communications in mathematical
match
and in computer chemistry

no. 44, October 2001
MATCDY (44) 361 - 386 (2001)
ISSN 0340 - 6253

ON THE BALABAN-LIKE TOPOLOGICAL INDICES*

Sonja Nikolić, Dejan Plavšić and Nenad Trinajstić

The Rugjer Bošković Institute, P.O. Box 180, HR-10002 Zagreb, Croatia

Looking back, I think that there may be some truth in the motto we had on the wall of our laboratory: "For weaker characters, difficulties provide opportunities for excuse; *for stronger ones, they are challenges stimulating efforts to overcome them."*

A. T. Balaban [1]

Abstract. The Balaban index, Balaban-like topological indices (the complement Balaban index, the Harary-Balaban index, the quotient Balaban indices of the first and second kind), their variable counterparts, and vertex- and edge-connectivity indices are used in the comparative study of the structure-motor octane number modeling. Variable indices produced slightly better linear models than fixed indices. The best models obtained are quadratic models with the Harary-Balaban index and the quotient Balaban index of the second kind. All models found in the literature, but the model based on the Wiener number, are much poorer than models produced in the present study.

*Dedicated to Professor Alexandru T. Balaban on the occasion of his 70th anniversary.

INTRODUCTION

There are more than 3000 experimental and theoretical molecular descriptors available in the literature [2]. Among the theoretical descriptors there is a large class of graph-theoretical descriptors called topological indices [3,4]. A topological index is a single number that is used to characterize the graph corresponding to a molecule [5]. In the vast class of topological indices only a few are used successfully in the quantitative structure-property/activity relationships (QSPR/QSAR) [6]. Or as Milan Randić is fond of saying "*All descriptors are equal, but some are more equal than others*" paraphrasing George Orwell's well-known quote from "Animal Farm".

It appears that practitioners of the empirical QSPR/QSAR modeling are rather skeptical about the use of topological indices [7]. This is remindful of the attitude of organic chemists towards the quantum chemistry during its early days. However, a recent work by Lahana and co-workers [8] convincingly demonstrated the power of topological indices in the structure-property-activity modeling and predicting new drugs. These authors described a rational design of immunosuppressive peptides without relying on information regarding their receptors or mechanisms of action. Their design strategy uses a variety of topological indices and shape descriptors in combination with an analysis of molecular dynamic trajectories for the identification of potential drug candidates. Their approach started from 27 descriptors such as molecular volume, lipophilicity, the connectivity index [9], the Wiener index [10], a number of indicator variables (the number of carbon atoms, the number of hydrogen atoms, the number of methyl groups, the number of amino groups) and the Balaban index [11]. Statistical regression analysis showed that several of descriptors correlated with each other (the correlation coefficient being greater than 0.75). This allowed a reduction of the number of descriptors used to differentiate between active and inactive peptides. Thirteen independent descriptors were used as static filters to screen a virtual combinatorial library. The Balaban index also appeared in this set of thirteen descriptors.

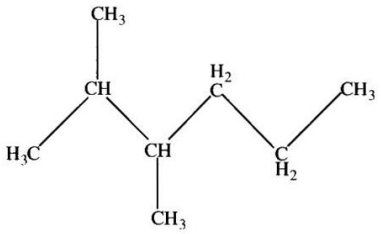
Lahana and co-workers generated a virtual combinatorial library for the peptides of the general form RXXXRXXXXY with seven variable positions. Use of 35 amino acids, 20 natural and 15 unnatural, leads to 35^7 combinations - 64 billion compounds, well above the present computing power. All the amino acids were characterized by means of their physicochemical properties (lipophilicity, acidity, aromaticity, etc.) and also by topological indices. The above number of compounds in the library was reduced to 6^7 (279,936 compounds) by taking into account lipophilicity distribution, considered critical for the activity studied. Screening the library of 279,936 compounds, using two types of filters: static and dynamic, resulted in the identification of 26 peptides satisfying all constraints. Biological activity of these peptides was tested in a heterotopic mouse heart allograft model. The molecule predicted to be the most potent displayed an immunosuppressive activity approximately 100 times higher than the lead compound.

The above work vindicate those who like Basak and his group [e.g.,12-15] have used consistently the Balaban index and its variants in the set of descriptors they employed in their work. The Balaban index, therefore, appears to be an important topological index and is consequently integrated in a number of computer programs for calculating topological indices such as POLLY [16], CODESSA [17], TAM [18], etc. Balaban himself and his former Ph.D. student (now Dr.) Ivanciuc also prepared computer program for calculating his index for heterosystems [19].

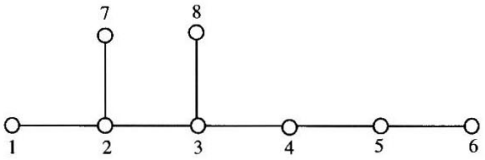
The aim of this report is two-fold: (i) to see the behavior of the Balaban-like indices regarding the weights of outer and inner bonds in alkanes and (ii) to compare structure-property models based on different Balaban-like indices. The property considered will be the motor octane number. This property is used because it has also been employed by Balaban in his early structure-property studies using his index.

In this paper we will use by choice graph-theoretical concepts and terminology [3,20]. Since in this paper we will be concerned only with alkanes, they will be represented by hydrogen-depleted trees. In Figure 1 we give as an example a tree representing 2,3-dimethylhexane.

(i) 2,3-dimethylhexane



(ii) Labeled hydrogen-depleted tree



T

FIGURE 1. A labeled tree T representing the hydrogen-depleted carbon skeleton of 2,3-dimethylhexane.

DEFINITION OF THE BALABAN INDEX

Randić introduced 25 years ago [9] an index he called the branching index, denoted by χ, that was soon after renamed into the connectivity index [21] and recently into the vertex-connectivity index [22,23] after a connectivity index related to graph-edges, the edge-connectivity index, was proposed [24]. The vertex-connectivity index was popularized by Kier and Hall [25,26] and has found a considerable use in QSPR and QSAR [3,4,6]. The vertex-connectivity index is given by:

$$\chi = \sum_{\text{edges}} [d(v_i)\, d(v_j)]^{-0.5} \qquad (1)$$

where $d(v_i)$ is the degree of a vertex v_i.

Balaban has shown that the vertex-connectivity index is associated with the vertex-adjacency matrix of a (molecular) graph [11]. Sum of elements in each row of the vertex-adjacency matrix is equal to the degree of a vertex, that is, elements in the i-th row give $d(v_i)$. Thus, one can straightforwardly obtain the vertex-connectivity index from the vertex-adjacency matrix.

The vertex-adjacency matrix **A** of a labeled connected graph with V vertices is the V x V matrix whose elements are defined as:

$$(A)_{ij} = \begin{cases} 1 \text{ if vertices i and j are adjacent} \\ 0 \text{ otherwise} \end{cases} \quad (2)$$

Estrada [24], when he introduced the edge-connectivity index, denoted by ε, used implicitly Balaban's idea. He used the edge-adjacency matrix ${}^{E}\mathbf{A}$ [3] that defined as:

$$({}^{E}A)_{ij} = \begin{cases} 1 \text{ if vertices i and j are adjacent} \\ 0 \text{ otherwise} \end{cases} \quad (3)$$

The edge-connectivity index is given by:

$$\varepsilon = \sum_{\substack{\text{adjacent} \\ \text{edges}}} [d(e_i)\, d(e_j)]^{-0.5} \quad (4)$$

where $d(e_i)$ is the degree of an edge e_i.

Sum of elements in each row of the edge-adjacency matrix is equal to the degree of an edge, that is, elements in the i-th row give $d(e_i)$. Thus, the edge-connectivity index can be obtained simply from the edge-adjacency matrix. The edge-connectivity index has also found use in the structure-property-activity modeling [22,23,27-31].

Calculations of the vertex-connectivity and edge-connectivity indices, using the Balaban algorithm, for a tree representing the carbon skeleton of 2,3-dimethylhexane are shown in Table 1 and Table 2.

If one uses some other matrix rather the vertex/edge-adjacency matrix we can obtain an index that would formally resemble the vertex/edge-connectivity index. This was idea that guided Balaban to propose his index [11]. He used graph-theoretical distance matrix [32] instead of the vertex-adjacency matrix and obtained a distance-based connectivity index he called the average distance-sum connectivity and denoted by J:

$$J = F \sum_{\text{edges}} (d_i\, d_j)^{-0.5} \qquad (5)$$

where d_i's are the distance-sums (or distasums, for brevity [6]).

TABLE 1. Calculation of the vertex-connectivity index χ for a tree representing 2,3-dimethylhexane given in Figure 1.

		1	2	3	4	5	6	7	8	
	1	0	1	0	0	0	0	0	0	$d(v_1)=1$
	2	1	0	1	0	0	0	1	0	$d(v_2)=3$
	3	0	1	0	1	0	0	0	1	$d(v_3)=3$
A =	4	0	0	1	0	1	0	0	0	$d(v_4)=2$
	5	0	0	0	1	0	1	0	0	$d(v_5)=2$
	6	0	0	0	0	1	0	0	0	$d(v_6)=7$
	7	0	1	0	0	0	0	0	0	$d(v_7)=1$
	8	0	0	1	0	0	0	0	0	$d(v_8)=1$

$$\chi = 3\,(1\cdot3)^{-0.5} + (3\cdot3)^{-0.5} + (3\cdot2)^{-0.5} + (2\cdot2)^{-0.5} + (2\cdot1)^{-0.5} = 3.6807$$

TABLE 2. Calculation of the edge-connectivity index ε for a tree representing 2,3-dimethylhexane given in Figure 1.

$$
{}^{E}\mathbf{A} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array}
\begin{array}{c}
\begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\
\left[\begin{array}{ccccccc}
0 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 0 & 1 & 1 \\
0 & 1 & 0 & 1 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 1 & 0 & 0 & 0 & 0
\end{array}\right]
\end{array}
\begin{array}{l}
d(e_1)=2 \\ d(e_2)=4 \\ d(e_3)=3 \\ d(e_4)=2 \\ d(e_5)=1 \\ d(e_6)=2 \\ d(e_7)=2
\end{array}
$$

$$\varepsilon = 3\,(2\cdot4)^{-0.5} + (2\cdot2)^{-0.5} + (4\cdot3)^{-0.5} + (3\cdot2)^{-0.5} + (2\cdot1)^{-0.5} = 3.3729$$

Distance sums can also be viewed as distance degrees [33]. F is given by:

$$F = E/(\mu + 1) \tag{6}$$

where E is the number of edges in a molecular graph and μ is its cyclomatic number which is equal to the minimum number of edges necessary to be removed from a polycyclic graph to convert it to a related acyclic graph [34]. Consequently, F=E for acyclic graphs. Since for isomeric acyclic graphs F=constant, the Balaban indices can be computed for isomers using the following simple expression and remembering that the number of number edges needs to be taken into account if non-isomeric acyclic structures are considered:

$$J = \sum_{\text{edges}} (d_i\, d_j)^{-0.5} \tag{7}$$

Note that the distance matrix **D** of a labeled connected graph is a symmetric V × V matrix whose elements are defined as [3,6,32]:

$$(D)_{ij} = \begin{cases} l_{ij} \text{ if } i \neq j \\ 0 \text{ if } i = j \end{cases} \tag{8}$$

where ℓ_{ij} is the length of the shortest path (i.e., the minimum number of edges) between the vertices i and j.

In Table 3 it is exemplified the computation of the Balaban index for a tree given in Figure 1.

TABLE 3. Calculation of the Balaban index J for a tree given in Figure 1.

		1	2	3	4	5	6	7	8	
	1	0	1	2	3	4	5	2	3	d_1=20
	2	1	0	1	2	3	4	1	2	d_2=14
	3	2	1	0	1	2	3	2	1	d_3=12
D =	4	3	2	1	0	1	2	3	2	d_4=14
	5	4	3	2	1	0	1	4	3	d_5=18
	6	5	4	3	2	1	0	5	4	d_6=24
	7	2	1	2	3	4	5	0	3	d_7=20
	8	3	2	1	2	3	4	3	0	d_8=18

$$J = 2\,(20 \cdot 14)^{-0.5} + 2\,(14 \cdot 12)^{-0.5} + (12 \cdot 18)^{-0.5} + (14 \cdot 18)^{-0.5} + (18 \cdot 24)^{-0.5}$$
$$= 0.4530$$

The Balaban index has many nice properties such as a very low degeneracy, an elegant extension to heterosystems and applicability in the QSPR/QSAR modeling [1,8,11,19,33]. However, there is one problem with the Balaban index. Unlike the connectivity index it gives greater weights to the inner (interior) CC bonds and smaller weights to the outer (terminal) CC bonds of an alkane. This appears to oppose intuitive reasoning that the outer more exposed bonds should have greater weights than inner bonds because the outer bonds are associated

with larger parts of molecular surface and consequently are expected to make a greater contribution to physicochemical properties. Because of this observation we investigated the behavior of the Balaban-like indices regarding the weights of outer and inner bonds.

BALABAN-LIKE TOPOLOGICAL INDICES

Balaban in his seminal paper [11] on the distance-sum connectivity index introduced a procedure which is quite general for deriving new indices: Any Balaban-like topological index (BTI) can be obtained using row-sums r_i and r_j of a given matrix in the inverse square root algorithm making contribution $(r_i \cdot r_j)^{-0.5}$ for the edge i-j.

$$BTI = F \sum_{edges} (r_i\, r_j)^{-0.5} \tag{9}$$

The use of F for acyclic, isomeric and polycyclic structures is discussed above.

Based on this procedure, a whole family of Balaban-like topological indices can be generated. In this sense the vertex- and edge-connectivity indices are also Balaban-like topological indices. Here we will report on several novel Balaban-like indices.

Index Based on the Complement of the Distance Matrix

The complement of the distance matrix, introduced by Randić [35], is a symmetric V x V matrix, called the distance-complement matrix and denoted by $^{c}\mathbf{D}$, that can be simply obtained from the distance matrix. The distance-complement matrix elements $(^{c}\mathbf{D})_{ij}$ can be expressed in terms of the distance matrix elements $(\mathbf{D})_{ij}$:

$$(^{c}D)_{ij} = \begin{cases} V - (D)_{ij} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases} \tag{10}$$

The corresponding index we call the complement Balaban index and denote it by ^{c}J. Randić and Pompe in their recent paper called this index the reversed Balaban index [36]. In

Table 4 we give as an example the computation of the complement Balaban index for a 2,3-dimethylhexane tree given in Figure 1.

TABLE 4. Calculation of the complement Balaban index ^cJ for a tree given in Figure 1.

$$^{c}\mathbf{D} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \end{array} \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{bmatrix} 0 & 7 & 6 & 5 & 4 & 3 & 6 & 5 \\ 7 & 0 & 7 & 6 & 5 & 4 & 7 & 6 \\ 6 & 7 & 0 & 7 & 6 & 5 & 6 & 7 \\ 5 & 6 & 7 & 0 & 7 & 6 & 5 & 6 \\ 4 & 5 & 6 & 7 & 0 & 7 & 4 & 5 \\ 3 & 4 & 5 & 6 & 7 & 0 & 3 & 4 \\ 6 & 7 & 6 & 5 & 4 & 3 & 0 & 5 \\ 5 & 6 & 7 & 6 & 5 & 4 & 5 & 0 \end{bmatrix} \end{array} \begin{array}{l} r_1=36 \\ r_2=42 \\ r_3=44 \\ r_4=42 \\ r_5=38 \\ r_6=32 \\ r_7=36 \\ r_8=38 \end{array}$$

$$^{c}J = 2\,(36\cdot42)^{-0.5} + 2\,(42\cdot44)^{-0.5} + (44\cdot38)^{-0.5} + (42\cdot38)^{-0.5} + (38\cdot32)^{-0.5}$$
$$= 0.1761$$

This index, unlike the Balaban index, gives in accordance with intuitive reasoning greater weights to the outer (terminal) CC bonds and smaller weights to the inner (interior) CC bonds of alkanes.

Index Based on the Reciprocal Distance Matrix

The reciprocal distance matrix, denoted by $^{r}\mathbf{D}$, was introduced independently by Plavšić *et al.* [37] and Balaban *et al.* [38]. It is a symmetric V × V matrix that can be simply obtained from the distance matrix **D** by replacing all distance matrix elements $(\mathbf{D})_{ij}$ by their reciprocals:

$$(^{r}D)_{ij} = \begin{cases} 1/(D)_{ij} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases} \tag{11}$$

The reciprocal distance matrix has been used to generate the Harary index in the same way as the Wiener index is generated from the distance matrix [5] and is so named in honor of Frank Harary, an important contributor to graph theory and chemical graph theory.

The corresponding index we call the Harary-Balaban index and denote it by ^{r}J. Randić and Pompe called this index the Harary-connectivity index [36]. In Table 5 we give as an example the computation of the Harary-Balaban index for a tree given in Figure 1.

It can be easily seen that the Harary-Balaban index gives also in accordance with intuitive reasoning greater weights to the outer (terminal) CC bonds and smaller weights to the inner (interior) CC bonds of alkanes.

Index Based on the Quotient Matrix $\mathbf{D/^{c}D}$

The quotient matrices have been introduced by Randić [39] and used by Plavšić *et al.* [40]. The quotient matrix $\mathbf{D/^{c}D}$ is a symmetric V x V matrix that can be simply obtained from the elements of the distance matrix $(\mathbf{D})_{ij}$:

$$(D/^{c}D)_{ij} = \begin{cases} (D)_{ij}/V - (D)_{ij} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases} \qquad (12)$$

This matrix is used to generate the Balaban-like index denoted by J' that we call the quotient Balaban index of the first kind. In Table 6 we exemplified the computation of the quotient Balaban index of the first kind J' for a tree given in Figure 1.

TABLE 5. Calculation of the Harary-Balaban index ^{r}J for a 2,3-dimethylhexane tree given in Figure 1.

$^{r}\mathbf{D} =$

	1	2	3	4	5	6	7	8	
1	0	1	0.5	0.33	0.25	0.2	0.5	0.33	r_1=3.12
2	1	0	1	0.5	0.33	0.25	1	0.5	r_2=4.58
3	0.5	1	0	1	0.5	0.33	0.5	1	r_3=4.83
4	0.33	0.5	1	0	1	0.5	0.33	0.5	r_4=4.17
5	0.25	0.33	0.5	1	0	1	0.25	0.33	r_5=3.67
6	0.2	0.25	0.33	0.5	1	0	0.2	0.25	r_6=2.73
7	0.5	1	0.5	0.33	0.25	0.2	0	0.33	r_7=3.12
8	0.33	0.5	1	0.5	0.33	0.25	0.33	0	r_8=3.25

$$^{r}J = 2\,(3.12 \cdot 4.58)^{-0.5} + (4.58 \cdot 4.83)^{-0.5} + (4.83 \cdot 4.17)^{-0.5} + (4.83 \cdot 3.25)^{-0.5} + (4.17 \cdot 3.67)^{-0.5} + (3.67 \cdot 2.73)^{-0.5}$$
$$= 1.7885$$

This index has lost the nice feature that had the previous two Balaban-like indices: It gives similarly to the original Balaban index counterintuitively greater weights to the inner (interior) CC bonds and smaller weights to the outer (terminal) CC bonds of alkanes.

TABLE 6. Calculation of the quotient Balaban index of the first kind J' for a 2,3-dimethylhexane tree given in Figure 1.

$$
\mathbf{D/^cD} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \end{array}
\begin{array}{c}
\begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\
\left[\begin{array}{cccccccc}
0 & 1.14 & 0.33 & 0.6 & 1 & 1.67 & 0.33 & 0.6 \\
1.14 & 0 & 0.14 & 0.33 & 0.6 & 1 & 0.14 & 0.33 \\
0.33 & 0.14 & 0 & 0.14 & 0.33 & 0.6 & 0.33 & 0.14 \\
0.6 & 0.33 & 0.14 & 0 & 0.14 & 0.33 & 0.6 & 0.33 \\
1 & 0.6 & 0.33 & 0.14 & 0 & 0.14 & 1 & 0.6 \\
1.67 & 1 & 0.6 & 0.33 & 0.14 & 0 & 1.67 & 1 \\
0.33 & 0.14 & 0.33 & 0.6 & 1 & 1.67 & 0 & 0.6 \\
0.6 & 0.33 & 0.14 & 0.33 & 0.6 & 1 & 0.6 & 0
\end{array}\right]
\end{array}
\begin{array}{l}
\\
r_1=4.68 \\
r_2=2.70 \\
r_3=2.03 \\
r_4=2.49 \\
r_5=3.72 \\
r_6=6.41 \\
r_7=4.68 \\
r_8=3.61
\end{array}
$$

$$
\begin{aligned}
J' &= 2\,(4.68 \cdot 2.70)^{-0.5} + (2.70 \cdot 2.03)^{-0.5} + (2.03 \cdot 2.49)^{-0.5} \\
&\quad + (2.03 \cdot 3.61)^{-0.5} + (2.49 \cdot 3.82)^{-0.5} + (3.82 \cdot 6.41)^{-0.5} \\
&= 2.3327
\end{aligned}
$$

Index Based on the Quotient Matrix $^cD/D$

The quotient matrix $^cD/D$ is a symmetric V x V matrix that can be simply obtained from the elements of the distance matrix $(D)_{ij}$:

$$
(D/^cD)_{ij} = \begin{cases} V/(D)_{ij} - 1 & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases} \tag{13}
$$

Therefore, the elements of the quotient matrix $^cD/D$ can be obtained from the elements of the reciprocal distance matrix ^rD:

$$
(^cD/D)_{ij} = V\,(^rD)_{ij} - 1 \tag{14}
$$

However, the simplest way of constructing the quotient matrix ^{c}**D/D** is by taking the reciprocal values of the elements of the quotient matrix **D/**^{c}**D**:

$$(^{c}\mathbf{D/D})_{ij} = 1/(\mathbf{D/}^{c}\mathbf{D})_{ij} \tag{15}$$

The quotient matrix ^{c}**D/D** is used to generate the Balaban-like index denoted by J" that we call the quotient Balaban index of the second kind. In Table 7 we exemplify the computation of the quotient Balaban index of the second kind J" for a 2,3-dimethylhexane tree given in Figure 1.

TABLE 7. Calculation of the quotient Balaban index of the second kind J" for a tree given in Figure 1.

$^{c}\mathbf{D/D} =$

	1	2	3	4	5	6	7	8	
1	0	7	3	1.67	1	0.6	3	1.67	r_1=17.93
2	7	0	7	3	1.67	1	7	3	r_2=29.67
3	3	7	0	7	3	1.67	3	7	r_3=31.67
4	1.67	3	7	0	7	3	1.67	3	r_4=26.33
5	1	1.67	3	7	0	7	1	1.67	r_5=22.33
6	0.6	1	1.67	3	7	0	0.6	1	r_6=14.87
7	3	7	3	1.67	1	0.6	0	1.67	r_7=17.93
8	1.67	3	7	3	1.67	1	1.67	0	r_8=19.00

$$J'' = 2\,(17.93 \cdot 29.67)^{-0.5} + (29.67 \cdot 31.67)^{-0.5} + (31.67 \cdot 26.33)^{-0.5} + (31.67 \cdot 19.00)^{-0.5} + (26.33 \cdot 22.33)^{-0.5} + (22.33 \cdot 14.87)^{-0.5} = 0.2909$$

This index has regain the nice feature that have the complement Balaban index ^{c}J and the Harary-Balaban index ^{r}J, that is, it gives in accordance with intuitive reasoning greater weights to the outer (terminal) CC bonds and smaller weights to the inner (interior) CC bonds of alkanes.

THE STRUCTURE-MOTOR OCTANE NUMBER MODELING USING A SINGLE BALABAN-LIKE INDEX

In Figure 2 we give trees corresponding to octane isomers and in Table 8 we list the Balaban-like indices, vertex- and edge-connectivity indices and motor octane numbers of isomeric octanes. Octane numbers are related to the ability of alkanes to form radicals by cracking a high pressures and temperatures [41] and represent an important indicator for the quality of an alkane as fuel.

The linear and quadratic regressions between the motor octane numbers and considered topological indices for octanes were carried out. In Tables 9 and 10 we summarize the regression statistics for all seven topological indices listed in Table 8.

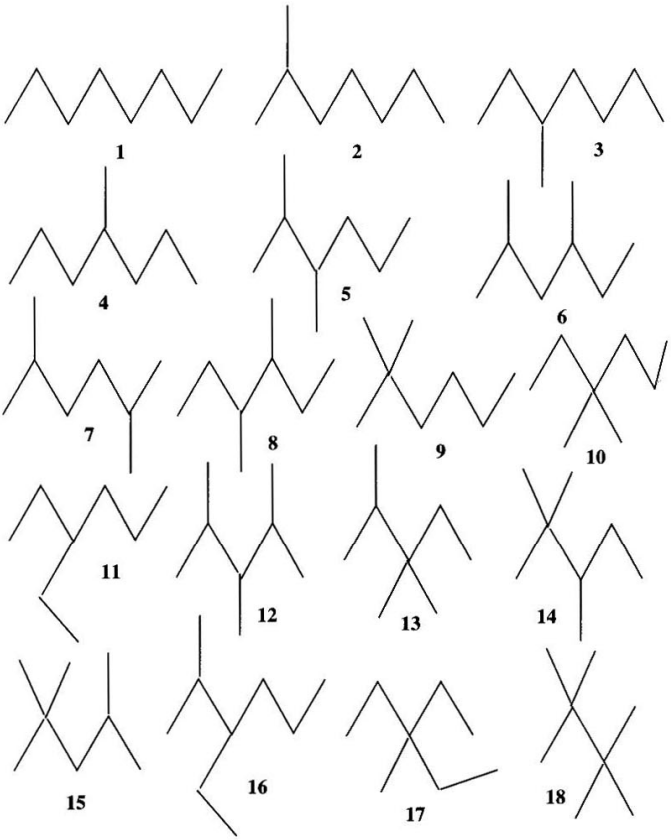


FIGURE 2. Trees corresponding the carbon skeletons of isomeric octanes.

TABLE 8. The Balaban-like indices, vertex- and edge-connectivity indices and motor octane numbers of isomeric octanes.

Octane[a]	χ	ε	J	^{c}J	^{r}J	J'	J''	MON[b]
n-octane	3.914	3.414	0.361	0.197	1.997	1.442	0.335	-
2-methylheptane	3.770	3.432	0.388	0.189	1.910	1.704	0.316	23.8
3-methylheptane	3.808	3.342	0.409	0.184	1.885	1.912	0.311	35.0
4-methylheptane	3.808	3.380	0.417	0.183	1.879	2.002	0.310	39.0
2,3-dimethylhexane	3.681	3.373	0.453	0.176	1.789	2.333	0.291	78.9
2,4-dimethylhexane	3.664	3.377	0.443	0.177	1.798	2.245	0.293	69.9
2,5-dimethylhexane	3.626	3.449	0.418	0.181	1.823	2.005	0.297	55.7
3,4-dimethylhexane	3.719	3.256	0.470	0.174	1.768	2.510	0.287	81.7
2,2-dimethylhexane	3.561	3.427	0.445	0.178	1.774	2.208	0.288	77.4
3,3-dimethylhexane	3.621	3.299	0.482	0.173	1.737	2.590	0.281	83.4
3-ethylhexane	3.846	3.270	0.439	0.179	1.851	2.250	0.304	52.4
2,3,4-trimethylpentane	3.553	3.372	0.495	0.170	1.700	2.734	0.272	95.9
2,3,3-trimethylpentane	3.504	3.283	0.530	0.166	1.646	3.042	0.262	99.4
2,2,3-trimethylpentane	3.481	3.334	0.518	0.167	1.658	2.919	0.264	99.9
2,2,4-trimethylpentane	3.417	3.471	0.484	0.171	1.690	2.604	0.270	100.0
3-ethyl-2-methylpentane	3.719	3.272	0.480	0.172	1.760	2.622	0.285	88.1
3-ethyl-3-methylpentane	3.682	3.116	0.512	0.169	1.703	2.903	0.274	88.7
2,2,3,3-tetramethylbutane	3.250	3.414	0.574	0.162	1.549	3.427	0.242	-

[a]Octane trees are given in Figure 2;

[b]MON = motor octane numbers are taken from Balaban [33].

TABLE 9. The regression statistics for seven mono-parametric linear structure-motor octane number models for octanes. Symbols used are TI = topological index, R = correlation coefficient, S = standard error of estimate and F = Fisher ratio.

TI	R	S	F
χ	0.778	16.0	21.5
ε	0.271	24.6	1.1
J	0.928	9.5	86.4
^{c}J	0.966	6.6	193.1
^{r}J	0.963	6.9	180.6
J'	0.921	10.0	178.0
J''	0.965	6.7	190.0

TABLE 10. The regression statistics for seven quadratic models for octanes.

TI	R	S	F
χ	0.795	16.1	11.2
ε	0.275	25.5	0.5
J	0.963	7.1	83.6
^{c}J	0.974	6.1	118.0
^{r}J	0.990	3.7	335.2
J'	0.951	8.2	61.5
J"	0.989	3.8	303.4

The best linear regression involves the complement Balaban index ^{c}J, but regressions with the Harary-Balaban index ^{r}J and the quotient Balaban index of the second kind J" are close to it. A very poor model is obtained with the edge-connectivity index and a somewhat better model is obtained with the vertex-connectivity index. Between these two sets of models are models based on the original Balaban index J and the quotient Balaban index of the first kind J'. The best quadratic models are based on the Harary-Balaban index ^{r}J and the quotient Balaban index of the second kind J". They are far the best stricture-motor octane number models obtained. The rest of quadratic models are slightly better than the corresponding linear models.

VARIABLE BALABAN-LIKE INDICES

We also decided to investigate whether the use of variable Balaban-like indices leads to better models. The variable Balaban-like indices are based on augmented distance matrices. The augmented distance matrices can be obtained from the standard distance matrices by replacing diagonal elements, which are equal to zero, with the variable x [36,42]. The variable x depends on a physicochemical property considered [43]. The optimum value of x will be attained when the standard error of estimate will be at minimum. In Table 11 we give the augmented distance matrix for 2,3-dimethylhexane and the variable Balaban index. We denote

the variable topological indices with superscript f as suggested by Randić [36,42,43] since they are functions of variable x.

We considered all Balaban-like topological indices in their variable form. The resulting linear models are only slightly better than the corresponding models based on fixed indices. The regression statistics are summarized in Table 12.

TABLE 11. Calculation of the variable Balaban index J^f for the 2,3-dimethylhexane tree given in Figure 1.

$$\mathbf{D} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \end{array} \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{bmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 2 & 3 \\ 1 & 0 & 1 & 2 & 3 & 4 & 1 & 2 \\ 2 & 1 & 0 & 1 & 2 & 3 & 2 & 1 \\ 3 & 2 & 1 & 0 & 1 & 2 & 3 & 2 \\ 4 & 3 & 2 & 1 & 0 & 1 & 4 & 3 \\ 5 & 4 & 3 & 2 & 1 & 0 & 5 & 4 \\ 2 & 1 & 2 & 3 & 4 & 5 & 0 & 3 \\ 3 & 2 & 1 & 2 & 3 & 4 & 3 & 0 \end{bmatrix} \end{array} \begin{array}{l} \\ d_1=20 \\ d_2=14 \\ d_3=12 \\ d_4=14 \\ d_5=18 \\ d_6=24 \\ d_7=20 \\ d_8=18 \end{array}$$

$$J = 2\,(20\cdot14)^{-0.5} + 2\,(14\cdot12)^{-0.5} + (12\cdot18)^{-0.5} + (14\cdot18)^{-0.5} + (18\cdot24)^{-0.5}$$
$$= 0.4530$$

TABLE 12. The regression statistics for five mono-parametric linear structure-motor octane number models for octanes based on the variable Balaban-like indices.

TI	R	S	F
J^f x = 130	0.958	7.3	156.6
$^cJ^f$ x = - 24	0.974	5.8	260.0
$^rJ^f$ x = - 2	0.967	6.5	201.5
$(J')^f$ x = 100	0.965	6.6	195.1
$(J'')^f$ x = - 9	0.967	6.5	202.0

Results reported in Table 12 parallel results in Table 9, that is, the best linear regression involves the variable complement Balaban index ${}^{c}J^{f}$, but regressions with the variable Harary-Balaban index ${}^{r}J^{f}$ and the variable quotient Balaban index of the second kind $(J'')^{f}$ are close to it. However, the improvement is really negligible, and if one takes into account a labor needed to carry out the optimization of x, there is no advantage in this case to use the variable Balaban-like indices.

We also carried out the linear regression analysis with the squares of all Balaban-like indices computed using the optimum values of x (see Table 12). The improved models were obtained only for the variable Haray-Balaban index ${}^{r}J^{f}$ (x=-2) (R=0.974, S=5.8, F=259.4) and the variable quotient Balaban index of the second kind $(J'')^{f}$ (x=- 9) ((R=0.974, S=5.8, F=259.6).

Finally, we carried out the quadratic regression with all Balaban-like indices computed using the optimum values of x given in Table 12. The quadratic models with the variable Balaban index J^{f} (x=130) (R=0.967, S=6.7, F=94.2), the variable Harary-Balaban index ${}^{r}J^{f}$ (x=-2) (R=0.983, S=4.8, F=189.2) and the variable quotient Balaban index of the second kind $(J'')^{f}$ (x=- 9) (R=0.983, S=45.8, F=189.8) show improvement in the values of R and S, but the F values are much lower than for the linear models. The quadratic models with the variable Haray-Balaban index ${}^{r}J^{f}$ (x=-2) and the variable quotient Balaban index of the second kind $(J'')^{f}$ (x=- 9) are still poorer than the quadratic models with the corresponding fixed indices.

COMPARISON WITH THE STRUCTURE-MOTOR OCTANE NUMBER MODELS FROM THE LITERATURE

There are several structure-(research and motor) octane number models reported in the literature [33, 44-48]. In Table 13 we present statistical characteristics of some of these models in the manner they were reported.

TABLE 13. Statistical characteristics of several linear structure-motor octane number models for octanes taken from the literature.

TI	R	S	Ref.
χ	0.745	24.7	33
J	0.932	24.7	33
W^a	0.957	7.4	45
$\lambda_1{}^b$	0.925	9.7	45
J	0.931	9.3	45
$a_N{}^c$	0.917	10.2	45
GAI^d	0.907	10.8	45

[a] W = Wiener number [10]; [b] λ_1 = largest eigenvalue of the vertex-adjacency matrix [3];

[c] a_N = determinant of the vertex-adjacency matrix of the derivative graph [45];

[d] GAI = generalized a_N index [45].

Among the models listed in Table 13, the model based on the Wiener index is comparable to our model based on the quadratic model with the Balaban index J. It is better than our several models (e.g., the linear and quadratic models with the vertex- and edge-connectivity indices, the linear models with the Balaban index and the quotient Balaban index of the first kind J', the linear model with the variable Wiener index), and but worse than many of models presented here (e.g., the linear and quadratic models based on complement Balaban index ^cJ, Harary-Balaban index ^rJ and the quotient Balaban index of the second kind J'', the linear models based on the variable complement Balaban index $^cJ^f$, Harary-Balaban index $^rJ^f$ and the quotient Balaban indices of the first and second kind $(J')^f$, $(J'')^f$). All other models in Table 13 are inferior to most of our models.

CONCLUDING REMARKS

A suite of structure-motor octane number models based on the Balaban index, Balaban-like topological indices (the complement Balaban index, the Harary-Balaban index, the quotient Balaban indices of the first and second kind), their variable counterparts, and vertex-

and edge-connectivity indices are compared among themselves and with several models from the literature. Variable indices produced in all cases slightly better linear models than fixed indices. All linear models found in the literature, but the structure-motor octane number model based on the Wiener index, are much poorer than most models produced here. The best models obtained are quadratic models with the Harary-Balaban index and the quotient Balaban index of the second kind. This is very gratifying because these two Balaban-like indices give greater weights to the outer (terminal) CC bonds and smaller weights to the inner (interior) CC bonds of alkanes. This agrees with the intuitive reasoning according to which the outer more exposed bonds are associated with larger parts of molecular surface and consequently expected to make a greater contribution to physicochemical properties than the inner, less exposed, CC bonds. Therefore, the structure-property modeling with the Harary-Balaban index and the quotient Balaban index of the second kind and their variable counterparts warrants a further study using a variety of physicochemical properties of diverse chemicals to establish the range of their applicability.

Acknowledgment. This research was supported by the Ministry of Science and Technology of the Republic of Croatia through Grant Nos. 00980606 (SN, NT) and 00980607 (DP).

REFERENCES

[1] A. T. Balaban, Confessions and Reflections of a Graph-Theoretical Chemist, *Commun. Math. Chem.* (*MATCH*), **1993**, 29, 3-17.

[2] R. Todeschini and V. Consonni, *The Handbook of Molecular Descriptors*, Wiley - VCH, New York, 2000.

[3] N. Trinajstić, *Chemical Graph Theory*, 2nd revised ed., CRC Press, Boca Raton, FL, 1992.

[4] M. Karelson, *Molecular Descriptors in QSAR/QSPR*, Wiley-Interscience, New York, 2000.

[5] H. Hosoya, Topological Index. A Newly Proposed Quantity Characterizing the Topological Nature of Structural Isomers of Saturated Hydrocarbons, *Bull. Chem. Soc. Jpn.* **1971**, *44*, 2332-2339.

[6] J. Devillers and A. T. Balaban, Eds., *Topological Indices and Related Descriptors in QSAR and QSPR*, Gordon & Breach Sci. Publ., Amsterdam, 1999.

[7] H. Kubinyi, Quantitative Structure-Activity Relationships in Drug Design. In: *The Encyclopedia of Computational Chemistry*, Eds.: P. von R. Schleyer, N.L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer III and P. R. Schreiner. Wiley, Chichester, 1999, pp. 2309-2319.

[8] G. Grassy, B. Calas, A. Yasri, R. Lahana, J. Woo, S. Iyer, M. Kazcorek, R. Floc'h and R. Buelow, Computer Assisted Rational Design of Immunosuppressive Compounds, *Nature Biotehnol.* **1998**, *16*, 748-752.

[9] M. Randić, On Characterization of Molecular Branching, *J. Amer. Chem. Soc.* **1975**, *97*, 6609-6615.

[10] H. Wiener, Structural Determination of Paraffin Boiling Points, *J. Amer. Chem. Soc.* **1947**, *69*, 17-20.

[11] A. T. Balaban, Highly Discriminating Distance-Based Topological Index, *Chem. Phys. Lett.* **1982**, *89*, 399-404.

[12] S. C. Basak, B. D. Gute and G. D. Grunwald, Estimation of the Normal Boiling Points of Haloalkanes Using Molecular Similarity, *Croat. Chem. Acta*, **1996**, *69*, 1159-1173.

[13] S.C. Basak, G.D. Grunwald and G.J. Niemi, Use of Graph-Theoretic and Geometrical Molecular Descriptors in Structure-Activity Relationships. In: *From Chemical Topology to Three-Dimensional Geometry*, Ed.: A.T. Balaban. Plenum Press, New York, 1997, pp. 73-116.

[14] S. C. Basak, B. D. Gute and G. D. Grunwald, A Hierarchical Approach to the Development of QSAR Models using Topological, Geometrical and Quantum Chemical Parameters. In: *Topological Indices and Related Descriptors in QSAR and QSPR*, Eds.: J. Devillers and A.T. Balaban. Gordon and Breach Science Publishers, The Netherlands, 1999, pp. 675-696.

[15] S. C. Basak, B. D. Gute, B. Lučić, S. Nikolić and N. Trinajstić, A Comparative QSAR Study of Benzamidines Complement-Inhibitory Activity and Benzene Derivatives Acute Toxicity, *Comput. Chem.* **2000**, *24*, 181-191.

[16] S. C. Basak, D. K. Harriss and V. R. Magnuson, POLLY 2.3, Copyright of the University of Minnesota, 1988.

[17] A. R. Katritzky, V. Lobanov and M. Karelson,.CODESSA (COmprehensive DEscriptors for Structural and Statistical Analysis): Reference Manual - Version 2, University of Florida, Gainesville, FL., 1996. Professor Katritzky said recently during his lecture (June 2) at the Second Indo-US Workshop on Mathematical Chemistry (Duluth: May 30 - June 3, 2000) that CODESSA will soon available on web.

[18] Vedrina, S. Marković, M. Medić-Šarić and N. Trinajstić, TAM: A Program for the Calculation of Topological Indices in QSPR and QSAR Studies, *Comput. Chem.*, **1997**, *21*, 355-361. TAM is available on request from Professor Marica Medić-Šarić.

[19] A. T. Balaban and O. Ivanciuc, FORTRAN 77 Computer Program for Calculating the Topological Index J for Molecules Containing Heteroatoms. In: *MATH/CHEM/COMP* **1988**, Ed.: A. Graovac. Elsevier, Amsterdam, 1989, pp. 193-212.

[20] J. A. Bondy and U.S.R. Murty, *Graph Theory with Applications*, North Holland/Elsevier, Amsterdam, 19786.

[21] L. B. Kier, L. H. Hall, W. J. Murray and M. Randić, Molecular Connectivity I: Relationship to Nonspecific Local Anesthesia, *J. Pharm. Sci.*, **1975**, *64*, 1971-1974.

[22] N. Trinajstić, S. Nikolić, D. Babić and Z. Mihalić, The Vertex- and Edge-Connectivity Indices of Platonic and Archimedean Molecules. *Bull. Chem. Technol. Macedonia*, **1997**, *16*, 43-51.

[23] S. Nikolić, N: Trinajstić and I. Baučić, Comparison between the Vertex- and Edge-Connectivity Indices for Benzenoid Hydrocarbons. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 42-46.

[24] E. Estrada, Edge Adjacency Relationships and a Novel Topological Index Related to Molecular Volume. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 31-33.

[25] L. B. Kier and L. H. Hall, *Molecular Connectivity in Chemistry and Drug Design*. Academic Press, New York, 1976.

[26] L. B. Kier and L. H. Hall, *Molecular Connectivity in Structure-Activity Analysis*, Research Studies Press Ltd., Letchworth, England, 1986.

[27] E. Estrada, Edge Adjacency Relationships in Molecular Graphs Containing Heteroatoms: A New Topological Index Related to Molecular Volume. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 701-707.

[28] G. G. Cash, Correlation of Physicochemical Properties of Alkylphenols with Their Graph-Theoretical ε Parameter. *Chemosphere*, **1995**, *31*, 4307-4315.

[29] E. Estrada and L. Rodriguez, Edge-Connectivity Indices in QSPR/QSAR Studies. 1. Comparison to other Topological Indices in QSPR Studies. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 1037-1041.

[30] E. Estrada, Edge-Connectivity Indices in QSPR/QSAR Studies. 2. Accounting for Long Range Bond Contributions. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 1042-1048.

[31] S. C. Basak, S. Nikolić, N. Trinajstić, D. Amić and D. Bešlo, QSPR Modeling: Graph Connectivity Indices versus Line Graph Connectivity Indices. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 927-933.

[32] Z. Mihalić, D. Veljan, D. Amić, S. Nikolić, D. Plavšić and N. Trinajstić, The Distance Matrix in Chemistry, *J. Math. Chem.* **1992**, *11*, 223-258.

[33] A. T. Balaban, Topological Indices Based on Topological Distances in Molecular Graphs, *Pure Appl. Chem.* **1983**, *55*, 199-206.

[34] R. J. Wilson, *Introduction to Graph Theory*, Oliver & Boyd, Edinburgh, 1972, p. 46.

[35] M. Randić, Linear Combinations of Path Numbers as Molecular Descriptors. *New J. Chem.* **1997**, *21*, 945-951.

[36] M. Randić and M. Pompe, The Variable Molecular Descriptors Based on Distance Related Matrices, (reprint).

[37] D. Plavšić, S. Nikolić, N. Trinajstić and Z. Mihalić, On the Harary Index for Characterization of Chemical Graphs. *J. Math. Chem.* **1993**, *12*, 235-250.

[38] O. Ivanciuc, T. S. Balaban and A. T. Balaban, Design of Topological Indices. Part 4. Reciprocal Distance Matrix, Related Local Vertex Invariants and Topological Indices. *J. Math. Chem.* **1993**, *12*, 309-318.

[39] M. Randić, On Characterization of Cyclic Structures, *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 1063-1071.

[40] D. Plavšić, N. Trinajstić, D. Amić and M. Šoškić, Comparison between the Structure-Boiling Point Relationships with Different Descriptors for Condensed Benzenoids. *New J. Chem.* **1998**, *22*, 1075-1078.

[41] Kirk-Othmer, *Encyclopedia of Chemical Technology*, Wiley, New York, 1994, Vol. 12, p. 348.

[42] M. Randić and S.C. Basak, Variable Molecular Descriptors, (reprint).

[43] M. Randić and S.C. Basak, Construction of High-Quality Structure Property-Activity Regressions: The Boiling Points of Sulfides, *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 899-905.

[44] A. T. Balaban, L. B. Kier and N. Joshi, Structure-Property Analysis of Numbers for Hydrocarbons (Alkanes, Cycloalkanes and Alkenes), *Commun. Math Chem.* (*MATCH*), **1992**, *28*, 13-27.

[45] L. Xu, H.-Y. Wang and Q. Su, A Newly Proposed Molecular Topological Index for the Discrimination of *Cis/Trans* Isomers and for the Studies of QSAR/QSPR, *Comput. Chem.* **1992**, *16*, 187-194.

[46] J. M. Andrade, S. Muniategui, P. López and D. Prada, Costs, Laboratory Safety, Productivity and Faster Research Octane Number and Motor Octane Number Determinations in Industrial Chemistry Laboratories, *Analyst*, **1995**, *120*, 249-253.

[47] E. S. Blurock, Automatic Learning of Chemical Concepts: Research Octane Numbers and Molecular Substructures, *Comput. Chem.* **1995**, *19*, 91-99.

[48] I. Gutman, W. Linert, I. Lukovits and Ž. Tomović, The Multiplicative Version of the Wiener Index, *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 113-116.