Comparison between the structure–boiling point relationships with different descriptors for condensed benzenoids

Dejan Plavšić,[a] **Nenad Trinajstić,***[,a] **Dragan Amić**[b] **and Milan Šoškić**[c]

[a] *The Rugjer Bošković Institute, P. O. Box 1016, HR-10001 Zagreb, Croatia*
[b] *Faculty of Agriculture, The Josip Juraj Strossmayer University, HR-31001 Osijek, Croatia*
[c] *Faculty of Agriculture, University of Zagreb, HR-10001 Zagreb, Croatia*

We compare the structure–boiling point models for benzenoid hydrocarbons based on three distance descriptors (the detour index, the Wiener sum index and its variant) and two connectivity indices (the vertex- and edge-connectivity indices) with those based on the geometric distance indices. The conclusions of this study are that the best model is based on the edge-connectivity index and that the models based on the geometric distance indices are good but not as good as models based on the connectivity indices.

Randić[1] has recently reported in this journal a quantitative structure–boiling point relationship for benzenoid hydrocarbons, using novel geometric distance-dependent descriptors. This paper stimulated us to analyze the performance of several newly introduced molecular descriptors in structure–boiling point relationships for condensed benzenoids. We also compare the structure–boiling point relationships obtained by us with those obtained by Randić. In the analysis we use the same set of benzenoid hydrocarbons and the same experimental values of boiling points[2] as Randić.

In order to simplify the presentation we use the graph theoretical terminology[3] in referring to benzenoids and their structural characteristics, that is, instead of benzenoid hydrocarbons, atoms, bonds, valencies, *etc*., we often use the terms benzenoid graphs, vertices, edges, degrees, *etc*., because the molecular descriptors that are considered were obtained using (chemical) graph theoretical concepts.[4]

Brief Description of the Molecular Descriptors used

We have used five molecular descriptors in our work: three recently introduced descriptors, the edge-connectivity index[5] ε, the detour index[6] ω and the Wiener sum (WS) of the quotient matrix[7] $\mathbf{D}/\Delta$, as well as the vertex-connectivity index χ, which has been in use for more than twenty years[8] and a variant of the WS index denoted by ws. The vertex- and edge-connectivity indices will be called 'connectivity indices' for short, whilst the detour index, the Wiener sum and its variant 'distance indices'.

The vertex-connectivity index

The vertex-connectivity index $\chi = \chi(G)$ of a (molecular) graph G is defined as:[8]

$$\chi = \sum_{\text{edges}} [d(v_i)\, d(v_j)]^{-1/2} \tag{1}$$

where $d(v_i)$ is the vertex-degree of a vertex v_i and $[d(v_i)\, d(v_j)]^{-1/2}$ is the weight of the i-j edge. Note that the degree $d(v_i)$ of the vertex v_i is equal to the number of adjacent vertices. Two vertices of a graph are adjacent if there is an edge joining them.

For benzenoid graphs (representing carbon skeletons of benzenoid hydrocarbons), eqn. (1) can be given in a closed form: in benzenoid graphs only two types of vertices appear, that is, vertices with degree 2 and vertices with degree 3. This leads to the following three edge-weights in benzenoids: $1/2 = 0.5$, $1/\sqrt{6} = 0.408$ and $1/3 = 0.333$. The numbers of the corresponding edges are denoted as e_{22}, e_{23} and e_{33}, respectively. Introducing these numbers and the corresponding weights into eqn. (1), one obtains:[9]

$$\chi = 0.5\, e_{22} + 0.408\, e_{23} + 0.333\, e_{33} \tag{2}$$

Therefore, all one needs to compute the vertex-connectivity indices of benzenoid graphs is to enumerate their edge-types.

The edge-connectivity index

The edge-connectivity index $\varepsilon = \varepsilon(G)$ of a (molecular) graph G is defined in a similar way as the vertex-connectivity index, but it uses edge-degrees $d(e_i)$ instead of vertex-degrees $d(v_i)$:[5]

$$\varepsilon = \sum_{\substack{\text{adjacent}\\ \text{edges}}} [d(e_i)\, d(e_j)]^{-1/2} \tag{3}$$

where $[d(e_i)\, d(e_j)]^{-1/2}$ is the weight of adjacent edges e_i and e_j. Note that the degree $d(e_i)$ of an edge e_i is equal to the number of adjacent edges. Two edges of a graph G are adjacent if they are incident with the same vertex.

For benzenoid graphs eqn. (3) can also be given in a closed form. In benzenoid graphs three types of edge-degrees appear, that is, edges with degree 2, degree 3 and degree 4. This leads to the following five weights of adjacent edges in benzenoids: $1/2 = 0.5$, $1/\sqrt{6} = 0.408$, $1/3 = 0.333$, $1/2\sqrt{3} = 0.289$ and $1/4 = 0.25$. The numbers of adjacent edges with the above weights are denoted by n_{22}, n_{23}, n_{33}, n_{34} and n_{44}, respectively. Introducing these numbers and the corresponding weights into eqn. (3), one obtains:[10]

$$\chi = 0.5\, n_{22} + 0.408\, n_{23} + 0.333\, n_{33} + 0.289\, n_{34} + 0.25\, n_{44} \tag{4}$$

Hence, all one needs to compute the edge-connectivity indices of benzenoid graphs is to enumerate the numbers of adjacent edges and to classify them according to their weights.

The detour index

The detour index ω is defined in a similar way to the Wiener index W,[11,12] that is, the detour index is equal to the half-sum of the elements of the detour matrix Δ:[6]

$$\omega = (1/2) \sum_i \sum_j (\Delta)_{ij} \tag{5}$$

---

* E-mail: trina@rudjer.irb.hr

The detour matrix $\Delta = \Delta(G)$ of a labelled connected graph G with N vertices is a real symmetric $N \times N$ matrix whose (i,j) entry is the length of the longest path[4] (elongation) from vertex v_i to vertex v_j.[13,14] By definition the elements on the main diagonal of the matrix are zero.

The Wiener sum index

The Wiener sum index WS is defined as the half-sum of the quotient matrix $\mathbf{D}/\Delta$ of a graph G:[7]

$$\mathrm{WS} = (1/2) \sum_i \sum_j (\mathbf{D}/\Delta)_{ij} \tag{6}$$

where $\mathbf{D}$ is the distance matrix and Δ is the detour matrix. The distance matrix $\mathbf{D} = \mathbf{D}(G)$ of a labelled connected graph G with N vertices is a real symmetric $N \times N$ matrix whose (i,j) entry is the length of the shortest path[4] from vertex v_i to vertex v_j.[15] By definition the diagonal entries $(\mathbf{D}/\Delta)_{ii}$ $(i = 1, \ldots, N)$ are equal to 0.

A variant of the Wiener sum index

A variant of the Wiener sum index ws is defined in a similar way to the Wiener sum index except the elements in the quotient matrix are inverted:

$$\mathrm{ws} = (1/2) \sum_i \sum_j (\Delta/\mathbf{D})_{ij} \tag{7}$$

By definition the diagonal entries $(\Delta/\mathbf{D})_{ii}$ $(i = 1, \ldots, N)$ are equal to 0.

Results and Discussion

All five indices, reviewed above, for 22 benzenoid hydrocarbons were computed using our own computer program. This program is available on request. The values of these indices are given in Table 1 together with the boiling points (in °C) of the considered benzenoids.

The values of the Wiener sum indices for naphthalene, phenanthrene, anthracene, triphenylene, chrysene, benzphenanthrene and tetracene differ from those given by Randić.[7] If these are not typographical errors, then this may indicate that Randić made a systematic error in computing the detour matrix. The computation of the distance matrix is fairly simple,[16,17] but computing the detour matrix is rather involved[18,19] and if it is done by hand it is error-prone. The problem of finding the detour matrix is said to be NP-complete, that is, the computer time required for computing the detour matrix increases exponentially with the increasing size of a molecule. We devised a computer program that is based on path-tracing and is efficient for modestly sized molecules.

A structure–boiling point modelling was carried out by simple and multivariate regression analysis. We also considered two sets of benzenoid hydrocarbons: one set (complete set) containing all 22 benzenoids and another without benzene (set without benzene) because strictly speaking benzene does not belong to the condensed benzenoids.[20,21] Below we list the best linear, quadratic and multivariate structure–boiling point models of benzenoids for each set. The comparison between the models is made on the basis of their standard errors. In order to make the comparison between different models easier, the statistical parameters are collected in Table 2.

Complete set

Linear correlations. The best linear correlation with distance indices is:

$$\mathrm{bp} = (80.32 \pm 27.40) + (9.46 \pm 0.63)\mathrm{WS} \tag{8}$$

Regression coefficients and their standard errors are shown in parentheses.

This is not a particularly good correlation. It improves considerably when the vertex-connectivity indices are used:

$$\mathrm{bp} = (-59.86 \pm 9.85) + (55.02 \pm 0.99)\chi \tag{9}$$

A comparable result is obtained when the edge-connectivity indices are used:

$$\mathrm{bp} = (-14.97 \pm 10.15) + (42.37 \pm 0.86)\varepsilon \tag{10}$$

The above is not surprising since χ and ε are closely related connectivity indices, related through the expression:

$$d(e_k) = d(v_i) + d(v_j) - 2 \tag{11}$$

Eqn. (11) is based on the following simple argument. Since every edge in G connects two vertices, then the vertices of edge

Table 1 Vertex-connectivity indices (χ), edge-connectivity indices (ε), Wiener sums of $\mathbf{D}/\Delta$ (WS) and $\Delta/\mathbf{D}$ (ws) matrices, detour indices (ω) and boiling points (bp in °C) of the studied benzenoid hydrocarbons

Benzenoid hydrocarbon	Molecular descriptors					
	χ	ε	WS	ws	ω	bp
Benzene	3.000	3.000	7.200	45.0	63	80.1
Naphthalene	4.966	5.454	14.859	190.1	345	218.0
Phenanthrene	6.950	7.926	25.058	474.8	1019	338.0
Anthracene	6.933	7.942	26.094	470.2	1015	340.0
Chrysene	8.933	10.397	37.484	922.1	2275	431.0
Benzanthracene	8.916	10.414	38.434	915.5	2265	425.0
Triphenylene	8.950	10.414	36.609	929.6	2229	429.0
Tetracene	8.899	10.430	39.866	906.7	2257	440.0
Benzo[*a*]pyrene	9.916	11.897	41.024	1225.1	3166	496.0
Benzo[*e*]pyrene	9.933	11.897	38.862	1242.5	3190	493.0
Perylene	9.933	11.897	39.526	1222.9	3146	497.0
Anthanthrene	10.899	13.397	44.693	1637.7	4341	547.0
Benzoperylene	10.916	13.380	42.837	1709.1	4433	542.0
Dibenzo[*a,c*]anthracene	10.916	12.902	51.288	1562.0	4215	535.0
Dibenzo[*a,h*]anthracene	10.899	12.885	54.028	1539.7	4261	536.0
Dibenzo[*a,i*]anthracene	10.899	12.885	52.956	1544.0	4265	531.0
Picene	10.916	12.868	52.574	1551.5	4299	519.0
Coronene	11.899	14.863	48.140	2148.8	5790	590.0
Dibenzo[*a,k*]pyrene	11.916	14.385	53.091	1992.0	5640	592.0
Dibenzo[*a,h*]pyrene	11.899	14.385	56.007	1970.4	5644	596.0
Dibenzo[*a,g*]pyrene	11.899	14.385	56.063	1970.2	5642	594.0
Dibenzo[*a,d*]pyrene	11.916	14.390	52.286	2000.1	5640	595.0

Table 2 Collected statistical parameters (r = correlation coefficient, s = standard error of estimate and F = Fisher ratio) for both sets of hydrocarbons and all models

	Complete set ($n = 22$)				Set without benzene ($n = 21$)		
Model[a]	r	s	F	Model[a]	r	s	F
Linear correlation							
(8)	0.958	38.5	223.4	(17)	0.974	23.4	346.3
(9)	0.997	10.8	3075.6	(18)	0.997	8.1	3044.8
(10)	0.996	12.1	2448.5	(19)	0.997	7.9	3178.5
Quadratic correlation							
(12)	0.987	22.6	344.0	(20)	0.994	12.0	683.2
(13)	0.998	8.9	2241.6	(21)	0.997	8.2	1459.9
(14)	0.998	7.9	2897.7	(22)	0.998	7.3	1847.8
Multivariate regression							
(15)	0.998	7.8	2956.5	(23)	0.998	7.2	1944.3
(16)	0.996	12.2	1201.3	(24)	0.996	8.9	1260.2

[a] Numbers in parentheses correspond to equations in the text.

e_k can be expressed in terms of degrees of its incident edges v_i and v_j. As already stressed χ and ε are closely related indices, but, except for benzene, they are not identical indices. All condensed benzenoid hydrocarbons possess different values of the χ and ε indices because a different number of terms enter into the formulae for computing χ and ε. For example, in the case of phenanthrene, 16 terms enter into eqn. (2), that is, $e_{22} = 7$, $e_{23} = 6$ and $e_{33} = 3$ and 22 terms enter into eqn. (4), that is, $n_{22} = 4$, $n_{23} = 6$, $n_{33} = 2$, $n_{34} = 8$ and $n_{44} = 2$.

Linear correlations with Randić's first- (^1D) and second-order (^2D) geometric distance indices are not particularly good,[1] the standard error in the former case being 32.2 and in the latter case 56.7.

Quadratic correlations. The best quadratic correlation with distance indices is:

$$\text{bp} = (121.97 \pm 16.80) + (42.37 \pm 3.16)\text{ws} + (-0.97 \pm 0.13)(\text{ws})^2 \qquad (12)$$

In this and the following equations [eqn. (15), (17), (20), (23)] numerical values of ws were divided by 100 to be of the same order of magnitude as those of other indices used in the regression analysis. Again, this is not a particularly good correlation, but somewhat better than model (8) above. The vertex- and edge-connectivity indices considerably improve the correlation:

$$\text{bp} = (-114.25 \pm 18.97) + (70.29 \pm 4.88)\chi + (-0.94 \pm 0.30)\chi^2 \qquad (13)$$

$$\text{bp} = (-79.87 \pm 13.92) + (58.10 \pm 3.02)\varepsilon + (-0.82 \pm 0.16)\varepsilon^2 \qquad (14)$$

Multivariate regression analysis. We investigated all possible two-parameter combinations between the considered descriptors. In general, combinations of two different descriptors usually produced lower standard errors. The best multivariate model obtained is:

$$\text{bp} = (-71.67 \pm 12.39) + (54.09 \pm 2.24)\varepsilon + (-6.07 \pm 1.13)\text{ws} \qquad (15)$$

Model (15) is so far the best model. In Fig. 1 we plot the computed boiling points from model (15) and the experimental boiling points for the considered benzenoid hydrocarbons.

A model containing the first- and second-order geometric distance indices is not as good as model (15), but is much better (in terms of standard deviation) than either of the linear models with the first- or second-order geometric distance indices:

$$\text{bp} = (19.88 \pm 10.85) + (14.60 \pm 0.72)\,^1D + (-3.45 \pm 0.31)\,^2D \qquad (16)$$

The best model (one based on the ε and ws indices) predicts the boiling point of benzenoid hydrocarbons within the error range of 0.1–9.7%.

Set without benzene

Linear correlations. The best linear correlation with distance indices is:

$$\text{bp} = (257.26 \pm 13.49) + (17.36 \pm 0.93)\text{ws} \qquad (17)$$

This is not a good correlation, but it is much better than model (8) for the complete set. The correlations with the vertex- and edge-connectivity indices are much better:

$$\text{bp} = (-34.38 \pm 9.66) + (52.58 \pm 0.95)\chi \qquad (18)$$

$$\text{bp} = (13.86 \pm 8.61) + (40.07 \pm 0.71)\varepsilon \qquad (19)$$

Linear correlations with first- (^1D) and second-order (^2D) geometric distance indices are again not particularly good, the standard error in the former case being 26.7 and in the latter case 46.6.

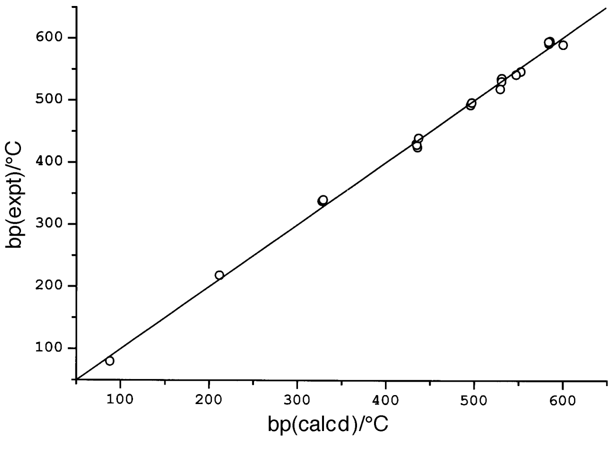

Fig. 1 A comparison of computed [using eqn. (15)] and experimental boiling points of 22 benzenoid hydrocarbons (the complete set)

Quadratic correlations. The best quadratic correlation with distance indices is:

$$\mathrm{bp} = (181.89 \pm 12.38) + (32.84 \pm 2.16)\mathrm{ws} + (-0.63 \pm 0.09)(\mathrm{ws})^2 \quad (20)$$

This is again not a particularly good correlation, but is better than model (17). Introduction of the quadratic term in the model based on the vertex-connectivity index does not improve the correlation:

$$\mathrm{bp} = (-50.26 \pm 35.35) + (56.33 \pm 8.07)\chi + (-0.21 \pm 0.45)\chi^2 \quad (21)$$

In the case of the edge-connectivity index the addition of the quadratic term reduces the standard error of estimate from 7.9 °C [eqn. (19)] to 7.3 °C:

$$\mathrm{bp} = (-35.80 \pm 25.90) + (50.02 \pm 4.98)\varepsilon + (-0.47 \pm 0.23)\varepsilon^2 \quad (22)$$

Multivariate regression analysis. Again, we investigated all possible two-parameter combinations between the considered indices. The best multivariate model is the following:

$$\mathrm{bp} = (-32.32 \pm 21.68) + (47.99 \pm 3.53)\varepsilon + (3.57 \pm 1.56)\mathrm{ws} \quad (23)$$

The combination of linear terms of the vertex- and edge-connectivity indices gives a model of comparable quality ($r = 0.998$, $s = 7.3$). Model (23) is overall the best model by virtue of the smallest value of the standard error. In Fig. 2 we plot the computed boiling points from model (23) and the experimental boiling points for the studied benzenoids.

A model containing the first- and second-order geometric distance indices is not as good as model (23), but is much better than either of the linear models with the first- or second-order geometric distance indices:

$$\mathrm{bp} = (48.83 \pm 10.45) + (13.39 \pm 0.59)\,{}^1D + (-3.05 \pm 0.25)\,{}^2D \quad (24)$$

The best model (the one based on the edge-connectivity index ε and the variant of the Wiener sum index ws) predicts the boiling point of benzenoid hydrocarbons within the error range of 0.0–2.4%. It should be noted that the connectivity indices are heavily dependent on the molecular size. Thus, the structure–boiling point correlations based on the molecular size in terms of the number of carbon atoms are comparable to correlations with connectivity indices. For example, relevant statistical parameters for the complete set and the set without benzene, respectively, are: (*a*) linear correlations: $r = 0.997, 0.997$; $s = 10.57, 8.0$ and (*b*) quadratic correlations: $r = 0.998, 0.997$, $s = 8.8, 8.2$.

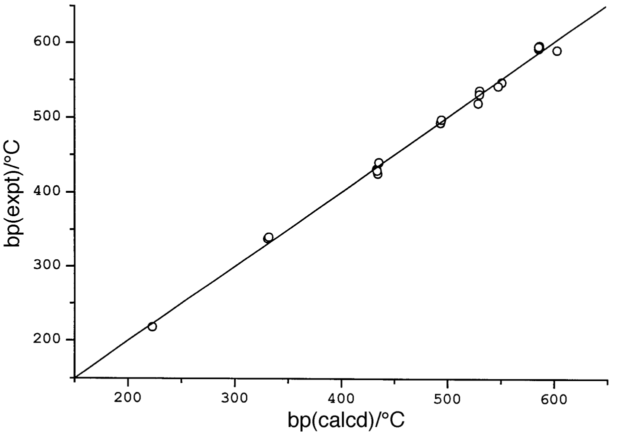

Fig. 2 A comparison of computed [using eqn. (23)] and experimental boiling points of 21 benzenoid hydrocarbons (the set without benzene)

The following conclusions can be drawn from this comparative study:(1) Distance indices are less useful than connectivity indices for predicting the boiling points of benzenoid hydrocarbons. (2) Among the three distance indices analyzed in this work the best models are obtained by using the variant of the Wiener sum index. (3) Correlations with the edge-connectivity index produce the best models. (4) In all cases the geometric distance indices produced models inferior to models containing the connectivity indices. (5) In general, the exclusion of benzene from the set of condensed benzenoid hydrocarbons gives models with significantly lower standard errors.

Concluding remarks

A comparative study of structure–boiling point models based on various descriptors has shown that the boiling points of condensed benzenoid hydrocarbons can be predicted within the error range 0.0–2.4%. If benzene is included in this set than the error range increases (0.1–9.4%). The best models are found to be those with the edge-connectivity index, but those containing the vertex-connectivity index are also very good. Models based on the geometric distance indices are good but somewhat inferior to models based on the connectivity indices. Models based on the distance indices are not so good. These indices show some potential as supporting descriptors in multivariate regression models.

Acknowledgements

This work was supported by the Ministry of Science and Technology of the Republic of Croatia through Grants no. 00980606 (D.P., N.T.) and 079301 (D.A.). We thank referees for helpful comments.

References

1 M. Randić, *New J.Chem.*, 1996, **20**, 1001.
2 S. C. Basak, G. D. Grunwald and G. J. Niemi, in *From Chemical Topology to Three-Dimensional Geometry*, ed. A. T. Balaban, Plenum Press, New York, 1997, pp. 73–116.
3 F. Harary, *Graph Theory*, Addison-Wesley, Reading, MA, 1971.
4 N. Trinajstić, *Chemical Graph Theory*, CRC Press, Boca Raton, FL, 2nd edn., 1992.
5 E. Estrada, *J. Chem. Inf. Comput. Sci.*, 1995, **35**, 31.
6 I. Lukovits, *Croat. Chem. Acta*, 1996, **69**, 873.
7 M. Randić, *J. Chem. Inf. Comput. Sci.*, 1997, **37**, 1063.
8 M. Randić, *J. Am. Chem. Soc.*, 1975, **97**, 6609.
9 M. Randić, Ž. Jeričević, A. Sabljić and N. Trinajstić, *Acta Phys. Polonica*, 1988, **A74**, 317.
10 S. Nikolić, N. Trinajstić and I. Baučić, *J. Chem. Inf. Comput. Sci.*, 1998, **38**, 42.
11 H. Hosoya, *Bull Chem. Soc. Jpn*, 1971, **44**, 2332.
12 S. Nikolić, N. Trinajstić and Z. Mihalić, *Croat. Chem. Acta*, 1995, **68**, 105.
13 D. Amić and N. Trinajstić, *Croat. Chem. Acta*, 1995, **68**, 53.
14 N. Trinajstić, S. Nikolić, B. Lučić, D. Amić and Z. Mihalić, *J. Chem. Inf. Comput. Sci.*, 1997, **37**, 631.
15 Z. Mihalić, D. Veljan, D. Amić, S. Nikolić, D. Plavšić and N. Trinajstić, *J. Math. Chem.*, 1992, **11**, 223.
16 M. Bersohn, *J. Comput. Chem.*, 1983, **4**, 110.
17 W. R. Müller, K. Szymanski, J. V. Knop and N. Trinajstić, *J. Comput. Chem.*, 1987, **8**, 170.
18 I. Lukovits and M. Razinger, *J. Chem. Inf. Comput. Sci.*, 1997, **37**, 283.
19 N. Trinajstić, S. Nikolić and Z. Mihalić, *Int. J. Quantum Chem.*, 1997, **65**, 415.
20 I. Gutman and S. J. Cyvin, *Introduction to the Theory of Benzenoid Hydrocarbons*, Springer-Verlag, Berlin, 1989.
21 N. Trinajstić, *J. Math. Chem.*, 1990, **5**, 171.

Received in Montpellier, France, 14th April 1998;
Paper 8/03004I