

The Variable Molecular Descriptors Based on Distance Related Matrices[†]

Milan Randić^{*,‡,§} and Matevž Pompe[#]

National Institute of Chemistry, Hajdrihova 19, 1000 Ljubljana, Slovenia, and Department of Chemistry and Chemical Technology, Aškerčeva 5, The University of Ljubljana, 1000 Ljubljana, Slovenia

Received July 22, 2000

Recently variable molecular connectivity index and variable paths have been tested as molecular descriptors in several structure–property regressions. Here we outline the construction of several variable molecular descriptors, derived from the distance matrix and the “reversed” distance matrix. This includes the variable Balaban J index and the “reversed” Balaban index $1/J$ as well as a novel index $1/JJ$ derived from J and $1/J$. All the variable descriptors mentioned were constructed by augmenting the distance matrix by replacing the diagonal zeroes with the variables x , y , z , ...

INTRODUCTION

Mathematical descriptors derived from molecular graphs have been widely used in quantitative structure–property and structure–activity relationships.^{1–3} They are often known as topological indices (TIs) and can assume integer values or real number values. The former are often a consequence of some structural enumeration, like the Wiener number,⁴ which in acyclic graphs counts all the separations between any pair of vertices in a graph, or the Hosoya Z index,⁵ which counts all possible disjoint bonds in a structure. Representative real number TIs are the connectivity index⁶ ($^1\chi$) and Balaban's J index.⁷ As Balaban has pointed out the connectivity index can be constructed from the rows of the adjacency matrix using the algorithm $1/\sqrt{(R_i R_j)}$.⁷ Here R_i and R_j are the row sums of the adjacency matrix, and summation is made over all bonds (that is, i, j are labels of adjacent vertices). In the case of the adjacency matrix the row sums give the graph theoretical valence, which have the physical meaning of differentiating the primary, the secondary, the tertiary, and the quaternary carbon atoms. From the mathematical point of view the maximal and the minimal row sums, according to the well-known Frobenius-Perron theorem⁸ of Linear Algebra, represent the upper and the lower bound on the leading eigenvalue of (adjacency) matrix. On the other hand, according to Lovasz and Pelikan,⁹ the leading eigenvalue of the adjacency matrix can be viewed as an index of molecular branching. When the diagonal entries of the adjacency matrix are augmented (vide infra) by a variable parameter, one alters the relative roles of the primary, the secondary, the tertiary, and the quaternary carbon atoms, relative to each other.

Balaban's J index is similarly obtained from the row sums of the distance matrix using the same procedure. The simple

interpretation of the rows of the matrix has been lost because the distance matrix no longer represents the initial graph but a complete graph with weighted edges. Nevertheless, the extreme row sums remain to signify the upper and the lower bound on the leading eigenvalues of the distance matrix. However, the leading eigenvalue of the distance matrix has not yet found plausible structural interpretation, although the so-called D/D matrix,¹⁰ the elements of which are constructed using elements of the distance matrix, has been interpreted as an index of the degree of folding of a molecule. In the present paper we will ignore the constant factor $q/(1 + \mu)$ which Balaban introduced when calculating J values, because we will consider only acyclic systems (for which $\mu = 0$), and will consider octane isomers, which have the same number of edges, thus q is constant and does not influence regressions.

In the case of heteroatoms the “simple” (hydrocarbon) TIs have to be modified as outlined by Kier and Hall^{11,12} for the connectivity indices and by Balaban and collaborators¹³ for the J index. However, as has been recently pointed out¹⁴ the proposed “modifications” may be suitable for some properties and may make the modified descriptors even less suitable than the corresponding “simple” (hydrocarbon) descriptors in other applications. This came to light with the development of the variable molecular descriptors,¹⁵ the variable connectivity index,^{15–21} and the variable path numbers.^{22–24}

VARIABLE MOLECULAR DESCRIPTORS

Variable connectivity indices have been proposed as an alternative route for optimal characterization of heteroatoms in molecules. Along the main diagonal of the adjacency matrix one replaces the zero entries by variables x , y , z , ... for each type of (hetero)atom and then one uses the row sums to obtain contributions for individual bonds using the same algorithm $1/\sqrt{(R_i R_j)}$. In Table 1 we present the augmented adjacency matrix for 2-pentanol (assuming the standard numbering of carbon atoms) and the calculation of the variable connectivity index $^1\chi$ for 2-pentanol. As we see the resulting expression is a function of two variables, x , y , which describe the distinct contributions of carbon atoms and oxygen in alcohols. Hence, $^1\chi$ is a function of two variables,

* Corresponding author phone: (515)292-7411; fax: (515)292-8629; e-mail: milan.randic@ki.si.

[†] On the occasion of his 70th anniversary this paper is dedicated to Professor A. T. Balaban, one of the few chemists of today who continues to share his time between the theoretical research in mathematical chemistry and the synthesis of novel organic compounds.

[‡] National Institute of Chemistry.

[§] The University of Ljubljana.

[#] On leave from the Department of Mathematics & Computer Science, Drake University, Des Moines, IA 50311.

Table 1. Augmented Adjacency Matrix and the Corresponding Row Sums for 2-Pentanol and the Variable Connectivity Index ${}^1\chi^f$ ^a

	1	2	3	4	5	6	row sum
1	<i>x</i>	1	0	0	0	0	1 + <i>x</i>
2	1	<i>x</i>	1	0	0	1	3 + <i>x</i>
3	0	1	<i>x</i>	1	0	0	2 + <i>x</i>
4	0	0	1	<i>x</i>	1	0	2 + <i>x</i>
5	0	0	0	1	<i>x</i>	0	1 + <i>x</i>
6	0	1	0	0	0	<i>y</i>	1 + <i>y</i>

^a The variable connectivity Index ${}^1\chi^f$: ${}^1\chi^f = 1/\sqrt{(1+x)(3+x)} + 1/\sqrt{(3+x)(2+x)} + 1/(2+x) + 1/\sqrt{(2+x)(1+x)} + 1/\sqrt{(3+x)(1+y)} = f(x, y)$.

Table 2. Augmented Distance Matrix and the Corresponding Row Sums for 2-Pentanol and the Variable Balaban Index J^* ^a

	1	2	3	4	5	6	row sum
1	<i>x</i>	1	2	3	4	2	12 + <i>x</i>
2	1	<i>x</i>	1	2	3	1	8 + <i>x</i>
3	2	1	<i>x</i>	1	2	2	8 + <i>x</i>
4	3	2	1	<i>x</i>	1	3	10 + <i>x</i>
5	4	3	2	1	<i>x</i>	4	14 + <i>x</i>
6	2	1	2	3	4	<i>y</i>	12 + <i>y</i>

^a The variable Balaban index J^* : $J^* = 1/\sqrt{(12+x)(8+x)} + 1/(8+x) + 1/\sqrt{(8+x)(10+x)} + 1/\sqrt{(10+x)(14+x)} + 1/\sqrt{(8+x)(12+y)} = f(x, y)$.

$f(x, y)$, and the superscript f is a reminder of two properties of variable descriptors: ${}^1\chi^f$ to be a function and to have flexibility. The values of the variables x, y are determined during the regression so that the standard error for property considered is as small as possible. It is therefore possible, and in fact ought to be anticipated, that the values of the variables x and y may critically depend on the property considered.

In this paper we will consider several structurally related variable connectivity indices obtained from the distance matrix and modified distance matrices. First, we consider the variable Balaban's J index designated as J^* , which can be obtained from the row sums of the distance matrix in which along the main diagonal variables x, y, \dots replace zeroes. In the case of alcohols the variable x will modify the graph theoretical valence of carbon atoms and the variable y will modify the graph theoretical valence of oxygen atom. In Table 2 is shown the augmented distance matrix for 2-pentanol as well as the expression for the variable J^* index for 2-pentanol. We obtained for the topological index a function of two variables, the value of which is to be determined when considering a particular structure–property application.

The flexibility of the variable descriptors is not only in that they will minimize the standard error for a particular structure–property–activity study but also that they will adapt and assume different optimal values when different molecular properties are considered. In Table 3 we show the J^* values for the 18 isomers of octane for five different values of the variable x (characterizing carbon atoms): $x = -5$; $x = -2$; $x = -1$; $x = +5$; and $x = +10$. We included also the $x = 0$ case, which gives the Balaban's J index as originally proposed. As we see from Table 3 for the negative values of x , the magnitudes of J^* index increase relative to J , and for the positive values of x the opposite is the case. What is more important, however, is that the relative

Table 3. Variations of J^* for Octane Isomers as x Varies from -5 to $+10$

isomer	$x = -5$	$x = -2$	$x = -1$	$x = 0$	$x = +5$	$x = +10$
<i>n</i>	.49385 A	.40457	.38176	.36144 A	.28588	.23676
2M	.54413 B	.43795	.41141	.38798 B	.30241	.24808
3M	.58795 C	.46510	.43512	.40887 C	.31470	.25619
4M	.60665 E	.47605	.44456	.41709 D	.31934	.25918
2,5M	.60463 D	.47668	.44552	.41826 E	.32074	.26041
3E	.65624 F	.50531	.46985	.43920 F	.33204	.26747
2,4M	.65951 G	.50911	.47352	.44269 G	.33462	.26937
2,2M	.66455 H	.51175	.47572	.44454 H	.33551	.26987
2,3M	.68502 I	.52318	.48547	.45297 I	.34018	.27285
3,4M	.72669 J	.54667	.50555	.47035 J	.34981	.27900
2M3E	.74963 K	.55892	.51592	.47927 K	.35466	.28207
3,3M	.76878 M	.56291	.51915	.48191 L	.35582	.28269
2,2,4M	.75546 L	.56454	.52117	.48413 M	.35795	.28438
2,3,4M	.78449 N	.57958	.53380	.49489 N	.36360	.28788
3M3E	.84025 O	.60480	.55430	.51189 O	.37170	.29259
2,2,3M	.84623 P	.61146	.56051	.51761 P	.37546	.29517
2,3,3M	.88351 Q	.62900	.57497	.52976 Q	.38157	.29887
2233M	1.0090 R	.69195	.62751	.57434 R	.40473	.31312

Table 4. Augmented “Reversed” Distance Matrix and the Corresponding Row Sums for 2-Pentanol and the “Reversed” Variable Balaban Index $1/J^*$ ^a

	1	2	3	4	5	6	row sum
1	<i>x</i>	5	4	3	2	4	18 + <i>x</i>
2	5	<i>x</i>	5	4	3	5	22 + <i>x</i>
3	4	5	<i>x</i>	5	4	4	22 + <i>x</i>
4	3	4	5	<i>x</i>	5	3	20 + <i>x</i>
5	2	3	4	5	<i>x</i>	2	16 + <i>x</i>
6	4	5	4	3	2	<i>y</i>	18 + <i>y</i>

^a The “reversed” variable Balaban index $1/J^*$: $1/J^* = 1/\sqrt{(18+x)(22+x)} + 1/(22+x) + 1/\sqrt{(22+x)(20+x)} + 1/\sqrt{(20+x)(16+x)} + 1/\sqrt{(22+x)(18+y)} = f(x, y)$.

magnitudes of the variable J^* index change as x changes. For example, the quotient of J^* for *n*-octane and 2-methylheptane is 0.90760 when $x = -5$ and 0.95437 when $x = +0.10$. Although these changes may appear small, this minor variation of the variable descriptors can reduce the standard error sometime by a factor of 2 or more.

It is also of interest to observe that ordering of isomers based on the relative magnitudes of the variable index J^* may depend on the value of the variable x . In Table 3 we have ordered isomers by the increasing values of the index J ($x = 0$) and we assigned labels A–R to the 18 octanes. As we see the ordering A, B, C, D, E, F, ... holds for cases $x = -2$ to $x = +10$, but in the case $x = -5$ the alphabetic order has been perturbed in positions involving isomers D and L.

THE REVERSE DISTANCE MATRIX

The distance matrix has an apparent conceptual disadvantage: One expects that as the distance between atoms in a molecule increases the role of more distant neighbors will decrease. But in the row sums of the distance matrix more distant neighbors make larger contributions. One way to deal with this was introduction of the inverse distance matrix by Plavšić et al.²⁵ and Ivanciuc et al.²⁶ An alternative route is to consider “reversed” weights, giving to the shortest distances the larger weight.²⁷ In Table 4 we show the reverse distance matrix with augmented diagonal elements that have variables as entries. The corresponding row sums are shown because in construction of J index the row sums are used in the computation. The entries in the “reverse” distance matrix

Table 5. Augmented $1/[\text{Distance}]$ Matrix and the Corresponding Row Sums for 2-Pentanol and the Variable Harary-Connectivity Index H^* ^a

	1	2	3	4	5	6	row sum
1	x	1	1/2	1/3	1/4	1/2	$2.5833 + x$
2	1	x	1	1/2	1/3	1	$3.8333 + x$
3	1/2	1	x	1	1/2	1/2	$3.5000 + x$
4	1/3	1/2	1	x	1	1/3	$3.1667 + x$
5	1/4	1/3	1/2	1	x	1/4	$2.3333 + x$
6	1/2	1	1/2	1/3	1/4	y	$2.5833 + y$

^a The variable Harary-Connectivity index H^* : $H^* = 1/\sqrt{\{(2.5833 + x)(3.8333 + x)\} + 1/\sqrt{\{(3.8333 + x)(3.5000 + x)\} + 1/\sqrt{\{(3.5000 + x)(3.1667 + x)\} + 1/\sqrt{\{(3.1667 + x)(2.3333 + x)\} + 1/\sqrt{\{(2.3333 + x)(2.5833 + y)\} = f(x, y)}$.

Table 6. Augmented $1/[\text{Distance}]^2$ Matrix and the Corresponding Row Sums for 2-Pentanol and the Variable Harary-Connectivity Index H^{2*} ^a

	1	2	3	4	5	6	row sum
1	x	1	1/4	1/9	1/16	1/4	$1.6736 + x$
2	1	x	1	1/4	1/9	1	$3.3611 + x$
3	1/4	1	x	1	1/4	1/4	$2.7500 + x$
4	1/9	1/4	1	x	1	1/9	$2.4722 + x$
5	1/16	1/9	1/4	1	x	1/16	$1.4961 + x$
6	1/4	1	1/4	1/9	1/16	y	$1.6736 + y$

^a The variable Harary-Connectivity Index H^{2*} : $H^{2*} = 1/\sqrt{\{(1.6736 + x)(3.3611 + x)\} + 1/\sqrt{\{(3.3611 + x)(2.7500 + x)\} + 1/\sqrt{\{(2.7500 + x)(2.4722 + x)\} + 1/\sqrt{\{(2.4722 + x)(1.4961 + x)\} + 1/\sqrt{\{(1.4961 + x)(1.6736 + y)\} = f(x, y)}$.

and the corresponding entries in the distance matrix add to a constant, the longest length for compounds of formula $C_5H_{11}OH$, which is six, the graph theoretical separation of terminal non-hydrogen atoms in 1-hexanol. A similar modification of the distance matrix was recently considered by Balaban and co-workers.²⁸

The reversed distance matrix leads to a topological index for which we use symbol $1/J$. Hence, we will designate the variable analogue of $1/J$ as $1/J^*$. At the bottom of Table 4 we gave the expression for $1/J^*$ illustrated on 2-pentanol. Additional variable indices can be similarly constructed from the reciprocal distance matrix,²⁹ the reciprocal square distance matrix, and the hyper-Wiener matrix.^{30,31} In Tables 5–7 we show the corresponding augmented matrices for 2-pentanol and the corresponding expressions for the variable indices derived from their respective row sums for the variable Harary indices and the variable hyper-Wiener index.

REGRESSIONS OF THE MOTOR OCTANE NUMBERS OF OCTANE ISOMERS WITH THE VARIABLE INDICES J^* AND $1/J^*$

Balaban's J index was tested on motor octane numbers (MON) and found as a suitable descriptor.³² The variable J^* index when $x = 0$ corresponds to the Balaban J index, which when correlated for MON (Motor Octane Numbers) gives the standard error $s = 9.54$, the correlation coefficient $r = 0.9277$, and the Fisher ratio $F = 86.4$. This result may be considered fair in view that we have limited regression only to isomers of octane, i.e., molecules having the same size. The number of isomers used was $n = 16$. The missing data occur for n -octane and 2,2,3,3-tetramethylbutane. In Table 8 we show for a selection of physicochemical properties of octanes some of the better regressions based

Table 7. Augmented Hyper-Wiener Matrix and the Corresponding Row Sums for 2-Pentanol and the Variable Hyper-Wiener Index WW^* ^a

	1	2	3	4	5	6	row sum
1	x	5	3	2	1	1	$12 + x$
2	5	x	9	6	3	5	$28 + x$
3	3	9	x	8	4	3	$27 + x$
4	2	6	8	x	5	2	$23 + x$
5	2	3	4	5	x	1	$14 + x$
6	2	5	3	2	1	y	$12 + y$

^a The variable Hyper-Wiener index WW^* : $WW^* = 1/\sqrt{\{(12 + x)(28 + x)\} + 1/\sqrt{\{(28 + x)(27 + x)\} + 1/\sqrt{\{(27 + x)(23 + x)\} + 1/\sqrt{\{(23 + x)(14 + x)\} + 1/\sqrt{\{(14 + x)(12 + y)\} = f(x, y)}$.

Table 8. Change of the Standard Error s for Motor Octane Numbers as the Variable Changes from $x = -5$ to $x = 100$ for Balaban Index $1/J^*$

property	r	descriptor ref
acentric factor	0.992	$^2\chi$
motor octane numbers	0.989	$\lambda\lambda_1$
density	0.979	$^3\chi$
molecular volume	0.978	$^3\chi$
molar refraction	0.970	$^3\chi$
surface tension	0.964	$^2\chi - ^3\chi$
entropy	0.964	R_1
heat of vaporization	0.958	Z
heat of atomization	0.931	$1/2\chi$
heat of formation	0.931	$1/2\chi$
^{13}C chemical shift sum	0.929	W/Z
critical temperature	0.889	$^1\chi - ^2\chi$
boiling points	0.888	Z
critical volume	0.849	$\chi(V)$
critical pressure	0.668	$1/2\chi$

Table 9. Selection of Better Regressions of Physicochemical Properties of Octanes Using a Single Molecular Descriptor

x	r	s	F
-5	0.9056	10.84	63.9
-2	0.9218	9.91	79.2
-1	0.9250	9.71	90.0
0	0.9277	9.54	86.4
+5	0.9361	8.99	99.2
+10	0.9408	8.66	107.7
+20	0.9457	8.31	118.5
+50	0.9511	7.90	132.6
+100	0.9537	7.69	140.7

on a single molecular descriptor. As we see for most properties the regression coefficients are above 0.900, and one of the largest values ($r = 0.9891$) obtained belongs to a quadratic regression for the motor octane numbers using the leading eigenvalue of the path matrix as the descriptor. The path matrix is constructed from path subgraphs so that an element i, j depicts all the shortest paths between vertices i and j .^{33,34}

In Table 9 we show how the standard error s decreases as we alter the values of the variable J^* index from $x = -5$ to $x = 100$. The best value we show is when $x = +100$ ($r = 0.9537$), but it is clear that the standard error would further decrease if we continued to increase the value of the variable x . The reason we did not wish to continue to increase x further and find the optimal value of x is that the "reversed" J index, $1/J$ when $x = 0$, already gives a better regression ($r = 0.9657$) than J^* for $x = +100$. In Table 10 we show how the regression of motor octane numbers improves as x in $1/J^*$ varies. This time, as we see x takes negative values

Table 10. Change of the Standard Error s for Motor Octane Numbers as the Variable Changes from $x = 0$ to $x = 25$ for Reverse Balaban Index $1/J^*$

x	r	s	F
0	0.9657	6.64	193.4
-5	0.9668	6.53	200.3
-10	0.9682	6.39	209.9
-20	0.9726	5.94	244.8
-25	0.9739	5.80	258.0

Table 11. Augmented $1/JJ$ Matrix and the Corresponding Row Sums for 2-Pentanol and the Variable $1/JJ$ -Connectivity Index^a

	1	2	3	4	5	6	row sum
1	x	5/1	4/2	3/3	2/4	4/2	$10.5 + x$
2	5/1	x	5/1	4/2	3/3	5/1	$18 + x$
3	4/2	5/1	x	5/1	4/2	4/2	$16 + x$
4	3/3	4/2	5/1	x	5/1	3/3	$14 + x$
5	2/4	3/3	4/2	5/1	x	2/4	$9 + x$
6	4/2	5/1	4/2	3/3	2/4	y	$10.5 + y$

^a The variable $1/JJ^*$ index: $1/JJ^* = 1/\sqrt{(10.5 + x)(18 + x)} + 1/\sqrt{(18 + x)(16 + x)} + 1/\sqrt{(16 + x)(14 + x)} + 1/\sqrt{(14 + x)(9 + x)} + 1/\sqrt{(18 + x)(10.5 + y)} = f(x, y)$.

Table 12. List of J , $1/J$, $1/JJ$, and the Optimal $1/JJ^*$ (for $x = +20$) for the Octane Isomers

octane isomers	$J(x=0)$	$1/J(x=0)$	$1/JJ(x=0)$	$1/JJ^*(x=+20)$
<i>n</i> -octane	.361437	.196761	.335371	.169976
2M	.387978	.188610	.316042	.164753
3M	.408867	.184414	.311263	.163119
4M	.417088	.183107	.310140	.162692
3E	.439196	.178876	.304367	.160944
22MM	.444538	.177486	.288032	.156197
23MM	.452974	.176123	.290849	.157142
24MM	.442690	.177233	.292542	.157857
25MM	.418260	.181070	.297240	.159559
33MM	.481912	.172447	.282242	.154176
34MM	.470354	.173545	.284181	.155236
2M3E	.479268	.172233	.285023	.155308
3M3E	.511888	.168717	.273774	.151518
223MMM	.517612	.167383	.260103	.147779
224MMM	.484132	.170791	.269985	.150964
233MMM	.529761	.166233	.261800	.147964
234MMM	.494890	.169619	.275475	.152470
2233MMMM	.574342	.161165	.242276	.141658

and when $x = -25$, we have already reduced the standard error to 5.80. Again, rather than pursuing the optimal value of x we will examine a novel index obtained by combining the schemes for construction of the indices J and $1/J$.

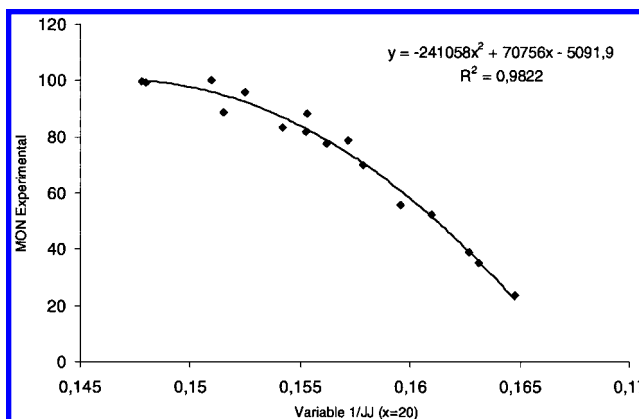
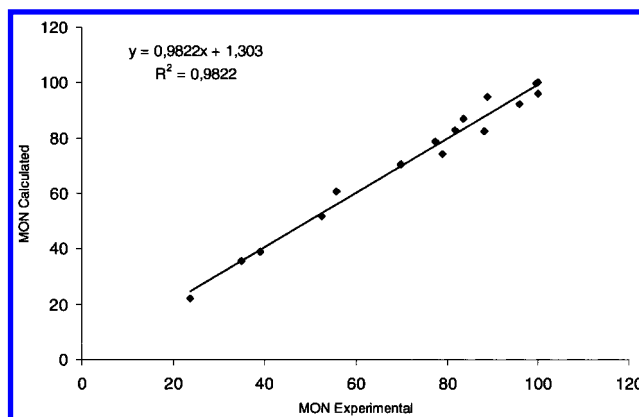
REGRESSIONS OF MOTOR OCTANE NUMBERS WITH NOVEL VARIABLE INDEX $1/JJ^*$

The novel index $1/JJ$ is obtained by combining the elements of the distance matrix with the corresponding elements of the reverse distance matrix and obtaining the quotient of the corresponding matrix elements as is illustrated in Table 11 again on a graph of 2-pentanol. In Table 12 we have listed J , $1/J$, and $1/JJ$ for octane isomers. In the last column we give the value of $1/JJ^*$ when $x = +20$, which is the optimal value for the variable x . In Table 13 we show how statistical values vary when x varies. When $x = +20$ we have obtained the minimum value for the standard error associated with $r = 0.991$ and F over 355. In Figure 1 we plotted motor octane numbers against $1/JJ^*$ ($x = +20$), and in Figure 2 we show a plot of calculated MON against the

Table 13. Change of the Standard Error s for Motor Octane Numbers as the Variable Changes from $x = -5$ to $x = +20$ the Index $1/JJ^*$ ^a

	r	s	F
$x = -5$	0.9838	4.76	195.3
$x = 0$	0.9895	3.84	303.3
$x = +5$	0.9904	3.67	332.0
$x = +10$	0.9908	3.60	347.0
$x = +20$	0.9910	3.54	358.0

^a $\text{MON} = -241,057.966 (1/JJ^*)^2 + 70,76.463 (1/JJ^*) - 5,091.925$.

**Figure 1.** The plot of the motor octane numbers against $1/JJ^*$ ($x = +20$).**Figure 2.** The plot of the calculated MON against the experimental MON values.

experimental MON values. At the bottom of Table 13 we show the regression equation for the optimal choice of the variable $1/JJ^*$.

If we compare the correlation coefficient $r = 0.991$ with those of Table 8, we see that this is one of the best simple regressions found for octanes. At the same time this result improves, though only slightly, the best previously obtained regression for motor octane numbers having $r = 0.989$ and based on the leading eigenvalue of the path matrices is $\lambda\lambda_1$.

In Table 14 we summarized the regression statistic for J , reverse J , and the combination of J and $1/J$. In the top part we show the results obtained assuming linear regressions, and in the bottom part we show the corresponding statistics based on quadratic regressions. In the case of J and $1/J$ ($x = 0$) there is no dramatic improvement between the linear and the corresponding quadratic regressions. However for $1/JJ$ the standard error has been reduced dramatically by inclusion of the quadratic term. It is well-known that nonlinear regression may drastically improve a correlation

Table 14. Summary of the Regression Statistic for J , the Reverse J , and the Combination $1/JJ^a$

Linear Regression				
J	$x = 0$	$r = 0.9277$	$s = 9.54$	$F = 86.4$
$1/J$	$x = 0$	$r = 0.9657$	$s = 6.64$	$F = 193.4$
$1/JJ$	$x = 0$	$r = 0.9595$	$s = 7.20$	$F = 162.3$
$1/JJ^*$	$x = +20$	$r = 0.9561$	$s = 7.49$	$F = 149.0$
Quadratic Regression				
J	$x = 0$	$r = 0.9632$	$s = 7.13$	$F = 83.5$
$1/J$	$x = 0$	$r = 0.9735$	$s = 6.07$	$F = 117.7$
$1/JJ$	$x = 0$	$r = 0.9895$	$s = 3.84$	$F = 303.3$
$1/JJ^*$	$x = +20$	$r = 0.9910$	$s = 3.54$	$F = 358.$

^a The results based on the optimal variable $1/JJ^*$ are shown in the last row of the table.

for a size-dependent property. Here we see that even when considering molecules of the same size various topological indices may display a different functional dependence. Thus J and $1/J$ offer fair linear regressions, but in the case of $1/JJ$ only the quadratic regression offer superior result. This should be borne in mind when comparing results based on different descriptors. Otherwise one may incorrectly assume that an inferior descriptor is better than descriptors using nonlinear regression offering the smallest standard error. For example, from the upper part of Table 14 one would say that the reversed Balaban's index $1/J$ is better than J or $1/JJ$, being associated with best statistical parameters. However, as we can see from the lower part of the same table it is $1/JJ$ that is the best among the three indices considered.

Illustrations of such erroneous claims can be found in recent papers of Ren³⁵ and Bonchev,³⁶ who based their conclusion on linear regressions between their corresponding topological indices against linear regression for the same compounds using the connectivity indices. Indeed their indices are better when one restricts simple regression (using a single variable) to linear dependence between the property and structure. However, when the same data are analyzed using quadratic expressions, the regressions based on the connectivity indices showed visibly smaller standard errors than the corresponding quadratic regressions based on the Xu index of Ren and the "overall connectivities" of Bonchev. That one can get a misleading result when confining an analysis to a linear model is also illustrated when one compares the linear and the quadratic models using $1/JJ$ ($x = 0$) and $1/JJ^*$ ($x = +20$). In linear model $1/JJ$ for $x = 0$ gives slightly better statistics than the $1/JJ^*$ for $x = +20$, but as we have seen this is no longer true when the quadratic regressions are compared. Clearly linear regression is simpler than a quadratic, but there are not theoretical or conceptual arguments that demand that the structure-property relationship ought to be linear. In fact most structure-property correlations are nonlinear, particularly when the domain is extended to include diverse molecules and molecules of widely different sizes.

ON INTERPRETATION OF THE DISTANCE RELATED DESCRIPTORS

In recent years we have continued to witness construction of novel molecular descriptors. These include, for example, the variable connectivity indices,¹⁵ the path/walks shape descriptors,³⁷ the 2D planar chirality index,³⁸ etc. Also we see some diversification of use of topological indices. For

example, a very important new application was illustrated by Lahana and co-workers,³⁹ who employed two dozen topological indices in screening a virtual combinatorial library having about 280,000 dipeptides in search for novel immunosuppressive compound. By restricting the permitted intervals for topological indices based on the values of the lead compounds they succeeded in reducing the initial pool of compounds to about two dozen, which were further examined narrowing attention to four compounds of the initial 280,000. After synthesizing these compounds they found a compound having about 100 times stronger activity than the lead compound!

The efforts on construction of novel indices and their use continue, but no similar efforts seem to be directed toward interpretation of topological indices. Occasionally a paper appears in which attempts are made to explain why a particular index performs so well. Thus recently Kier and Hall⁴⁰ tried to relate the relative magnitudes of the connectivity index $^1\chi$ to intermolecular accessibility.

The Wiener index,⁴ the oldest nontrivial topological index, still remains mostly a mystery. However, Randić and Zupan⁴¹ have recently examined the partition of a selection of topological indices into bond contributions and have shown that topological indices attribute to different bonds in a molecule with different relative weights. Thus, for instance, the connectivity index $^1\chi$ and the Hosoya topological index⁵ Z in the case of hydrocarbons assign to terminal CC bonds greater weights than to the internal CC bonds. In contrast the Wiener number and here considered Balaban's index J do just the opposite: they assign a smaller role to terminal bonds and a dominant role to internal CC bonds. The following question can then be raised: which is better? The answer may depend on the property considered. As is known already from the early literature¹² most physicochemical properties of hydrocarbons apparently prefer descriptors that give a greater role to terminal CC bonds.

From the lower part of Table 14 we have seen that of the descriptors considered Balaban's J is not as good as the reversed J index, which in turn is not as good as combination $1/JJ$. This is a result of the search for optimal solution. Can this result be better understood? In Figure 3 we show a molecular diagram of n -octane, 2-methylheptane, and 3-methylheptane and the corresponding partitioning of the indices J , $1/J$, $1/JJ$, and $1/JJ^*$ for each CC bond in these compounds. As we see in the case of the Balaban's J index peripheral CC bonds have smaller weights. Because $1/J$, $1/JJ$, and $1/JJ^*$ gave a better regression than J it stands to reason that physicochemical properties such as motor octane numbers prefer descriptors that give greater weight to the peripheral CC bonds.

The difference between $1/J$, $1/JJ$, and $1/JJ^*$ is in the relative magnitudes of the weights (partitions) between different bonds. In all three cases the terminal CC bonds make a greater contribution to the molecular topological index. The relative weights for $1/J$, $1/JJ$, and $1/JJ^*$ of terminal bonds, C_1-C_2 and C_7-C_8 in n -octane, bonds C_1-C_2 , C_2-C_8 , and C_6-C_7 in 2-methylheptane, and bonds C_1-C_2 , C_3-C_8 , and C_6-C_7 in 3-methylheptane, are larger than the weights of all internal bonds, with a single exception in 3-methylheptane where the weights of C_3-C_8 and C_5-C_6 (which is next to the terminal bond) are of similar magnitude but C_5-C_6 has a slightly larger weight. Clearly the relative

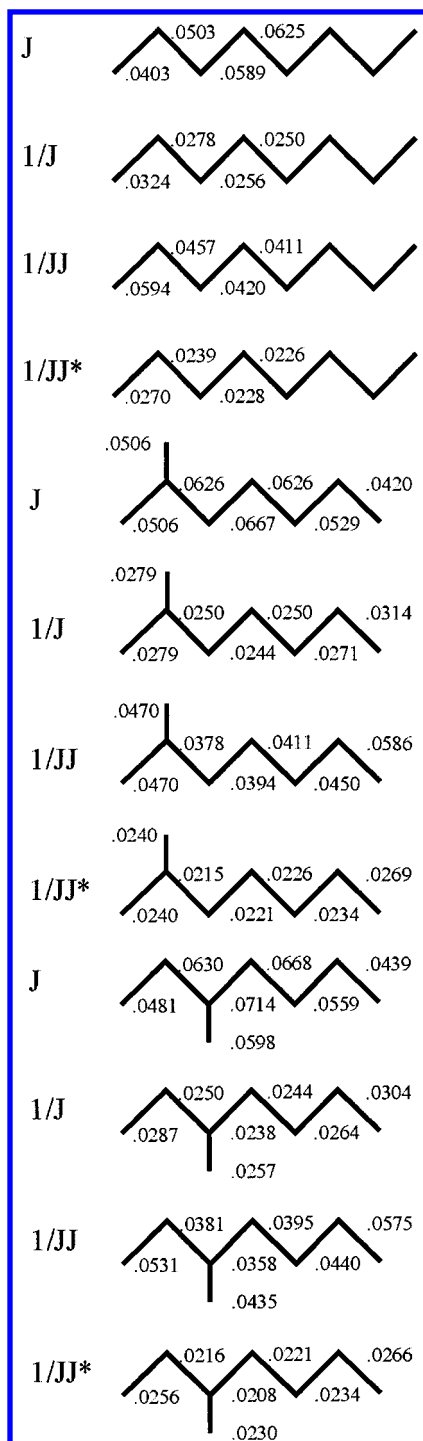


Figure 3. The partition of topological indices J , $1/J$, $1/JJ$, and $1/JJ^*$ for CC bonds in (a) n -octane, (b) 2-methylheptane, and (c) 3-methylheptane.

weights in $1/J$, $1/JJ$, and $1/JJ^*$ better describe variations in contributions of different bonds in a molecule.

CONCLUDING REMARKS

In the present paper we described construction of variable indices related to Balaban's J index using the distance matrix, the reversed distance matrix, and a matrix with elements obtained as quotients of the reverse distance and the distance matrix. When so constructed indices are augmented by inserting variable x along the main diagonal, they need to be optimized. Optimization will in general depend on the

property considered, and we found for the case of motor octane numbers (MON) optimal value for x for $1/JJ$ to be +20. This formal mathematical approach offers some flexibility for interpretation of the results suggesting that variation of x is associated with variations in the partitioning of the topological index. In the case of MON for octane isomers we found that optimal descriptors suggest a dominant role of terminal CC bonds, but the difference between the role of terminal and internal bonds is less than that suggested by the use of nonoptimal descriptors.

ACKNOWLEDGMENT

We thank the Ministry of Science and Technology of Republic of Slovenia for the financial support.

REFERENCES AND NOTES

- (1) Randić, M. Topological Indices. In *The Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer III, H. F., Schreiner, P. R., Eds.; John Wiley & Sons: Chichester, 1998; pp 3018–3032.
- (2) Balaban, A. T. *Historical developments of topological indices*, in: *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds.; Gordon and Breach Sci. Publ: Amsterdam, 1999; pp 21–57.
- (3) Todeschini, R.; Consonni, V. *The Handbook of Molecular Descriptors*. In *The Series of Methods and Principles in Medicinal Chemistry*; Mannhold, R., Kubinyi, H., Timmerman, H., Eds.; Wiley-VCH: New York, 2000; Vol. 11, p 680.
- (4) Wiener, H. Structural determination of paraffin boiling points. *J. Am. Chem. Soc.* **1947**, *69*, 17–20.
- (5) Hosoya, H. 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) Randić, M. On the characterization of molecular branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609–6615.
- (7) Balaban, A. T. Highly discriminating distance-based topological index. *Chem. Phys. Lett.* **1982**, *89*, 399–404.
- (8) Gantmacher, F. *Theory of Matrices*; Chelsea Publishers: New York, 1959; Vol. II, Chapter 13.
- (9) Lovasz, L.; Pelikan, J. On the eigenvalues of trees. *Period. Math. Hung.* **1973**, *3*, 175–182.
- (10) Randić, M.; Kleiner, A. F.; DeAlba, L. M. Distance/distance matrices. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 277–286.
- (11) Kier, L. B.; Hall, L. H. Molecular connectivity VII: Specific treatment of heteroatoms. *J. Pharm. Sci.* **1976**, *65*, 1806–1809.
- (12) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: New York, 1976.
- (13) Ivanciuc, O.; Balaban, A. T. Design of Topological indices. Part 5. Precision and error in computing graph theoretical invariants for molecules containing heteroatoms and multiple bonds. *MATCH* **1994**, *30*, 117–139.
- (14) Randić, M. Variable connectivity index after 25 years. ACS Meeting, Washington, DC, August 20–24, 2000.
- (15) Randić, M.; Basak, S. C. Variable molecular descriptors. In *Some Aspects of Mathematical Chemistry*; Sinha, D. K., Basak, S. C., Mohanty, R. K., Busamallick, I. N., Eds.; Visva-Bharati University Press: Santiniketan, India, in press.
- (16) Randić, M. Novel graph theoretical approach to heteroatom in quantitative structure–activity relationship. *Chemometrics & Intel. Lab. Syst.* **1991**, *12*, 970–980.
- (17) Randić, M. On computation of optimal parameters for multivariate analysis of structure–property relationship. *J. Comput. Chem.* **1991**, *12*, 70–980.
- (18) Randić, M.; Dobrowolski, J. Cz. Optimal molecular connectivity descriptors for nitrogen containing molecules. *Int. J. Quantum Chem: Quantum Biol. Symp.* **1998**, *70*, 1209–1215.
- (19) Randić, M.; Mills, S. C. On use of variable connectivity index for characterization of amino acids. *Int. J. Quantum Chem.* In press.
- (20) Randić, M.; Basak, S. C. On construction of high quality structure–property–activity regressions: The boiling points of Sulfides. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 899–905.
- (21) Randić, M. High quality structure – property regressions. Boiling points of smaller alkanes. *New J. Chem.* **2000**, *24*, 165–171.
- (22) Randić, M.; Basak, S. C. Optimal molecular descriptors based on weighted path numbers. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 261–266.

- (23) Randić, M.; Pompe, M. On characterization of CC double bond in alkenes. *SAR QSAR Environ. Res.* **1999**, *10*, 451–471.
- (24) Randić, M.; Basak, S. C. Multiple regression analysis with optimal molecular descriptors. *SAR QSAR Environ. Res.* **2000**, *11*, 1–23.
- (25) Plavšić, D.; Nikolić, S.; Trinajstić, N.; Mihalić, Z. On the Harary index for characterization of chemical graphs. *J. Math. Chem.* **1993**, *12*, 235–250.
- (26) Ivanciuc, O.; Balaban, T. S.; Balaban, A. T. Design of topological indices. Part 4. Reciprocal distance matrix, related local vertex invariants and topological indices. *J. Math. Chem.* **1993**, *12*, 309–318.
- (27) Randić, M. Linear Combinations of Path Numbers as Molecular Descriptors. *New J. Chem.* **1997**, *21*, 945–951.
- (28) Balaban, A. T.; Mills, D.; Ivanciuc, O.; Basak, S. C. Reverse Wiener indices. *Croat. Chem. Acta* **2000**, *73*, 923–941.
- (29) Diudea, M. Indices of reciprocal properties or Harary indices. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 292–299.
- (30) Randić, M.; Guo, X.; Oxley, T.; Krishnapriyan, H. Wiener matrix: Source of novel graph invariants. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 709–716.
- (31) Randić, M.; Guo, X.; Oxley, T.; Krishnapriyan, H.; Naylor, L. Wiener matrix invariants. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 361–367.
- (32) Balaban, A. T.; Kier, L. B.; Joshi, N. Structure–property analysis of octane numbers for hydrocarbon (alkanes, cycloalkanes, alkenes). *MATCH* **1992**, *28*, 13–27.
- (33) Randić, M. On molecular branching. *Acta Chim. Slov.* **1997**, *44*, 57–77.
- (34) Randić, M.; Plavšić, D.; Razinger, M. Double invariants. *MATCH* **1997**, *35*, 243–259.
- (35) Ren, B. A new topological index for QSPR of alkanes. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 139–143.
- (36) Bonchev, D. Overall connectivity and topological complexity for QSAR/QSPR in: *Topological Indices and Related Descriptors QSAR and QSPR*; Delivers, J., Balaban, A. T., Eds.; Gordon and Breach Sci. Publ: Amsterdam, 1999; pp 361–401.
- (37) Randić, M. Novel shape descriptors for molecular graphs. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 607–613.
- (38) Randić, M. The first graph theoretical descriptor of 2-D chirality, *J. Chem. Inf. Comput. Sci.* **1999**, accepted for publication.
- (39) Grassy, G.; Calas, B.; Yasri, A.; Lahana, R.; Woo, J.; Iyer, S.; Kaczorek, M.; Floc'h, R.; Buelow, R. Computer-assisted rational design of immuno-suppressive compounds. *Nature Biotechnol.* **1998**, *16*, 748–752.
- (40) Kier, L. B.; Hall, L. H. Intermolecular accessibility: the meaning of molecular connectivity. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 792–795.
- (41) Randić, M.; Zupan, J. On the interpretation of well-known topological indices. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 550–560.

CI0001029