J.Serb.Chem.Soc. 65(8)577-580(2000)
JSCS-2779

UDC 54.02/.022/.084
Note

NOTE

On the application of line graphs in quantitative structure-property studies

IVAN GUTMAN* and ŽELJKO TOMOVIĆ#

Faculty of Science, University of Kragujevac, P. O. Box 60, YU-34000 Kragujevac, Yugoslavia

(Received 10 December 1999)

Let G be a molecular graph possessing $m_0(G)$ edges. Let $m_1(G)$ be the number of edges of the line graph $L(G)$ of G, known as the Bertz index. Let $m_2(G)$ be the number of edges of the line graph of $L(G)$, *etc.* We examine the applicability of the sequence $m_i(G)$, $i = 0, 1, 2, ...$, for predicting physicochemical properties of alkanes. Some earlier obtained results are extended and amended.

Keywords: line graph, Bertz index, QSPR, QSAR.

INTRODUCTION

The idea that quantitative structure-property relations (QSPR) and quantitative structure-activity relations (QSAR) could be designed on the basis of the invariants ("topological indices") of the line graph $L_1(G) = L(G)$ of the molecular graph G, instead of such invariants of the graph G itself, seems to have been first put forward by Bertz.[1] In particular, Bertz investigated the number $m_1(G)$ of edges of $L(G)$. The same author was also the first to consider[2,3] higher-order line graphs, namely $L_2(G) = L(L_1(G))$, $L_3(G) = L(L_2(G))$, *etc.*

In the following the number of edges of the i-th line graph $L_i(G)$ of the molecular graph G is denoted by $m_i(G)$, $i = 1, 2, ...$, and, in addition, the number of edges of G by $m_0(G)$. We call $m_i(G)$ the i-th Bertz index.

More details on line graphs, their chemical applications as well as an exhaustive list of references on this topic are to be found elsewhere;[4] for some recent works along these lines see the Refs.5–7.

In an earlier work[8] the surface tension of alkanes was approximated by means of a linear combination of the Bertz indices, namely by the model:

$$M(0,1,2,...,k) = a_0 m_0 + a_1 m_1 + a_2 m_2 + ... + a_k m_k + b \qquad (1)$$

where a_0, a_1, a_2, ..., a_k, b were determined by least squares fitting, using the stan-

Serbian Chemical Society active member.

dard Needham-Wei-Seybold data base.[9] In a subsequent work[4] analogous analyses were performed for six additional physicochemical properties (from the same data base[9]). The parameter k in Eq. (1) was determined by the requirement that the approximation $M(0,1,2, ..., k)$ is significantly more accurate than $M(0,1,2, ..., k-1)$, and that $M(0,1,2, ..., k+1)$ is not significantly more accurate than $M(0,1,2, ..., k)$. Statistical significance was assessed by the F-test, at the 99 % confidence level.

The model (1) is biased because it assumes that the best choice of the Bertz indices is the set of the first $k+1$ such indices, $m_0, m_1, m_2, .., m_k$, for some pertinently chosen value of k. Because the validity of this assumption is by no means obvious, we decided to test it by examining a more general QSPR approximation. Another bias in the works[4,8] was the assumption that $k \geq 1$. The present calculations clearly indicate (*cf.* Table I) that the possibility of $k = 0$ should not have been excluded from consideration.

THE GENERAL LINEAR MODEL BASED ON BERTZ INDICES

Instead of model (1) the physicochemical property considered are now approximated by the expression

$$M(i_0, i_1, i_2, ..., i_k) = a_0 m_{i0} + a_1 m_{i1} + a_2 m_{i2} + ... + a_k m_{ik} + b \qquad (2)$$

where $\{i_0, i_1, i_2, ..., i_k\}$ is a subset of $\{0, 1, 2, ..., 6\}$ and where $a_0, a_1, ..., a_k, b$ and k are determined in the same manner as in the case of (1). An approximation of the form (2) is now sought which is optimal within all possible choices of $k+1$ distinct Berzt indices, *i.e.*, within all possible permutations $(i_0, i_1, i_2, ..., i_k)$.

NUMERICAL WORK

The Needham-Wei-Seybold data base[9] contains the boiling points, critical temperatures and critical pressures for 74 alkanes, the molar volumes, molar refractions and heats of vaporization for 69 alkanes, as well as the surface tensions for 68 alkanes. The data base is often employed for testing the applicability of various topological indices for predicting physicochemical properties of non-polar organic compounds. The results communicated in this paper (as well as in earlier works[4,8]) pertain to this data base.

The first difficulty in connection with higher-order line graphs is that their size rapidly increases. As a consequence, the higher-order Birtz indices also rapidly become enormously large. A typical example is 3-ethyl-2-methylpentane, for which $m_0 = 7$, $m_1 = 8$, $m_2 = 14$, $m_3 = 41$, $m_4 = 213$, $m_5 = 2078$, $m_6 = 39449$, $m_7 = 1485676$. In view of this, the calculation of the Bertz indices was performed by a specially designed computer program, written in FORTRAN-90, which enabled the calculation of m_i, $i \leq 6$ for all alkanes from our sample, as well as m_7 and m_8 for some members of this sample.

In all cases studied, the one-parameter approximation $a_0 m_0 + b$ was found to be far better than any of the approximations $a_0 m_i + b$, $i = 1, 2, ..., 6$. For instance, for

the heats of vaporization (HV) the correlation coefficient for $a_0 m_0 + b$ is 0.96, whereas for $a_0 m_i + b$, $i = 1, 2, ..., 6$, the correlation coefficients are 0.59, 0.32, 0.23, 0.20, 0.18 and 0.17, respectively. The situation with the other physicochemical properties examined is similar.

Therefore in Eq. (2) $i_0 = 0$ was fixed and only the indices $i_1, i_2, ..., i_k$ were varied.

We then continued by setting $k = 1$ and eventually increasing it one-by-one. For any given value of k, model (2) was tested for all possible permutations ($i_1, i_2, ..., i_k$) and the best such model selected. The search was ended when no model of the type $M(0, i_1, i_2, ..., i_k, i_{k+1})$ was significantly better (as determined by the F-test at the 99 % confidence level), than the best model of the type $M(0, i_1, i_2, ... i_k)$.

RESULTS AND DISCUSSION

The optimal QSPR models found by the above described procedure, as well as the statistical data indicating their precision are given in Table I.

TABLE I. The best linear QSPR models, Eq. (2), found for some physicochemical properties of alkanes: boiling points (*BP*) at normal pressure [K], molar volumes (*MV*) at 293 K [cm^3/mol], molar refractions (*MR*) at 293 K [cm^3/mol], heats of vaporization (*HV*) at 298 K [kJ/mol], critical temperatures (*CT*) [K], critical pressures (*CP*) [10^2 kPa] and surface tensions (*ST*) [10^{-3} N/m]; experimental values taken from the Needham-Wei-Seybold data base[9]; for details see text

Property	Optimal model	k	Correlation coefficient	Average rel. error
BP	$a_0 m_0 + b$	0	0.986	1.7 %
MV	$a_0 m_0 + a_1 m_1 + a_2 m_2 + b$	2	0.997	0.65 %
MR	$a_0 m_0 + a_1 m_1 + a_2 m_2 + b$	2	0.9998	0.22 %
HV	$a_0 m_0 + a_1 m_1 + a_2 m_3 + b$	2	0.9985	0.62 %
CT	$a_0 m_0 + b$	0	0.977	1.9 %
CP	$a_0 m_0 + b$	0	0.929	4.6 %
ST	$a_0 m_0 + a_1 m_1 + a_2 m_2 + b$	2	0.974	1.7 %

Property	a_0	a_1	a_2	b
BP	28.897			181.777
MV	10.714	6.333	– 1.519	56.141
MR	4.265	0.454	– 0.109	7.033
HV	6.831	–2.010	0.040	5.383
CT	35.402			311.790
CP	– 2.549			44.490
ST	3.518	– 2.498	0.506	9.288

As can be seen from Table I, except in a single case (for the heats of vaporization, *HV*) the form of the optimal formula coincides with that given by Eq. (1). Therefore, the extended model, Eq. (2), put forward in this work, did not prove to be of great practical value. Yet, the newly found QSPR formula $HV \approx a_0 m_0 + a_1 m_1 + a_2 m_3 + b$ (average relative error 0.62 %, correlation coefficient 0.9985) is far more accurate than the earlier reported[4] approximation $HV \approx a_0 m_0 + a_1 m_1 + a_2 m_2 + b$ (average relative error 0.94 %, correlation coefficient 0.994).

As before,[4,8] there was no need to go beyond $k = 2$, that is to use linear combinations of more than three Bertz indices.

For three of the seven physicochemical properties studied (*BP*, *CT*, *CP*), the one-parameter approximation ($k = 0$ with $i_0 = 0$) suffices, at least when the *F*-test with 99 % confidence level is used as the criterion. This is remarkable since in the case of alkanes m_0 is just the number of carbon atoms minus one. Recall that in the earlier studies[4,8] it was required that $k \geq 1$.

Comparing the results of the present calculations with those reported earlier[4,8] some awkward differences may be noticed. After careful checking we regret to announce that some of the earlier results were not completely correct and that they neded to be substituted with those found in Table I.

ИЗВОД

О ПРИМЕНИ ГРАФОВА ГРАНА ПРИ ПРОУЧАВАЊУ КВАНТИТАТИВНИХ ОДНОСА ИЗМЕЂУ СТРУКТУРЕ И ФИЗИЧКОХЕМИЈСКИХ ОСОБИНА

ИВАН ГУТМАН и ЖЕЉКО ТОМОВИЋ

Природно-математички факултет у Крагујевцу

Нека је G молекулски граф са $m_0(G)$ грана. Нека је $m_1(G)$ број грана у графу грана $L(G)$ графа G, познат као Берцов индекс. Нека је $m_2(G)$ број грана графа гране графа $L(G)$, итд. Испитали смо могућност примене низа $m_i(G)$, $i = 0, 1, 2, \ldots$, за предвиђање физичкохемијских особина алкана. Неки раније добивени резултати су проширени и побољшани.

(Примљено 10. децембра 1999)

REFERENCES

1. S. H. Bertz, *J. Chem. Soc. Chem. Commun.* (1981) 818
2. S. H. Bertz, in *Chemical Application of Topology and Graph Theory*, R. B. King, Ed., Elsevier, Amsterdam, 1993, pp. 206–221
3. S. H. Bertz, *Discrete Math.* **19** (1988) 65
4. I. Gutman, L. Popović, E. Estrada, S. H. Bertz, *ACH Models in Chem.* **135** (1998) 147
5. E. Estrada, *J. Chem. Inf. Comput. Sci.* **39** (1999) 90
6. E. Estrada, L. Rodriguez, *J. Chem. Inf. Comput. Sci.* **39** (1999) 1037
7. E. Estrada, *J. Chem. Inf. Comput. Sci.* **39** (1999) 1042
8. I. Gutman, L. Popović, B. K. Mishra, M. Kuanar, E. Estrada, N. Guevara, *J. Serb. Chem. Soc.* **62** (1997) 1025
9. D. E. Needham, I. C. Wei, P. G. Seybold, *J. Am. Chem. Soc.* **110** (1988) 4186.