the best model was defined using correlation coefficient, coefficient of determination, standard error of estimate, mean square, the Fischer statistics, and Durbin-Watson significance values. (Table 2).

**Table 2.** MLR statistics: Predictors  $X_i$ , Pearson correlation coefficient  $R^2$ ,  $R^2_{\text{Adjust}}$ , standard error of estimate  $s$ , Fisher Coefficient  $F$ , Mean Square  $MS$ , Significance of models  $Sig$ .

	$X_i$	$R^2$	$R^2_{\text{Adj}}$	$s$	$F$	$MS$	$Sig$
<b>1</b>	X, H, Volume, Mx P A, W	0.992	0.991	4.009	1.485E3	23860.749	0.000
<b>2</b>	X, Volume, Mx P A, W	0.992	0.991	3.988	1.875E3	29824.436	0.000
<b>3</b>	X, Volume, W	0.992	0.991	4.001	2.483E3	39758.370	0.000

To predict the  $T_{bp}$  values, three models were used with  $\text{sig} = 0.000$ ,  $F: 1.485E3 < 1.875E3 < 2.483E3$ ,  $M-S: 23860.749 < 29824 < 39758.370$ ; respectively (Table 2).

$$B_p = 109.615 + 42.683x + 1.038V - 0.971H + 0.199\text{Max P A} - 0.383W \quad (18)$$

$$B_p = 116.715 + 43.730x + 0.837V + 0.303\text{Max P A} - 0.393W \quad (19)$$

$$B_p = 121.780 + 48.452X + 0.778V - 0.380W \quad (20)$$

Finally, one model with a compromise between the highest square correlation coefficient ( $R^2 > 0.99$ ), Fisher coefficient ( $F = 2.483E3$ ), standard error of estimate (4.001) with significance level = 0.000 and the lowest number of descriptors was opted for further analysis, as reported in MLR eq. (20). Therefore, the best model with  $R^2 = 0.992$ ,  $R^2_{\text{Adjust}} = 0.991$ ,  $F = 2.483E3$ ,  $\sigma = 4.002$  K,  $MS = 39758.370$  K,  $DW = 1.651$  was used for prediction of alkanes  $T_{bp}$ . The best two descriptors are shown as predictors in the third model in terms of non-standardized coefficients.

This equation has three common descriptors  $X$ ,  $V$ ,  $W$  with high calibration statistics and prediction ability.

It is known that, the adjust coefficient ( $R^2_{\text{Adjust}}$ ) indicates the percentage of dependent variable that is justified by the independent variable. The less distinction between  $R^2_{\text{adjust}}$  and  $R^2$  indicates that the independent variables added to the model have been chosen more appropriately. The slight difference between the above amounts in the proposed model verifies the precision and accuracy of the model for predicting the critical properties. The coefficient which has been used in the statistical method is called the significance level. The more the significance level closes to zero, the smaller the significance level and the more meaningful the linear model will be. Therefore, a higher Fisher coefficient leads to a smaller significance coefficient.

Durbin-Watson (DW) statistics is a number giving information about the autocorrelation in the residuals. The statistical coefficient of Durbin-Watson test ranges between 0 and 4. The more this coefficient closes to zero, the less correlation between the errors will be. However, the amount of this number indicates there is no caution using the proposed models.

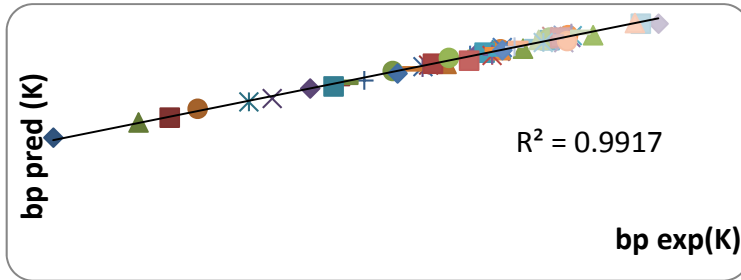
Table. 3 lists the  $T_{bp}$  of considered molecules predicted by eq. (20).

Figure 1 shows the linear correlation between the predicted  $T_{bp}$  by using eq. (20) compared with the experimental  $T_{bp}$ . The results indicate that the equation (20) can be used effectively for predicting the  $T_{bp}$  of alkanes. Therefore, the results are satisfactory.

MODELING BOILING POINTS OF ALKANE DERIVATIVES

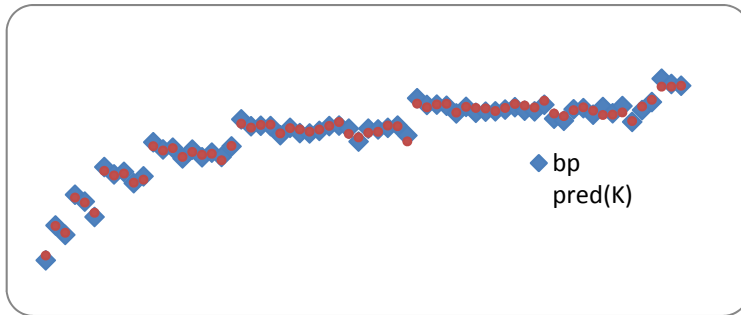
**Table 3.**  $T_{bp(Pred)}$  (K) with Eq. 20 and Residual (K) of Alkanes

No	$T_{bp(Pred)}$ (K)	Residual (K)	No	$T_{bp(Pred)}$ (K)	Residual (K)
1	236.93	-5.88	34	382.93	4.91
2	272.07	0.57	35	384.01	2.78
3	263.82	-2.37	36	391.46	-2.66
4	305.72	3.52	37	390.75	0.74
5	299.78	1.17	38	372.64	6.96
6	287.66	-5.01	39	417.43	6.46
7	337.50	4.64	40	413.09	2.86
8	331.96	1.48	41	416.60	-0.15
9	334.29	2.15	42	417.39	-2.38
10	323.53	-0.68	43	406.92	-1.07
11	327.02	4.12	44	413.68	-0.03
12	366.83	4.71	45	412.46	-5.8
13	361.79	1.35	46	411.79	-3.79
14	364.55	0.59	47	408.74	-0.39
15	354.15	-1.79	48	412.29	-1.83
16	360.05	2.90	49	417.17	-3.92
17	356.64	-2.98	50	415.22	-6.07
18	357.84	1.40	51	413.01	-4.66
19	350.09	3.95	52	420.89	-4.74
20	367.30	-0.65	53	405.72	-6.07
21	393.69	5.15	54	402.58	-5.43
22	388.99	1.76	55	409.75	1.14
23	392.11	-0.96	56	413.17	-1.02
24	392.49	-1.63	57	409.10	-4.65
25	382.05	-2.54	58	403.99	9.42
26	388.29	0.45	59	404.39	1.75
27	386.86	-4.31	60	406.96	7.69
28	384.23	-2.08	61	396.74	-0.89
29	386.60	-1.45	62	414.02	-4.13
30	391.07	-0.21	63	422.06	-2.71
31	395.59	-4.09	64	437.77	9.49
32	381.59	6.25	65	437.72	3.23
33	377.37	-4.97	66	438.89	-0.03



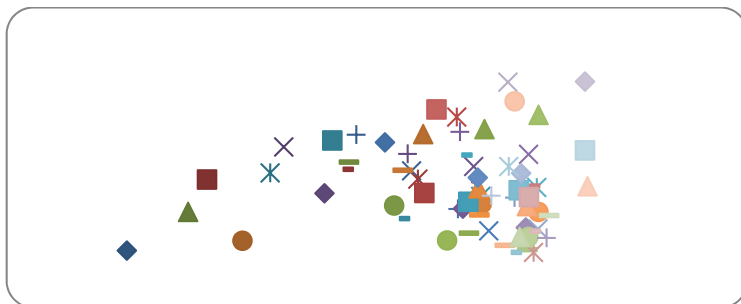
**Figure 1.** The plot of predicted bp vs experimental bp

Figure 2 shows the differences between the experimental and predicted boiling points of alkanes.



**Figure 2.** Comparison between the experimental and predicted of  $T_{bp}$  by MLR method

The residual values are represented at a fairly random pattern (Figure 3). Residuals are used to assess the normality of assumption. This random pattern indicates that a linear model provides a decent fit to the data.



**Figure 3.** The scatter plot of residuals against experimental values of alkanes boiling points



## CONCLUSIONS

Mathematical regression models are very important in QSAR modelling and the regression method, particularly MLR, can be used in this respect.

The results of this study indicated that the boiling points of alkane derivatives have a strong correlation with Wiener, Randić and volume descriptors.

## EXPERIMENTAL

To analyse the relationship between boiling point and molecular descriptors, the research data were collected in two stages:

First, the structures and the values of experimental available boiling points of 66 various types of alkanes and their derivatives used in the present investigation were taken from National Institute of Standard and Technology Chemistry web book and listed in Table 1.

Second, the values of Randic ( $\chi$ ), Harary (H), Balaban (J), Wiener (W), Platt (P) and HyperWiener (WW) topological indices related to the mentioned molecular graphs were calculated using the above formulas. Also the values of geometry descriptors such as the minimal z length ( $\text{Min.z.L/A}^\circ$ ), the maximal z length ( $\text{Max.z.L/A}^\circ$ ), the minimal projection area ( $\text{Min. P.A/A}^{\circ 2}$ ), the maximal projection area ( $\text{Max.P.A/A}^{\circ 2}$ ), the van der Waals volume ( $\text{V/A}^{\circ 3}$ ) for these 66 compounds were extracted from book and web book [20].

Third, the relationship between boiling points and 11 various types of molecular descriptors was investigated for the set of alkanes derivatives, using excel software, and the relevant equations were extracted.

Finally, the boiling points of these alkanes and their derivatives were estimated using SPSS software (version 16) by multiple linear regression method and backward procedure. The best molecular indices were determined for prediction of the boiling points of mentioned compounds in terms of important parameters such as Fisher statistics, Durbin Watson, correlation coefficient, square correlation coefficient, adjust square correlation coefficient, etc. (Table 2).

## REFERENCES

1. A. Arora, "Hydrocarbons, Alkanes, Alkenes and Alkynes", Discovery Publishing House, **2006**.
2. L. Blaha, J. Damborsky, M. Nemeč, *Chemosphere*, **1998**, 36(6), 1345.
3. N. Trinajstić, I. Gutman, *Croatica Chemica Acta*, **2002**, 75, 329.

4. N. Gutman, Trinajstić, *Chemical Physics Letters*, **1972**, 17, 535.
5. I. Gutman, B. Furtula (eds), "Novel molecular structure descriptors-theory and applications I and II", Univ. Kragujevac, **2010**, 3.
6. I. Gutman, B. Ruscic, N. Trinajstić, C.F. Wicox, *Journal of Chemical Physics*, **1975**, 62, 3399.
7. E. Mohammadinasab, and M. Goodarzi, *Fullerenes, Nanotubes, and Carbon Nanostructures*, **2011**, 19, 550.
8. A.A. Taherpour, E. Mohammadinasab, *Fullerenes, Nanotubes and Carbon Nanostructures*, **2010**, 18, 76.
9. M. Monajjemi, E. Mohammadinasab, F. Shafiei, *Journal Physical&Theoretical Chemistry*, **2009**, 6, 125.
10. M. Goodarzi, E. Mohammadinasab, *Fullerenes, Nanotubes and Carbon Nanostructures*, **2013**, 21, 102.
11. H. Wiener, *Journal of American Chemical Society*, **1947**, 69, 17.
12. M. Randić, *Chemical Physics Letters*, **1993**, 211, 478.
13. G. Cash, S. Klavzar, M. Petkovsek, *Journal Chemical Information and Computer Science*, **2002**, 42, 571.
14. A.T. Balaban, *Chemical Physics Letters*, **1982**, 89(5), 399.
15. M. Randić, *Journal American Chemical Society*, **1975**, 23, 6606.
16. D. Plavšić, S. Nikolić, N. Trinajstić and Z. Mihalić, *Journal of Mathematical Chemistry*, **1993**, 12(1), 235.
17. O. Ivanciuc, T.S. Balaban, and A.T. Balaban, *Journal of Mathematical Chemistry*, **1993**, 12, 309.
18. D. Ciubotariu, "Structură-reactivitate în clasa derivațiilor acidului carbonic", PhD Thesis, Timisoara, Romania, **1987**.
19. D. Ciubotariu, M. Medeleanu, V. Vlaia, T. Olariu, C. Ciubotariu, D. Dragos and C. Seiman (2004) *Molecules*, **2004**, 9, 1053.
20. Web search engine developed by ChemAxon; software available at [http:// www. Chemicalize. Org](http://www.Chemicalize.Org).