Ad hoc optimal topological indices for QSPR

E. A. Smolenskii, G. V. Vlasova, D. Yu. Platunov, and A. N. Ryzhov*

N. D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences,
47 Leninsky prosp., 119991 Moscow, Russian Federation.
E-mail: smolensk@ioc.ac.ru

Optimal topological indices for particular physicochemical properties can be designed using the chemical structure matrix. Taking the alkane boiling points as examples, the extrapolation properties of some well-known topological indices were compared with those of the optimal topological indices designed *ad hoc*. The advantages of the new indices are shown.

Key words: alkanes, descriptors, topological indices, chemical structure matrix, quantitative structure—property relationships, physicochemical properties.

One of the key problems in theoretical chemistry is establishment of quantitative relationships between the physicochemical properties and molecular structures of organic compounds. The properties are assumed to be functions of some arguments that depend on the molecular structure. Usually, these are various topological indices,[1] which can be treated as invariants of molecular graphs obtained ignoring hydrogen atoms.[2] The structures of alkanes are represented by trees, *i.e.*, graphs containing no rings and multiple edges. Consider a set of N alkanes from methane to C_nH_{2n+2} including all possible isomers and enumerate them from 1 to N. We will respectively denote the molecular graphs of methane, ethane, propane, butane, isobutane, *etc.* as g_1, g_2, g_3, g_4, g_5, *etc*. Let us define, for the ith alkane, a function $X_{j,i} = [g_j]_i$, where the quantity in square brackets denotes the number of different subgraphs g_j of the graph g_i of the ith alkane. Formally, the parameters $[g_j]_i$ can be treated as vector-type topological indices. Using them, for any class of organic compounds one can construct[3] a non-degenerate matrix called the chemical structure matrix (CSM). The columns of the CSM are linearly independent vectors that form a basis suitable for representation of any topological index. Let us denote the array of the values of a property P for the set of N alkanes as vector $\bar{P}$ and the optimal topological index as vector $\bar{A}$.

$$\bar{P} = \begin{pmatrix} P_1 \\ P_2 \\ \cdot \\ \cdot \\ \cdot \\ P_N \end{pmatrix}, \qquad \bar{A} = \begin{pmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_N \end{pmatrix}.$$

Then one has

$$S\bar{A} = \bar{P}, \qquad (1)$$

where S is the CSM. Having solved Eqn. (1), it is possible to obtain[3] quite unexpected expressions for the Randič index χ [4] and Hosoya index Z [5] as linear combinations of the vectors $[g_j]_i$. In this work we report on the results of computer experiments including the construction of linear models and prediction of (i) boiling points (T_b) for nine isomers of heptane and eighteen isomers of octane using the T_b values for thirteen alkanes C_1—C_6 and (ii) the Wiener index W,[6,7] the indices Z and χ, and the complexity index K.[8] The boiling points T_b for eighteen isomers of octane were predicted using analogous relationships and the T_b values for twenty-two alkanes C_1—C_7.

For the sake of comparison, similar computer experiments were carried out for the optimal topological index obtained by solving Eqn. (1) and some simplified modifications of this index. A comparative analysis of the results obtained was performed. Treatment of the initial data for the light alkanes as the training set and the data for the heavier alkanes as the test set allows one to disclose a fundamental difference between our approach and commonly accepted procedures (see, *e.g.*, Ref. 9), namely, we predict the properties of more complex compounds based on the properties of the compounds with simpler structures.

Topological index—property models: construction and investigation of extrapolation properties

In this work we will restrict ourselves to consideration of the boiling points of alkanes because this parameter is

Published in Russian in *Izvestiya Akademii Nauk. Seriya Khimicheskaya,* No. 9, pp. 1454—1461, September, 2006.
1066-5285/06/5509-1508

simple, very important, and used in all classical studies concerning topological indices.[4–6]

In order to compare different topological indices and to analyze their ability to predict the boiling points for hypothetical compounds, we will represent T_b as a linear function of the topological index:

$$(T_b)_i = A + B(\mathrm{TI})_i, \qquad (2)$$

where i is the number of an alkane included in the training set ($i = 1, ..., N$). By solving the system of N equations (2) for each topological index one can determine the coefficients A and B using the least-squares procedure. The results obtained are listed in Table 1. Parameters of other indices (corresponding design procedures are described below) are also presented in Table 1.

Solution of Eqn. (1) for T_b gives the following coefficients a_i:

$a_1 = -161.58$, $a_2 = 234.53$, $a_3 = -26.380$, $a_4 = -5.010$,

$a_5 = 10.150$, $a_6 = -4.987$, $a_7 = 3.022$, $a_8 = -3.040$,

$a_9 = -3.906$, $a_{10} = 0.833$, $a_{11} = 0.845$, $a_{12} = -0.365$,

$a_{13} = 0.750$.

From here for the T_b values of thirteen alkanes C_1—C_6 one gets

$$(T_b)_i = \sum_{j=1}^{13} a_j [g_j]_i. \qquad (3)$$

Considering $[g_j]_i$ as a topological index x_i (subscript j corresponds to the subgraph g_j, where i is the number of compound; $i = 1, ..., N$), the linear equation (3) derived using the coefficients a_j can formally be associated with a new topological index for T_b:

$$X_{13} = \sum_{j=1}^{13} a_j x_j. \qquad (4)$$

For each of the thirteen alkanes C_1—C_6 this topological index coincides with the corresponding boiling point. Formally, we obtained a trivial result. Indeed, we have solved the system of thirteen linear equations with thirteen unknowns and a nonzero determinant. As should be expected, the index X_{13} provides an exact description of the data ($s = 0$, $R^2 = 1$). However, the conclusions following from this "trivial" result are quite non-trivial. We can say that relationship (3) has the form (2) at $A = 0$ and $B = 1$. Now we will use expression (2) (remind that the coefficients A and B were determined assuming that the T_b values are only known for thirteen isomers of alkanes C_1—C_6) for prediction of T_b values for nine isomers of heptane and eighteen isomers of octane. Having denoted the index (4) as X_{13} now we will design an index X_9' by excluding the x_i values for which $|a_i| < 1$ (here, x_{10}—x_{13}). Of course, the coefficients A and B for the index X_9' will change (they are also listed in Table 1). The predictions obtained using expression (2) and different topological indices are listed in Table 2.

We carried out yet another computer experiment. Using the T_b values for twenty-two isomers of alkanes C_1—C_7 as the training set, we recalculated the parameters of relationship (2) (see Table 3). The index X_{13}' was designed as mentioned above by excluding nine smallest (in absolute value) coefficients a_i from the array of twenty-two values.

Thus, we introduced two new indices. The X_{22} index was designed using all twenty-two parameters and the X_{13}' index was derived by excluding nine small coefficients. Using the recalculated parameters of relationship (2), we predicted the boiling points for eighteen isomers of octane (the test set). The results of calculations and the predictions are listed in Table 4.

Data analysis shows that the indices W, Z, χ, and K are characterized by lower predictive power (extrapolation properties) compared to the new *ad hoc* indices. Even the Randič index, which behaves much better than the other well-known indices, can not compete with the new indices.

Table 1. Parameters of expression (2) for different indices[a] ($N = 13$)

Index	A	B	R^2	s	$\lvert\Delta\rvert_{max}$
W[b]	−83.90	4.98	0.814	29.40	77.682
Z	−102.62	15.59	0.804	30.18	74.549
χ	−163.73	44.15	0.963	13.17	25.438
K	−97.57	6.34	0.778	32.14	70.353
X_{13}	0.00	1.00	1.000	0.00	0.000
X_9'	0.15	1.00[c]	1.000[c]	0.36	0.618
$\ln W$[d]	−100.22	45.21	0.972	8.14	15.633
$\ln Z$	−146.73	88.35	0.984	8.71	14.854
$\ln K$	−160.91	66.89	0.956	14.28	29.979

[a] R^2 is the squared correlation coefficient, s is the root-mean-square deviation, and $|\Delta|_{max}$ is the maximum absolute deviation of the results of calculations from experimental data.

[b] Here and in Tables 2—6 and 8 the parameter W for methane was set to zero.

[c] The B and R^2 values for the X_9' index differ from unity by less than 0.001.

[d] Here and in Table 5 methane was not included in the set of descriptors $\ln W$ ($N = 12$), because formally for methane $W = 0$ and the logarithm of 0 does not exist.

Results and Discussion

Analysis of the results of the boiling point calculations and the values predicted for more complex compounds (see Tables 2 and 4) can lead to conclusion that most topological indices considered here, except the Randič index, are of no interest at all. As will be shown below, this is not the case.

Table 2. Boiling points calculated (C_1—C_6) and predicted (C_7, C_8) using expression (2) ($N = 13$)

Compound	T_b/°C	W	Z	χ	K	X_{13}	X_9'
Methane	−161.58	−77.68	−74.55	2.15	−70.35	0.00	0.19
Ethane	−88.63	−9.71	−17.19	−13.20	−10.08	0.00	0.04
Propane	−42.06	21.93	13.79	2.16	17.48	0.00	−0.06
n-Butane	−0.5	33.63	24.17	12.51	33.69	0.00	−0.15
2-Methylpropane	−11.72	27.39	28.54	−5.91	16.13	0.00	−0.12
n-Pentane	36.073	20.43	13.98	17.87	38.57	0.00	−0.22
2-Methylbutane	27.852	22.16	21.35	2.45	17.67	0.00	−0.21
2,2-Dimethylpropane	9.5	13.77	34.17	−25.44	−19.69	0.00	−0.17
n-Hexane	68.74	−21.56	−31.30	19.32	33,21	0.00	−0.29
2-Methylpentane	60.271	−15.09	−8.59	3.66	5.73	0.00	0.56
3-Methylpentane	63.282	−7.11	−21.17	6.67	2.40	0.00	0.57
2,2-Dimethylbutane	49.741	−5.72	12.06	−16.41	−42,83	0.00	−0.62
2,3-Dimethylbutane	57.988	−2.45	4.72	−5.82	−21.91	0.00	0.48
	s	29.4	30.2	13.2	32.1	0.0	0.4
	$\|\Delta\|_{max}$	77.68	74.55	25.44	70.35	0.00	0.57
n-Heptane	98.428	−96.38	−126.32	17.80	18.53	−2.98	−3.34
2-Methylhexane	90.052	−84.85	−87.93	2.22	−15.20	1.02	1.52
3-Methylhexane	91.851	−73,10	−101.72	4.02	−19.74	0.06	1.40
3-Ethylpentane	93.475	−61.52	−115.69	5.65	−37.13	−1.10	1.10
2,2-Dimethylpentane	79.198	−65.84	−36.43	−18.17	−83.09	0.36	2.18
2,3-Dimethylpentane	89.784	−55.26	−72.61	−5.24	−59.83	−0.32	2.62
2,4-Dimethylpentane	80.5	−74.50	−50.72	−14.53	−56.44	−0.65	2.37
3,3-Dimethylpentane	86.064	−49.02	−60.74	−11.30	−95.24	−0.62	0.01
2,2,3-Trimethylbutane	80.883	−44.25	−19.16	−23.68	−106.76	−4.18	−2.99
	s	73.2	87.1	14.3	67.5	1.9	2,3
	$\|\Delta\|_{max}$	96.38	126.32	23.68	106.76	4.18	3.34
n-Octane	125.665	−208.50	−301.74	13.79	−4.94	−8.41	−8.84
2-Methylheptane	117.646	−191.63	−231.82	−1.43	−44.64	−4.05	−3,62
3-Methylheptane	118.925	−175.42	−261.72	−0.15	−62.38	−1.63	−0.35
4-Methylheptane	117.709	−171.66	−247.34	−1.37	−69.93	−2.60	−0.48
3-Ethylhexane	118.534	−155.91	−277.70	−0.54	−81.78	−0.64	2,33
2,2-Dimethylhexane	106.840	−162.62	−149.09	−21.75	−112.49	3.15	4.92
2,3-Dimethylhexane	115.607	−148.88	−202.68	−10.67	−91.05	0.89	4.62
2,4-Dimethylhexane	109.429	−160.04	−193,27	−16.85	−116.24	0.66	4.47
2,5-Dimethylhexane	109.103	−175.29	−178.00	−17.18	−91.21	3.68	4.99
3,3-Dimethylhexane	111.969	−137.59	−175.14	−16.62	−158.07	0.09	3.18
3,4-Dimethylhexane	117.725	−136.81	−231.74	−8.55	−95.27	−0.60	4.81
2-Methyl-3-ethylpentane	115.650	−133,91	−218.22	−10.63	−103.68	−2.44	3.82
3-Methyl-3-ethylpentane	118.259	−116.37	−215.62	−10.33	−132.76	−0.06	1.51
2,2,3-Trimethylpentane	109.841	−119.81	−130.50	−25.95	−191.88	−4.87	1.31
2,2,4-Trimethylpentane	99.238	−145.34	−94.33	−36.56	−158.12	2.01	8.81
2,3,3-Trimethylpentane	114.760	−109.91	−141.17	−21.03	−199.64	−4.79	0.19
2,3,4-Trimethylpentane	113.467	−126.14	−158.05	−20.01	−150.23	−0.99	6.85
2,2,3,3-Tetramethylbutane	106.300	−98.46	−56.09	−39.01	−170.07	−15.20	−11.03
	s	155.7	208.1	19.3	127.4	4.9	5.3
	$\|\Delta\|_{max}$	208.50	301.74	39.01	199.64	15.20	11.03

Note. Here and in Tables 4—6 for different indices listed are the deviations of the results of calculations from experimental boiling point values.

Prediction of the boiling points for the nine isomers of heptane using expression (2) and the Randič index (the best index with respect to the parameters *s* and $|\Delta|_{max}$) gives the following results. Four calculated T_b values differ from the corresponding experimental values by less than 10 °C, other four T_b values differ by 10—20 °C, and one T_b value differs by more than 20 °C (namely, by 23.68 °C). For comparison, five theoretical values calcu-

Table 3. Parameters of expression (2) for different indices ($N = 22$)

Index	A	B	R^2	s	$\|\Delta\|_{max}$
W	−61.09	3.31	0.807	29.16	100.488
Z	−65.69	9.48	0.827	32.24	105.373
χ	−159.28	42.76	0.972	12.96	23.664
K	−67.75	4.30	0.808	33.97	98.130
X_{22}	0.00	1.00	1.000	0.00	0.000
X_{13}'	0.47	1.01	1.000[a]	0.91	1.775
$\ln W$[b]	−107.80	49.39	0.971	8.49	19.624
$\ln Z$	−137.39	81.27	0.979	9.61	24.191
$\ln K$	−161.80	68.23	0.948	15.08	33.110

[a] The R^2 values for the X_{13}' index differs from unity by less than 0.001.

[b] Here and in Table 6 methane was not included in the set of descriptors $\ln W$ ($N = 21$), because formally for methane $W = 0$ and the logarithm of 0 does not exist.

lated using the X_{13} index differ from the experimental data by at most 1 °C, three theoretical values differ by at most 2—3 °C, and only for the isomer with the highest degree of branching (2,2,3-trimethylbutane) we obtained a large deviation of 4 °C. This can be explained by the fact that the 2,2,3-trimethylbutane molecule contains a fragment that is absent in the training set.

It should be noted that the *ad hoc* index X_{13} is the optimal topological index for our training set (see Ref. 3). It is impossible to design an index that provides a higher accuracy. This is due to the fact that all basis vectors were involved in the index design procedure. Any index expressed using an incomplete basis can not provide a higher accuracy, which at best will be the same as that given by the optimal index. This reasoning is inapplicable to the extrapolation properties of the index, although one can hope that the accuracy of predictions will also be maximum for the new indices; however, this assumption should be tested in actual practice.

It is convenient to consider the advantages and drawbacks of other indices by analyzing the plots of T_b *vs.* particular index (Figs 1 and 2). Description using the Wiener and Hosoya indices and the complexity index K (see Fig. 1) is best with the logarithmic function, while the Randič index (see Fig. 2, *a*) requires the use of a combination of a linear function with certain functions for particular isomers. By considering the logarithms of the topological indices as some new indices we again carried out the computer experiments described above using expression (2). The parameters of this expression for the functions of logarithms are listed in Tables 1 and 3 and the results of calculations and predictions are given in Tables 5 and 6 for $N = 13$ and 22, respectively. From these data it follows that the use of logarithms of the indices instead of the indices themselves can significantly

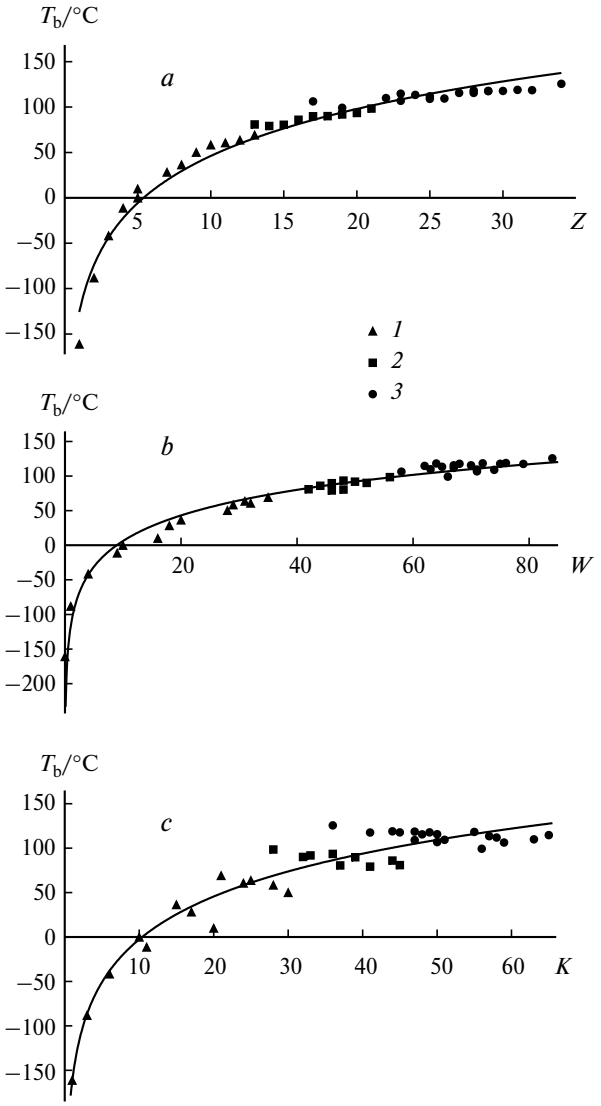

Fig. 1. Boiling points plotted *vs.* Hosoya index (*a*), Wiener index (*b*), and complexity index (*c*): isomers C_1—C_6 (*1*), C_7 (*2*), C_8 (*3*); solid lines denote corresponding logarithmic approximations.

improve the quality and predictive power of calculations. The accuracy of predictions becomes higher than in the case of calculations with the Randič index (this especially concerns the Wiener index). Figure 2, *a* shows that the boiling points of isomers are grouped near the straight lines that surprisingly pass in such a fashion that, at least in the first approximation, they intercept at the same point. Probably, with allowance for this fact in the case of the Randič index one will succeed in constructing such a model that its accuracy will be higher than that of our model with the *ad hoc* indices.

In Figs 2, *b*, *c* and in Fig. 3 the boiling points T_b are plotted *vs.* the optimal indices introduced in this work for

Table 4. Boiling points calculated (C_1—C_7) and predicted (C_8) using expression (2) ($N = 22$)

Compound	T_b/°C	W	Z	χ	K	X_{22}	X_{13}'
Methane	−161.58	−100.49	−105.37	−2.30	−98.13	0.00	0.84
Ethane	−88.63	−30.85	−41.91	−14.88	−33.79	0.00	0.25
Propane	−42.06	5.78	−4.82	1.45	−0.13	0.00	−0.13
n-Butane	−0.5	27.46	17.77	12.78	24.22	0.00	−0.47
2-Methylpropane	−11.72	19.55	16.04	−5.41	8.69	0.00	−0.37
n-Pentane	36.073	30.90	25.90	19.12	39.27	0.00	−0.76
2-Methylbutane	27.852	29.31	27.16	3.92	22.44	0.00	−0.69
2,2-Dimethylpropane	9.5	17.58	27.77	−23.66	−8.82	0.00	−0.55
n-Hexane	68.74	13.87	11.15	21.55	46.12	0.00	−1.03
2-Methylpentane	60.271	15.34	21.65	6.11	24.74	0.00	−0.12
3-Methylpentane	63.282	21.67	15.18	9.12	23.45	0.00	−0.13
2,2-Dimethylbutane	49.741	18.07	30.08	−13.66	−11.61	0.00	−1.24
2,3-Dimethylbutane	57.988	23.00	28.85	−3.14	5.24	0.00	−0.18
n-Heptane	98.428	−26.01	−35.02	21.00	45.68	0.00	−1.27
2-Methylhexane	90.052	−21.14	−14.95	5.66	20.09	0.00	−0.36
3-Methylhexane	91.851	−12.71	−22.64	7.45	17.59	0.00	0.53
3-Ethylpentane	93.475	−4.46	−30.49	9.08	6.30	0.00	1.33
2,2-Dimethylpentane	79.198	−12.11	12.13	−14.44	−29.50	0.00	1.40
2,3-Dimethylpentane	89.784	−1.53	−5.74	−1.58	−10.30	0.00	1.77
2,4-Dimethylpentane	80.5	−17.44	3.94	−10.87	−10.98	0.00	1.58
3,3-Dimethylpentane	86.064	1.38	0.03	−7.57	−35.54	0.00	−0.83
2,2,3-Trimethylbutane	80.883	2.82	23.29	−19.72	−45.03	0.00	0.41
	s	29.2	32.2	13.0	33.97	0.0	0.9
	$\|\Delta\|_{max}$	100.49	105.37	23.66	98.13	0.00	1.77
n-Octane	125.665	−91.54	−131.07	17.98	38.49	−2.45	−3.96
2-Methylheptane	117.646	−83.00	−91.67	2.98	8.95	0.89	0.31
3-Methylheptane	118.925	−71.78	−109.36	4.26	−2.68	0.27	0.59
4-Methylheptane	117.709	−69.68	−101.09	3.05	−8.20	0.27	1.49
3-Ethylhexane	118.534	−58.92	−119.23	3.87	−15.98	0.34	2.42
2,2-Dimethylhexane	106.840	−67.30	−45.58	−17.04	−40.59	−0.27	0.90
2,3-Dimethylhexane	115.607	−55.22	−74.74	−6.03	−23.21	0.08	2.60
2,4-Dimethylhexane	109.429	−64.71	−71.44	−12.21	−42.30	0.18	2.49
2,5-Dimethylhexane	109.103	−74.98	−62.28	−12.54	−25.41	−0.40	−0.07
3,3-Dimethylhexane	111.969	−48.92	−59.42	−11.91	−69.89	0.24	2.20
3,4-Dimethylhexane	117.725	−46.47	−91.59	−3.91	−25.40	−0.06	3.82
2-Methyl-3-ethylpentane	115.650	−45.24	−84.18	−5.99	−31.78	0.40	5.04
3-Methyl-3-ethylpentane	118.259	−32.69	−81.57	−5.62	−50.69	0.47	1.59
2,2,3-Trimethylpentane	109.841	−37.79	−33.09	−21.02	−93.53	−0.07	4.55
2,2,4-Trimethylpentane	99.238	−58.34	−15.25	−31.62	−74.01	3.24	7.94
2,3,3-Trimethylpentane	114.760	−29.56	−37.66	−16.10	−97.22	1.29	3.40
2,3,4-Trimethylpentane	113.467	−40.79	−48.43	−15.15	−64.09	0.95	5.89
2,2,3,3-Tetramethylbutane	106.300	−24.77	10.78	−33.78	−79.86	9.91	13.26
	s	60.3	80.3	15.9	54.1	2.6	4.8
	$\|\Delta\|_{max}$	91.54	131.07	33.78	97.22	9.91	13.26

the sets including thirteen and twenty-two alkanes. The results of T_b modeling for $N = 40$ alkanes are listed in Table 7.

Thus, we demonstrated the possibility of design of *ad hoc* topological indices. The design procedure is based on solution of Eqn. (1) and gives the best approximation to the "ideal" topological index. Moreover, analysis of the solution makes it possible to analyze the significance of particular subgraphs and thus decide which of them can be ignored in constructing a QSPR model for a given set and property P. The examples considered above allow one to hope that the *ad hoc* indices introduced in this work will also be most useful for prediction of the properties of hypothetical molecules.

Because all topological indices in the vector form can be expressed as linear combinations of the columns of the

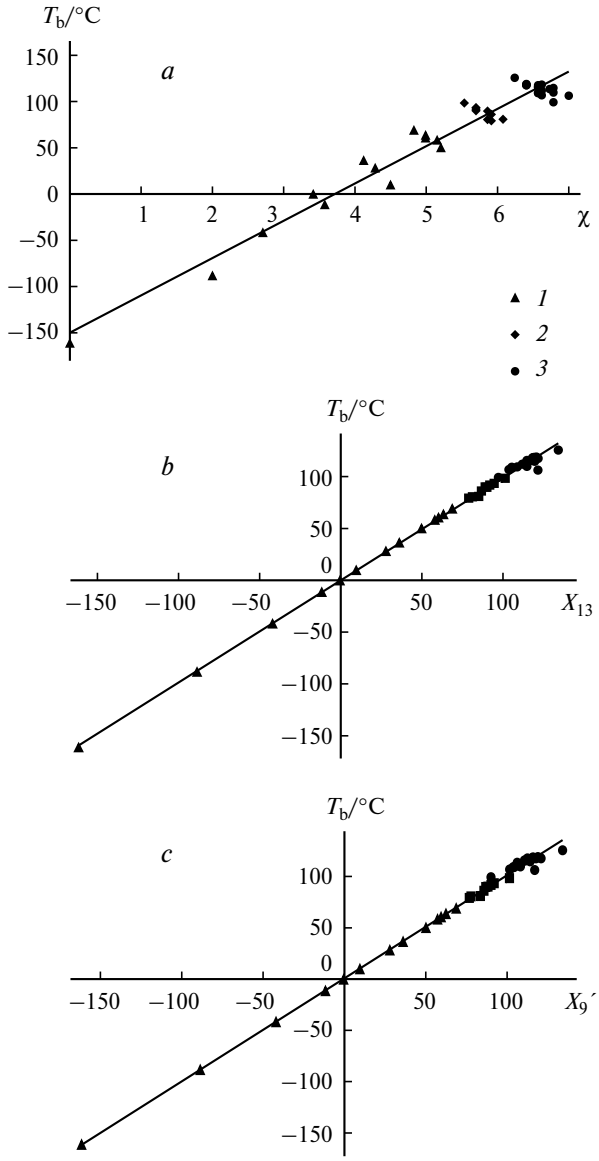

Fig. 2. Boiling points plotted *vs.* Randič index (*a*) and optimal indices X_{13} (*b*), X_9' (*c*): isomers C_1—C_6 (*1*), C_7 (*2*), and C_8 (*3*); solid line denotes the linear trend.

matrix *S*, Table 8 lists the coefficients at these vectors (because the columns of the CSM are the basis vectors[3]), which show their significance for the design of topological indices. For instance, the indices *W* and *Z* depend on a small number of the basis vectors that have small but different coefficients, the index χ depends on nine vectors, but the coefficients are independent of the property and therefore the use of this index seems to be unsuccessful and requires further studies. The index *K* depends on all basis vectors with the same coefficients, which assumes equivalence of all substructures with respect to the property under study.

Table 5. Boiling points calculated (C_1—C_6) and predicted (C_7, C_8) using logarithms of topological indices ($N = 13$)

Compound	T_b/°C	*W*	*Z*	*K*
Methane	−161.58	—	−14.85	−0.67
Ethane	−88.63	11.59	−3.14	−1.21
Propane	−42.06	−4.52	7.60	−1.01
n-Butane	−0.5	−4.38	4.03	6.39
2-Methylpropane	−11.72	−10.84	12.53	−11.21
n-Pentane	36.073	0.85	−0.92	15.84
2-Methylbutane	27.852	−2.61	2.66	−0.76
2,2-Dimethylpropane	9.5	−15.63	14.03	−29.98
n-Hexane	68.74	8.22	−11.15	26.00
2-Methylpentane	60.271	3.80	−4.86	8.60
3-Methylpentane	63.282	8.25	−9.54	8.88
2,2-Dimethylbutane	49.741	−0.69	2.34	−16.86
2,3-Dimethylbutane	57.988	5.97	1.28	−4.00
	s	8.1	8.7	14.3
	$\|\Delta\|_{max}$	15.63	14.85	29.98
n-Heptane	98.428	16.66	−23.83	36.44
2-Methylhexane	90.052	11.63	−18.59	19.13
3-Methylhexane	91.851	15.20	−21.57	18.87
3-Ethylpentane	93.475	18.67	−24.47	14.68
2,2-Dimethylpentane	79.198	6.32	−7.24	−8.30
2,3-Dimethylpentane	89.784	16.91	−13.81	5.63
2,4-Dimethylpentane	80.5	5.70	−12.03	−0.13
3,3-Dimethylpentane	86.064	15.20	−12.17	−6.16
2,2,3-Trimethylbutane	80.883	12.12	0.99	−12.84
	s	14.7	17.7	17.9
	$\|\Delta\|_{max}$	18.67	24.47	36.44
n-Octane	125.665	25.56	−39.17	46.87
2-Methylheptane	117.646	20.32	−33.13	30.15
3-Methylheptane	118.925	23.35	−37.74	26.71
4-Methylheptane	117.709	22.73	−36.06	23.99
3-Ethylhexane	118.534	25.40	−40.94	21.90
2,2-Dimethylhexane	106.840	14.34	−23.46	6.07
2,3-Dimethylhexane	115.607	23.75	−28.86	17.57
2,4-Dimethylhexane	109.429	16.93	−31.70	7.33
2,5-Dimethylhexane	109.103	14.73	−28.56	12.47
3,3-Dimethylhexane	111.969	22.09	−25.70	1.27
3,4-Dimethylhexane	117.725	27.18	−33.05	18.31
2-Methyl-3-ethylpentane	115.650	25.77	−32.03	14.88
3-Methyl-3-ethylpentane	118.259	30.45	−29.42	11.11
2,2,3-Trimethylpentane	109.841	22.75	−16.53	−6.39
2,2,4-Trimethylpentane	99.238	10.04	−14.18	−9.11
2,3,3-Trimethylpentane	114.760	28.39	−15.54	−3.56
2,3,4-Trimethylpentane	113.467	24.96	−20.59	3.93
2,2,3,3-Tetramethylbutane	106.300	22.94	2.71	−5.54
	s	23.6	29.7	19.2
	$\|\Delta\|_{max}$	30.45	40.94	46.87

Of course, even the best model for calculations and, moreover, prediction of T_b constructed in this work is far from being perfect. Nevertheless we believe that we have reached the limiting point for the index-based approach that ignores actual molecular geometry and details of

Table 6. Boiling points calculated (C_1—C_7) and predicted (C_8) using logarithms of topological indices ($N = 22$)

Compound	T_b/°C	W	Z	K
Methane	−161.58	—	−24.19	0.22
Ethane	−88.63	19.17	−7.57	−1.79
Propane	−42.06	−2.72	6.04	−2.52
n-Butane	−0.5	−6.41	6.09	4.19
2-Methylpropane	−11.72	−12.43	13.00	−13.54
n-Pentane	36.073	−4.07	4.46	13.09
2-Methylbutane	27.852	−7.09	7.09	−3.67
2,2-Dimethylpropane	9.5	−19.62	16.09	−33.11
n-Hexane	68.74	0.96	−2.33	22.80
2-Methylpentane	60.271	−3.08	2.78	5.22
3-Methylpentane	63.282	1.49	−1.28	5.45
2,2-Dimethylbutane	49.741	−7.02	8.56	−20.54
2,3-Dimethylbutane	57.988	−0.51	8.24	−7.58
n-Heptane	98.428	7.43	−11.62	32.86
2-Methylhexane	90.052	2.72	−7.46	15.37
3-Methylhexane	91.851	6.45	−10.06	15.07
3-Ethylpentane	93.475	10.09	−12.60	10.76
2,2-Dimethylpentane	79.198	−2.08	2.11	−12.39
2,3-Dimethylpentane	89.784	8.51	−3.09	1.60
2,4-Dimethylpentane	80.5	−2.88	−2.20	−4.09
3,3-Dimethylpentane	86.064	6.98	−1.88	−10.35
2,2,3-Trimethylbutane	80.883	4.10	9.81	−17.06
	s	8.5	9.6	15.1
	$\lvert\Delta\rvert_{max}$	19.62	24.19	33.11
n-Octane	125.665	14.65	−23.54	42.95
2-Methylheptane	117.646	9.66	−18.63	26.05
3-Methylheptane	118.925	12.85	−22.77	22.51
4-Methylheptane	117.709	12.29	−21.32	19.77
3-Ethylhexane	118.534	15.13	−25.74	17.62
2,2-Dimethylhexane	106.840	4.13	−10.60	1.71
2,3-Dimethylhexane	115.607	13.59	−14.86	13.26
2,4-Dimethylhexane	109.429	6.72	−17.97	2.94
2,5-Dimethylhexane	109.103	4.35	−15.11	8.19
3,3-Dimethylhexane	111.969	12.12	−12.24	−3.29
3,4-Dimethylhexane	117.725	17.14	−18.55	13.97
2-Methyl-3-ethylpentane	115.650	15.80	−17.77	10.52
3-Methyl-3-ethylpentane	118.259	20.67	−15.16	6.62
2,2,3-Trimethylpentane	109.841	13.03	−3.98	−11.06
2,2,4-Trimethylpentane	99.238	0.13	−2.67	−13.63
2,3,3-Trimethylpentane	114.760	18.74	−2.68	−8.28
2,3,4-Trimethylpentane	113.467	15.11	−7.43	−0.61
2,2,3,3-Tetramethylbutane	106.300	13.57	13.43	−10.13
	s	13.7	16.7	16.8
	$\lvert\Delta\rvert_{max}$	20.67	25.74	42.95

molecular interactions. Apparently, the boiling point and some other characteristics (*e.g.*, heat of melting, surface tension, melting point, *etc.*) are governed by the intermolecular interaction rather than molecular geometry. But because these interactions are implicitly originated from the molecular structure, the index-based approach can be improved using nonlinear (in particular, bilinear) functions[10] depending on complexity of the properties.[11]

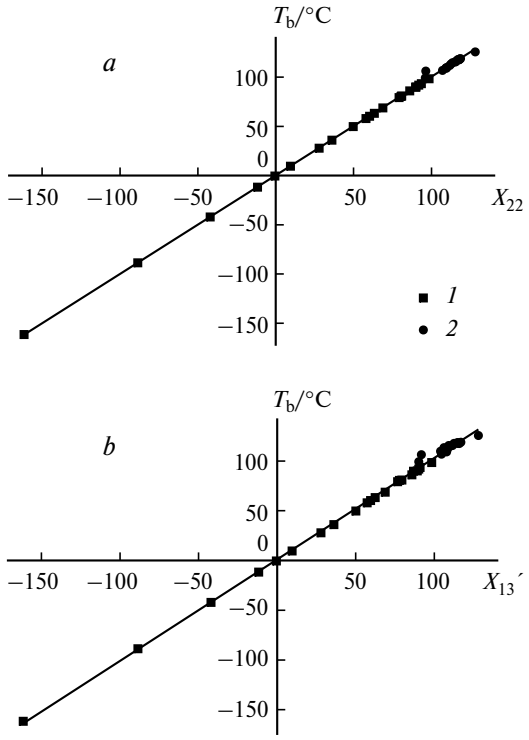

Fig. 3. Boiling points plotted *vs.* optimal indices X_{22} (*a*) and X_{13}' (*b*): isomers C_1—C_7 (*1*) and C_8 (*2*); solid line denotes the linear trend.

Table 7. Parameters of expression (2) for different indices ($N = 40$)

Index	A	B	R^2	s	$\lvert\Delta\rvert_{max}$
W	−36.43	2.28	0.797	27.7	125.15
Z	−27.56	5.61	0.721	32.5	139.63
χ	−149.92	40.34	0.955	13.0	26.16
K	−38.16	3.02	0.711	33.2	126.44
X_{13}	−0.10	0.99	0.998	3.1	13.44
X_9'	0.73	1.01	0.997	3.6	11.98
X_{22}	0.20	1.00	0.999	1.7	9.47
X_{13}'	1.01	1.02	0.998	2.8	11.65
X_{40}	0.00	1.00	1.000	0.0	0.00
X_{17}'	0.78	1.02	0.999	1.8	4.45
$\ln W$	−116.76	53.37	0.969	8.7	28.13
$\ln Z$	−125.89	74.71	0.972	10.3	35.69
$\ln K$	−162.90	69.59	0.939	15.2	39.20

Note. When using $\ln W$ as descriptor, methane was excluded ($N = 39$), because formally for methane $W = 0$ and the logarithm of 0 does not exist.

Now we will turn to the *ad hoc* topological indices. Following the design procedure for the X_9' and X_{13}' indices, we can always reduce the number of the basis vectors in a reasonable manner to retain good interpolation and extrapolation properties of the optimal topological indices. Moreover, by choosing the same basis vectors corre-

Table 8. Coefficients (a) of expansion of indices over the basis vectors as functions of the number of occurrences of the jth subgraph in the molecular graph

Index	a										
	1 (2)	3 (4)	5 (6)	7 (8)	9 (10)	11 (12)	13 (14)	15 (16)	17 (18)	19 (20)	21 (22)
W	— (1)	2 (3)	— (4)	— —	5 —	— —	— (6)	— —	— —	— —	— —
Z	1 —	— (1)	— (1)	— —	2 —	1 —	— (3)	— (1)	1 —	— —	— —
χ	— —	1.414 (−0.914)	−2.511 —	0.952 (3.557)	— —	— (−0.968)	−0.973 —	— —	— —	— —	— (0.982)
K	1 (1)	1 (1)	1 (1)	1 (1)	1 (1)	1 (1)	1 (1)	1 (1)	1 (1)	1 (1)	1 (1)
X_{13}	−161.58 (234.53)	−26.38 (−5.01)	10.15 (−4.987)	3.022 (−3.04)	−3.906 (0.833)	0.845 (−0.365)	0.75 —	— —	— —	— —	— —
X_9'	−161.58 (234.53)	−26.38 (−5.01)	10.15 (−4.987)	3.022 (−3.04)	−3.906 —	— —	— —	— —	— —	— —	— —
X_{22}	−161.58 (234.53)	−26.38 (−5.01)	−10.15 (−4.987)	3.022 (−3.04)	−3.906 (0.833)	0.845 (−0.365)	0.75 (−2.979)	1.02 (0.056)	−1.095 (0.359)	−0.325 (−0.648)	−0.621 (−4.185)
X_{13}'	−161.58 (234.53)	−26.38 (−5.01)	10.15 (−4.987)	3.022 (−3.04)	−3.906 —	— —	— (−2.979)	1.02 —	−1.095 —	— —	— (−4.185)

sponding to a certain index (*e.g.*, chains for the Wiener index or C_i—C_j subgraphs for the Randič index) we can obtain a modified index for a particular property by solving a "truncated equation" (1). Of course this will to some extent "improve" the index; however, the solution to Eqn. (1) is characterized by the minimum level of limitations and therefore the index obtained can be considered optimal.

References

1. M. I. Stankevich, I. V. Stankevich, and N. S. Zefirov, *Usp. Khim.*, 1988, **57**, 337 [*Russ. Chem. Rev.*, 1988, **57** (Engl. Transl.)].
2. T. G. Schmalz, D. J. Klein, and B. L. Sandleback, *J. Chem. Inf. Comput. Sci.*, 1992, **32**, 54.
3. E. A. Smolenskii, *Izv. Akad. Nauk, Ser. Khim.*, 2006, 1447 [*Russ. Chem. Bull., Int. Ed.*, 2006, 1501].
4. M. Randić, *J. Am. Chem. Soc.*, 1975, **97**, 6609.
5. H. Hosoya, *Bull. Chem. Soc. Jpn*, 1971, **44**, 2332.
6. H. Wiener, *J. Am. Chem. Soc.*, 1947, **69**, 17.
7. H. Wiener, *J. Am. Chem. Soc.*, 1947, **69**, 2636.
8. D. Bonchev, E. Marcel, and A. Dekmezian, *J. Chem. Inf. Comput. Sci.*, 2001, **41**, 1274.
9. A. R. Katrizky, R. Petrukhin, R. Jain, and M. Karelson, *J. Chem. Inf. Comput. Sci.*, 2001, **41**, 1521.
10. A. V. Kamernitzkii, E. A. Smolenskii, G. M. Makeev, I. V. Vesela, N. M. Mirsalikhova, A. M. Turuta, and N. S. Zefirov, *Bioorg. Khim.*, 2002, **28**, 269 [*Russ. J. Bioorg. Chem.*, 2002, **28** (Engl. Transl.)].
11. E. A. Smolenskii, *Dokl. Akad. Nauk*, 1999, **365**, 767 [*Dokl. Chem.*, 1999 (Engl. Transl.)].

Received November 7, 2005;
in revised form May 16, 2006