

On Variable Zagreb Indices*

Ante Miličević and Sonja Nikolić**

The Rugjer Bošković Institute, Bijenička 54, P.O. Box 180, HR-10002 Zagreb, Croatia

RECEIVED APRIL 15, 2003; REVISED JUNE 13, 2003; ACCEPTED JUNE 16, 2003

Key words
modified Zagreb indices
variable Zagreb indices
Zagreb indices

A novel variant of the Zagreb indices introduced in 1972, called the variable Zagreb indices, is presented. Their performance in the QSPR modeling of lower (C_3 – C_8) alkane boiling points is compared with the original Zagreb indices and their modifications. The variable Zagreb indices produce better single-descriptor models and participate in better two- and three-descriptor models.

INTRODUCTION

The Zagreb M_1 and M_2 indices were introduced¹ and applied to the branching problem² in the early seventies of the last century when only the Wiener index³ and the Hosoya Z-index⁴ were known and used in QSPR.^{5,6} Thirty years later there is still interest in Zagreb indices⁷ and consequently the overall Zagreb indices⁸ and their modified forms⁹ are proposed and used.

The Zagreb indices and their variants have been used to study molecular complexity^{10–14} and chirality¹⁵ whilst the overall Zagreb indices⁸ exhibit a potential applicability for deriving multilinear regression models. Various researchers also use the Zagreb indices in their QSPR and QSAR studies.^{5–8,15,16–24} Mathematical properties of the Zagreb indices were also subjects of several studies.^{25–27} These indices are referred to in most books reporting topological indices and their uses in QSPR and QSAR,^{5–7,28} and they are also included in a number of programs used for the routine computation of topological indices, e.g. Refs. 29–32.

Here we report a further development of the Zagreb indices, the so-called variable Zagreb indices.

DEFINITION OF ORIGINAL ZAGREB INDICES

The original Zagreb indices are defined as follows:^{1,2}

$$M_1 = \sum_{\text{vertices}} d(i) d(i) \quad (1)$$

$$M_2 = \sum_{\text{edges}} d(i) d(j) \quad (2)$$

where $d(i)$ is the degree of vertex i and $d(i) d(j)$ is the weight of edges i – j .

DEFINITION OF MODIFIED ZAGREB INDICES

A problem with the Zagreb indices, like with many other topological indices, is that their contributing parts give greater weights to the inner (interior) vertices and edges than to the outer (terminal) vertices and edges of a graph, as it can be seen from Figure 1.

This opposes intuitive reasoning that the outer atoms and bonds should have greater weights than the inner atoms and bonds because the outer atoms and bonds are associated with the larger part of the molecular surface and

* Dedicated to Professor Nenad Trinajstić, who is our scientific father, on the occasion of his 65th birthday.

** Author to whom correspondence should be addressed. (E-mail: sonja@irb.hr)

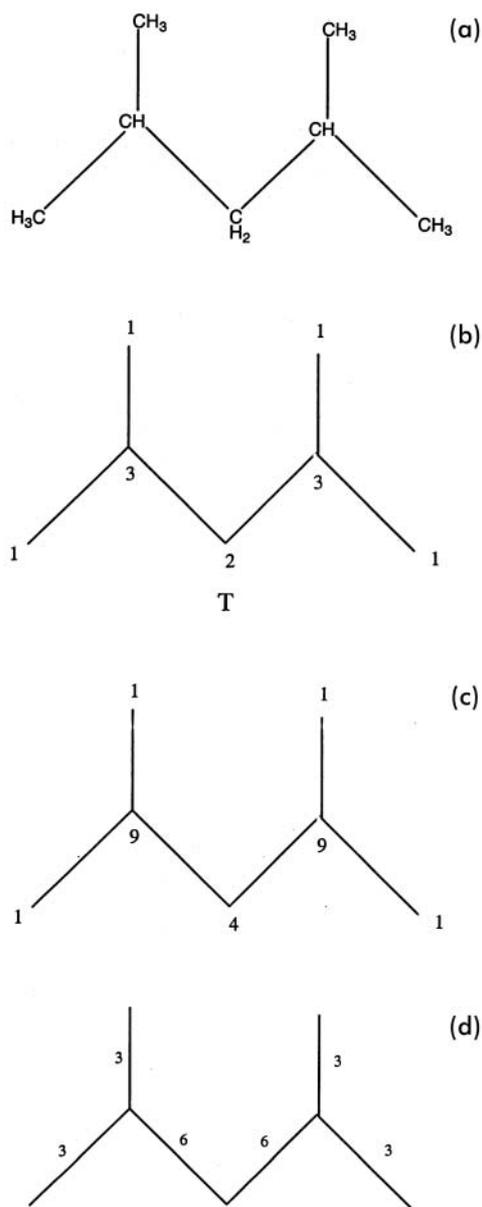


Figure 1. Contributing weights to the M_1 and M_2 indices of 2,4-dimethylpentane. (a) 2,4-dimethylpentane. (b) Hydrogen-suppressed tree T corresponding to 2,4-dimethylpentane and the vertex-degrees. (c) Vertex-weights contributing to the M_1 index of T. (d) Edge-weights contributing to the M_2 index of T.

are consequently expected to make a greater contribution to physical, chemical and biological properties.

Nikolić *et al.*⁹ amended the Zagreb indices by putting inverse values of the vertex-degrees into Eqs. (1) and (2). These indices were named the modified Zagreb indices and were denoted by mM_1 and mM_2 . They are given below:

$${}^mM_1 = \sum_{\text{vertices}} 1/[d(i) d(i)] = \sum_{\text{vertices}} [d(i) d(i)]^{-1} \quad (3)$$

$${}^mM_2 = \sum_{\text{edges}} 1/[d(i) d(j)] = \sum_{\text{edges}} [d(i) d(j)]^{-1} \quad (4)$$

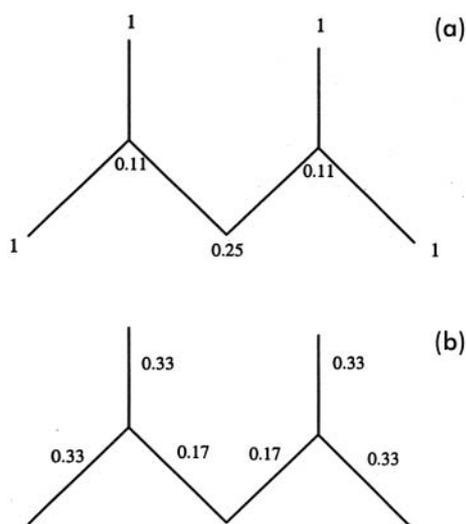


Figure 2. Contributing weights to the mM_1 and mM_2 indices of 2,4-dimethylpentane. (a) Vertex-weights contributing to the mM_1 index of T. (b) Edge-weights contributing to the mM_2 index of T.

Values mM_1 and mM_2 for the 2,4-dimethylpentane tree are given in Figure 2.

Comparing M_1 and M_2 indices with mM_1 and mM_2 for the 2,4-dimethylpentane tree we can see that modified Zagreb indices give greater weights to the outer vertices and edges than to the inner vertices and edges in accordance with the chemist's intuition.

VARIABLE ZAGREB INDICES

Eqs. (1) and (3) and Eqs. (2) and (4) can be collectively expressed as:

$${}^vM_1 = \sum_{\text{vertices}} [d(i) d(i)]^\lambda \quad (5)$$

$${}^vM_2 = \sum_{\text{edges}} [d(i) d(j)]^\lambda \quad (6)$$

where vM_1 and vM_2 denote variable Zagreb M_1 and M_2 indices. Obviously, for $\lambda = 1$ Eq. (5) reduces to Eq. (1) and Eq. (6) to Eq. (2), and for $\lambda = -1$ Eq. (5) reduces to Eq. (3) and Eq. (6) to Eq. (4). In this work we searched for the optimal value of λ . In order to determine λ , we used the set of C_3 – C_8 alkanes and their boiling points. This set of alkanes was selected because a number of papers report the modeling of their boiling points.^{9,33–35}

We considered the following a single-descriptor structure-boiling point relationship:

$$\text{bp} = a + b \text{ TI} \quad (7)$$

where bp stands for the boiling point and TI (topological index) for the variable Zagreb indices vM_1 and vM_2 .

We also considered the two-descriptor relationship:

$$\text{bp} = a + b \text{ TI} + c \text{ TI}' \quad (8)$$

where TI' was either p_3 or twc . p_3 is the polarity index³, which is equal to the number of paths of length 3 and encodes the steric aspects of an alkane.³⁶ twc is the total walk count³⁴ and is shown to be a very useful descriptor in building multivariate QSPR models.

Finally, we tested the three-descriptor relationship:

$$\text{bp} = a + b \text{ TI} + c p_3 + d \text{ twc} \quad (9)$$

We varied the exponent for ${}^{\vee}M_1$ and ${}^{\vee}M_2$ from $\lambda = 1$ to $\lambda = -1$ in order to detect that values of λ for which the variable Zagreb indices give the structure-boiling point models with the lowest value of the standard error of estimate s and the highest value of the correlation coefficient R . This is the origin of the term variable in the names of ${}^{\vee}M_1$ and ${}^{\vee}M_2$.

RESULTS AND DISCUSSION

We used data for C₃–C₈ alkanes from our previous paper.⁹ The structure-boiling point modeling was based on the CROMRsel procedure.^{37–40} This is a multivariate procedure, which selects the best possible model among the set of models obtained for a given number of parameters, the criterion being the standard error of estimate. The quality of models is expressed by fitted statistical parameters: the correlation coefficient r_{fit} , the standard error of estimate s_{fit} while F is the result of Fisher's test. The models were also cross-validated by using the leave-one-out procedure. Statistical parameters for the cross-validated models are denoted by r_{cv} and s_{cv} , where cv stands for the cross-validation procedure.

The standard errors of estimate, s_{fit} and s_{cv} , were computed using the following expressions:

$$s_{\text{fit}} = \sqrt{\left[\sum_{i=1}^N (P_i - P(\text{fit})_i^{\text{est}})^2 \right] / (N - I - 1)} \quad (10)$$

$$s_{\text{cv}} = \sqrt{\left[\sum_{i=1}^N (P_i - P(\text{cv})_i^{\text{est}})^2 \right] / (N - I - 1)} \quad (11)$$

where P_i denotes the values of a given experimental property, $P(\text{fit})_i^{\text{est}}$ estimated values based on the fit statistical procedure, $P(\text{cv})_i^{\text{est}}$ estimated values based on the leave-one-out cross-validation procedure, N is the number of data points used in the model building and I is the number of descriptors contained in the model.

The following results were obtained:

1. ${}^{\vee}M_1^{[\lambda]}$

Note that ${}^{\vee}M_1^{[\lambda]}$ is a variable Zagreb M_1 index with the optimum exponent λ .

(i) The best single-descriptor model

$$\text{bp} = -108.9 (\pm 4.5) + 21.84 (\pm 0.50) {}^{\vee}M_1^{[\lambda = 1/4]} \quad (12)$$

$$N = 38 \quad r_{\text{fit}} = 0.991 \quad r_{\text{cv}} = 0.989 \\ s_{\text{fit}} = 5.75 \quad s_{\text{cv}} = 6.15 \quad F = 1880$$

(ii) The best two-descriptor model:

– With p_3

$$\text{bp} = -103.8 (\pm 5.9) + 1.00 (\pm 0.74) p_3 + \\ 20.68 (\pm 0.99) {}^{\vee}M_1^{[\lambda = 1/4]} \quad (13)$$

$$N = 38 \quad r_{\text{fit}} = 0.991 \quad r_{\text{cv}} = 0.989 \\ s_{\text{fit}} = 5.68 \quad s_{\text{cv}} = 6.31 \quad F = 962$$

– With twc

$$\text{bp} = -118.1 (\pm 5.3) - 0.0071 (\pm 0.0025) \text{twc} + \\ 23.42 (\pm 0.73) {}^{\vee}M_1^{[\lambda = 1/4]} \quad (14)$$

$$N = 38 \quad r_{\text{fit}} = 0.992 \quad r_{\text{cv}} = 0.991 \\ s_{\text{fit}} = 5.26 \quad s_{\text{cv}} = 5.77 \quad F = 1125$$

(iii) The best three-descriptor model

$$\text{bp} = -110.3 (\pm 4.1) + 3.64 (\pm 0.64) p_3 - \\ 0.0154 (\pm 0.0023) \text{twc} + 21.11 (\pm 0.67) {}^{\vee}M_1^{[\lambda = 1/4]} \quad (15)$$

$$N = 38 \quad r_{\text{fit}} = 0.996 \quad r_{\text{cv}} = 0.995 \\ s_{\text{fit}} = 3.82 \quad s_{\text{cv}} = 4.32 \quad F = 1432$$

It is interesting to note that, in all cases, the best model was always obtained with ${}^{\vee}M_1^{[\lambda = 1/4]}$.

2. ${}^{\vee}M_2^{[\lambda]}$

Note that ${}^{\vee}M_2^{[\lambda]}$ is a variable Zagreb M_2 index with the optimum exponent λ .

(i) The best single-descriptor model

$$\text{bp} = -112.3 (\pm 3.6) + 50.49 (\pm 0.90) {}^{\vee}M_2^{[\lambda = -1/3]} \quad (16)$$

$$N = 38 \quad r_{\text{fit}} = 0.994 \quad r_{\text{cv}} = 0.993 \\ s_{\text{fit}} = 4.48 \quad s_{\text{cv}} = 4.87 \quad F = 3122$$

(ii) The best two-descriptor model:

– With p_3

$$\text{bp} = -103.8 (\pm 4.0) + 1.68 (\pm 0.49) p_3 + \\ 46.1 (\pm 1.5) {}^{\vee}M_2^{[\lambda = -1/3]} \quad (17)$$

$$N = 38 \quad r_{\text{fit}} = 0.996 \quad r_{\text{cv}} = 0.995 \\ s_{\text{fit}} = 3.92 \quad s_{\text{cv}} = 4.47 \quad F = 2046$$

– With twc

$$bp = -112.4 (\pm 4.4) - 0.0001 (\pm 0.0020) twc + 50.5 (\pm 1.3) {}^vM_2^{[\lambda = -1/3]} \quad (18)$$

$$N = 38 \quad r_{\text{fit}} = 0.994 \quad r_{\text{cv}} = 0.992 \\ s_{\text{fit}} = 4.54 \quad s_{\text{cv}} = 5.26 \quad F = 1518$$

Because of the negligible contribution from the twc , this model is practically identical to a single-parameter model (16).

(iii) The best three-descriptor model

$$bp = -105.9 (\pm 3.4) + 3.19 (\pm 0.56) p_3 - 0.0079 (\pm 0.0020) twc + 46.1 (\pm 1.2) {}^vM_2^{[\lambda = -1/3]} \quad (19)$$

$$N = 38 \quad r_{\text{fit}} = 0.997 \quad r_{\text{cv}} = 0.996 \\ s_{\text{fit}} = 3.29 \quad s_{\text{cv}} = 4.03 \quad F = 1942.8$$

Similarly as in the case of the variable Zagreb M_1 index, the best models always contain vM_2 with the exponent $\lambda = -1/3$.

The above results show that the variable Zagreb M_2 index produces slightly better models than the variable Zagreb M_1 index.

Table I gives a selection of the structure-boiling point models from the literature with up to three descriptors for C_3 – C_8 alkanes.

Apparently, the variable Zagreb indices, with the optimal λ value, produce better single-descriptor structure-boiling point models than the original Zagreb indices, their modified forms and the overall Zagreb indices, and many other indices. They are comparable to a single descriptor model⁸ based on the connectivity index ($s_{\text{fit}} = 5.75$ °C).⁴⁹ The two-descriptor model based on the variable Zagreb M_2 index with $\lambda = -1/3$ ($s_{\text{fit}} = 3.92$ °C) is better than most two-descriptor models in the literature, it is comparable to the model based on ${}^1\chi$ and M_2 indices⁹ ($s_{\text{fit}} = 3.84$ °C) and is poorer than the Rucker-Rucker model³⁴ based on $\chi^{1/3}$ and mwc_5 ($s_{\text{fit}} = 2.21$ °C). The three-descriptor model with ${}^vM_2^{[\lambda = -1/3]}$ ($s_{\text{fit}} = 3.29$ °C) is comparable to some models based on three descriptors, but is somewhat poorer than the model⁹ based on ${}^1\chi$, OM_1 and twc ($s_{\text{fit}} = 2.57$ °C) and much poorer than the model³⁴ based on p_3 , $W^{0.25}$ and V ($s_{\text{fit}} = 1.38$ °C). Models with four and five different descriptors produce the standard error of estimate even lower than 1 °C.

CONCLUDING REMARKS

We note that the variable vertex-connectivity index⁵⁰ is formally identical to the variable Zagreb M_2 index. Thus, one may say that the meeting point of these two topological indices is reached in their variable forms: ${}^vM_2 = {}^v({}^1\chi)$. The reader should be reminded that the road

TABLE I. Fitted statistical parameters of some structure-boiling point models with up to three descriptors for C_3 – C_8 alkanes taken from the literature

Descriptor	$I^{(a)}$	r_{fit}	s_{fit}	Ref.
${}^0M_1^{(b)}$	1	0.875	20.3	8
${}^0M_2^{(b)}$	1	0.985	7.2	8
TC ^(c)	1	0.753	27.6	41
TC1 ^(c)	1	0.732	28.6	41
$S^{(d)}$	1	0.795	25.4	41
${}^0\chi^{(e)}$	1	0.954	12.5	41
${}^1\chi^{(f)}$	1	0.990	5.75	9
${}^0\text{TC}, {}^1\text{TC}^{(g)}$	2	0.9933	4.92	41
${}^0M_1, {}^1M_1$	2	0.989	6.22	8
${}^0M_2, {}^1M_2$	2	0.990	6.12	8
${}^0\chi, {}^2\chi$	2	0.9927	5.13	41
$\chi^{0.33}, {}^{(h)}mwc_5^{(i)}$	2		2.21	34
${}^1\chi, M_2$	2	0.996	3.84	9
${}^0\text{TC}, {}^1\text{TC}, {}^1\text{TC1}^{(j)}$	3	0.9962	3.75	41
${}^0\chi, {}^2\chi, {}^5\chi$	3	0.9943	4.58	41
${}^1\chi, OM_1, twc$	3	0.998	2.57	9
$p_3, {}^{(k)}W^{0.25}, {}^{(l)}V^{(m)}$	3		1.38	34

(a) I = the number of descriptors in the model.

(b) ${}^0M_1, {}^0M_2$ = the overall Zagreb indices.⁸

(c) TC, TC1 = total values of topological complexity indices.⁴¹

(d) S = total number of connected subgraphs of a graph.⁴¹

(e) ${}^0\chi$ = the zero-order connectivity index.^{42,43}

(f) ${}^1\chi$ = the first-order connectivity index.⁴⁴

(g) ${}^e\chi$ ($e = 0, 1, \dots, E$) = the connectivity indices from the zero-order to the E -order.^{42,43} E is the total number of edges in a molecular graph.

(h) $\chi^{0.33}$ = cubic root of the connectivity index.⁴⁵

(i) mwc_5 = molecular walk count of order 5.^{34,46–48}

(j) ${}^e\text{TC}$ ($e = 0, 1, \dots, E$), ${}^e\text{TC1}$ ($e = 1, 2, \dots, E$) = e^{th} order topological complexity indices.⁴¹

(k) p_3 = number of paths of length 3.^{5,36}

(l) W^α = the fourth root of the Wiener index.^{3,49}

(m) V = the number of vertices in a molecular graph.

to the first-order connectivity index (${}^1\chi$) was⁹ the Zagreb M_2 index \Rightarrow modified Zagreb M_2 index \Rightarrow ${}^1\chi$. However, this road was not taken; the first order vertex-connectivity index was derived in quite a different way⁴⁴ than the Zagreb M_2 index.¹

Acknowledgments. – This work was supported in part by the Ministry of Science and Technology of the Republic of Croatia by Grant No. 0098034. We thank Dr Bono Lučić for comments and Professor Nenad Trinajstić for all he has taught us.

REFERENCES

- I. Gutman and N. Trinajstić, *Chem. Phys. Lett.* **17** (1972) 535–538.
- I. Gutman, B. Ruščić, N. Trinajstić and C. F. Wilcox, Jr., *J. Chem. Phys.* **62** (1975) 3399–3405.
- H. Wiener, *J. Am. Chem. Soc.* **69** (1947) 17–20.
- H. Hosoya, *Bull. Chem. Soc. Jpn.* **44** (1971) 2332–2339.
- J. Devillers and A. T. Balaban, (Ed.), *Topological Indices and Related Descriptors in QSAR and QSPR*, Gordon & Breach, Amsterdam, 1999.

6. M. Karelson, *Molecular Descriptors in QSAR/QSPR*, Wiley-Interscience, New York, 2000.
7. R. Todeschini and V. Consonni, *Handbook of Molecular Descriptors*, Wiley-VCH, Weinheim, 2000.
8. D. Bonchev and N. Trinajstić, *SAR QSAR Environ. Res.* **12** (2001) 213–236.
9. S. Nikolić, G. Kovačević, A. Miličević, and N. Trinajstić, *Croat. Chem. Acta* **76** (2003) 113–124.
10. S. H. Bertz and W. F. Wright, *Graph Theory Notes of New York* **35** (1998) 32–48.
11. S. Nikolić, I. M. Tolić, and N. Trinajstić, *MATCH – Commun. Math. Comput. Chem.* **40** (1999) 187–201.
12. S. Nikolić, N. Trinajstić, and I. M. Tolić, *J. Chem. Inf. Comput. Sci.* **40** (2000) 920–926.
13. S. Nikolić, I. M. Tolić, N. Trinajstić, and I. Baučić, *Croat. Chem. Acta* **73** (2000) 909–921.
14. S. Nikolić, N. Trinajstić, I. M. Tolić, G. Rücker, and C. Rücker, in: D. Bonchev and D. H. Rouvray (Eds.), *Complexity in Chemistry – Introduction and Fundamentals*, Taylor & Francis, London, 2003, pp. 29–89.
15. A. Golbraikh, D. Bonchev, and A. Tropsha, *J. Chem. Inf. Comput. Sci.* **41** (2001) 147–158.
16. A. T. Balaban, *Theor. Chim. Acta* **53** (1979) 355–375.
17. A. T. Balaban, *Chem. Phys. Lett.* **89** (1982) 399–404.
18. S. C. Basak and G. D. Grunwald, *SAR QSAR Environ. Res.* **3** (1995) 265–277.
19. S. C. Basak, G. D. Grunwald, and G. J. Niemi, in: A. T. Balaban (Ed.), *From Chemical Topology to Three-Dimensional Geometry*, Plenum Press, New York, 1997, pp. 73–116.
20. S. C. Basak, B. D. Gute, B. Lučić, S. Nikolić, and N. Trinajstić, *Comput. Chem.* **24** (2000) 181–191.
21. S. C. Basak, A. T. Balaban, G. D. Grunwald, and B. D. Gute, *J. Chem. Inf. Comput. Sci.* **40** (2000) 891–898.
22. X. Yao, X. Zhang, R. Zhang, M. Liu, Z. Hu, and B. Fan, *Comput. Chem.* **25** (2001) 475–482.
23. P. V. Khadikar, S. Karmarkar, and V. K. Agrawal, *J. Chem. Inf. Comput. Sci.* **41** (2001) 934–949.
24. S. Sardana and A. K. Madan, *MATCH – Commun. Math. Comput. Chem.* **45** (2002) 35–53.
25. E. Estrada, L. Rodriguez, and A. Gutierrez, *MATCH – Commun. Math. Comput. Chem.* **35** (1997) 145–156.
26. E. Estrada, *Chem. Phys. Lett.* **336** (2001) 248–252.
27. D. Vukičević and N. Trinajstić, *Croat. Chem. Acta* **76** (2003) 183–187.
28. N. Trinajstić, *Chemical Graph Theory*, 2nd revised ed., CRC Press, Boca Raton, FL, 1992, Chapter 10.
29. S. C. Basak, D. K. Harriss, and V. R. Magnuson, POLLY 2.3, Copyright of the University of Minnesota, 1988.
30. R. Todeschini, DRAGON, WebSite – <http://www.disat.unimib.it/chm/>
31. Cerius² 3.0 QSAR+, Molecular Simulation Inc., 9685 Scranton Road, San Diego, CA 92121, USA.
32. M. Vedrina, S. Marković, S. Medić-Šarić, and N. Trinajstić, *Comput. Chem.* **21** (1997) 355–361.
33. D. Bonchev, *SAR QSAR Environ. Res.* **7** (1997) 23–43.
34. G. Rücker and C. Rücker, *J. Chem. Inf. Comput. Sci.* **39** (1999) 788–802.
35. D. Bonchev, *J. Mol. Graphics Modell.* **20** (2001) 65–75.
36. J. R. Platt, *J. Phys. Chem.* **15** (1947) 328–336.
37. B. Lučić and N. Trinajstić, *J. Chem. Inf. Comput. Sci.* **39** (1999) 121–132.
38. B. Lučić, A. Miličević, S. Nikolić, and N. Trinajstić, *Croat. Chem. Acta* **75** (2002) 847–868.
39. S. C. Basak, B. D. Gute, B. Lučić, S. Nikolić, and N. Trinajstić, *Comput. Chem.* **24** (2000) 181–191.
40. B. Lučić, I. Bašić, D. Nadramija, A. Miličević, N. Trinajstić, T. Suzuki, R. Petrukhin, M. Karelson, and A. R. Katritzky, *Arkivoc* 2002, 45–49.
41. D. Bonchev, *SAR QSAR Environ. Res.* **7** (1997) 23–43.
42. L. B. Kier, L. H. Hall, W. J. Murray, and M. Randić, *J. Pharm. Sci.* **64** (1975) 1971–1974.
43. L. B. Kier, L. H. Hall, W. J. Murray, and M. Randić, *J. Pharm. Sci.* **65** (1976) 1226–1230.
44. M. Randić, *J. Am. Chem. Soc.* **97** (1975) 6609–6615.
45. D. Amić, D. Bešlo, B. Lučić, S. Nikolić, and N. Trinajstić, *J. Chem. Inf. Comput. Sci.* **38** (1998) 819–822.
46. G. Rücker and C. Rücker, *J. Chem. Inf. Comput. Sci.* **33** (1993) 683–695.
47. G. Rücker and C. Rücker, *J. Chem. Inf. Comput. Sci.* **40** (2000) 99–106.
48. G. Rücker and C. Rücker, *J. Chem. Inf. Comput. Sci.* **34** (1994) 534–538.
49. B. Lučić, A. Miličević, S. Nikolić, and N. Trinajstić, *Indian J. Chem., Sect. A*, in press.
50. D. Amić, D. Bešlo, B. Lučić, S. Nikolić, and N. Trinajstić, *J. Chem. Inf. Comput. Sci.* **38** (1998) 819–822.

SAŽETAK

O varijabilnim Zagrebačkim indeksima

Ante Miličević i Sonja Nikolić

Predložena je nova inačica Zagrebačkih indeksa, uvedenih 1972., koja je nazvana varijabilni Zagrebački indeksi. Njihova uloga u modeliranju točke vrelišta nižih (C₃–C₈) alkana uspoređena je s izvornim Zagrebačkim indeksima i njihovim modifikacijama. Varijabilni Zagrebački indeksi vode do boljih jedno-deskriptorskih modela nego druge vrste Zagrebačkih indeksa, a sudjeluju u gradnji dobrih dvo- i tro-deskriptorskih modela.