CROATICA CHEMICA ACTA
CCACAA, ISSN 0011-1643, e-ISSN 1334-417X
Croat. Chem. Acta **84** (1) (2011) 87–91.
CCA-3453

Original Scientific Article

Bond Additive Modeling 4. QSPR and QSAR Studies of the Variable Adriatic Indices

Damir Vukičević

Faculty of Natural Sciences and Mathematics, University of Split, Nikole Tesle 12, HR-21000 Split, Croatia
(E-mail: vukicevi@pmfst.hr)

RECEIVED JANUARY 13, 2010; REVISED FEBRUARY 10, 2011; ACCEPTED FEBRUARY 11, 2011

Abstract. In this paper we analyze the set of the variable Adriatic indices. We show that three of these indices show very good predictive properties. Namely, the *inverse sum -1.95-deg* index is well correlated with the standard enthalpy of formation of octane isomers $\left(R^2=0.75\right)$, the *inverse sum 0.43-lodeg* index is well correlated with the total surface area of octane isomers $\left(R^2=0.92\right)$ and the *sum 0.37-exdeg* index is well correlated with the octanol-water partition coefficient $\left(R^2=0.99\right)$. (doi: 10.5562/cca1666)

Keywords: molecular descriptor, molecular index, QSAR, QSPR, algorithm

INTRODUCTION

Let G be a simple connected graph. Denote by $V(G)$ the set of its vertices and by $E(G)$ the set of its edges, respectively. Let us observe two vertex invariants:

1) d_v degree of vertex v - number of edges incident to vertex v;
2) $D_v=\sum_{u\in V(G)} d(u,v)$, where $d(u,v)$ is distance between vertices u and v, *i.e.* the length of the shortest path between vertices u and v.

The set of 48 *variable Adriatic indices* is defined in Ref. 1 by the following procedure (motivation of the definition, choice of functions, and restrictions on a are commented in due detail in Ref. 1):

Procedure. Variable Adriatic Indice

1) Choose invariant p_v to be d_v or D_v;
2) Choose numbers $i\in\{1,2,3\}$ and $j\in\{1,2,...,8\}$;
3) Choose $a\in\mathbb{R}\setminus\{0\}$ if $i=2$ and $j\in\{1,2,...,5\}$; otherwise if $i=1$ or 2 choose $a\in\mathbb{R}^+$ and if $i=3$ choose $a\in\mathbb{R}^+\setminus\{1\}$.
4) Calculate $A(G)=\sum_{uv\in E(G)} \gamma_i\left(\phi_{j,a}(p_u),\phi_{j,a}(p_v)\right)$.

where:

- $\phi_{1,a}(x)=\log^a(x)$, $a>0$;
- $\phi_{2,a}(x)=x^a$, $a\in\mathbb{R}\setminus\{0\}$;
- $\phi_{3,a}(x)=a^x$, $a>0$;
- $\gamma_1(x,y)=x\cdot y$;
- $\gamma_2(x,y)=x+y$;
- $\gamma_3(x,y)=\begin{cases}\dfrac{1}{x+y}, & x+y\neq 0;\\ 0, & x+y=0;\end{cases}$
- $\gamma_4(x,y)=|x-y|$;
- $\gamma_5(x,y)=\begin{cases}\dfrac{1}{|x-y|}, & x\neq y;\\ 0, & x=y;\end{cases}$
- $\gamma_6(x,y)=\begin{cases}\dfrac{\min\{x,y\}}{\max\{x,y\}}, & \max\{x,y\}\neq 0;\\ 0, & \max\{x,y\}=0;\end{cases}$
- $\gamma_7(x,y)=\begin{cases}\dfrac{\max\{x,y\}}{\min\{x,y\}}, & \min\{x,y\}\neq 0;\\ 0, & \min\{x,y\}=0;\end{cases}$
- $\gamma_8(x,y)=\begin{cases}\dfrac{x}{y}+\dfrac{y}{x}, & x,y\neq 0;\\ 0, & \text{otherwise.}\end{cases}$

One can note that some famous molecular descriptors such as the Randić index[2] and the Zagreb index[3] are variable Adriatic indices. Namely, the Randić index is obtained by the use of the functions $\psi_{2,-1/2}$ and γ_1; and the Zagreb index is obtained using functions $\psi_{2,1}$ and γ_1.

In order to analyze the predictive properties of these indices, we use (similarly as in Ref. 1) the benchmark sets[4] proposed by the International Academy of Mathematical Chemistry.[5]

Namely, we observe four sets of chemical compounds:

1) the set of 18 octane isomers
2) the set of 82 polyaromatic hydrocarbons (PAH)
3) the set of 209 polychlorobiphenyls (PCB)
4) the set of 22 phenetylamines (Phenet)

16 properties and 102 descriptors are given for the set of octane isomers; 3 properties and 112 descriptors are given for PAHs; 8 properties and 106 descriptors for PCBs; and one property and 110 descriptors for the phenetylamines.

We exclude melting point from our observations since it does not predominantly depend on graph of the molecule.

We shall compare the best coefficient of determination of the one-parameter linear models based on the variable Adriatic indices with

1) the best coefficient of determination R^2 (equivalently correlation coefficient R) of the one-parameter linear model based on the descriptors in the benchmark sets;
2) the best coefficient of determination R^2 (equivalently correlation coefficient R) of the one-parameter linear model based on the discrete Adriatic indices.[1,6]

Note that this comparison is not completely fair. Namely the linear one-parameter models based on "non-variable" descriptors des depend on only two parameters to predict observed property $prop$; namely $prop \approx k \cdot des + l$ depends solely on k and l. On the other hand, variable descriptor $des(a)$ depends on three parameters to predict observed property $prop$, namely $prop \approx k \cdot des(a) + l$ depends on a, k and l. Hence, the same R^2 does not imply equally good predictive properties, because here we have one fitted parameter more.

Moreover, suppose that we observe the situation in which some discrete Adriatic index has better predictive properties than the benchmark descriptor. In this case, it is very much expected that the corresponding vaiable Adriatic index will make some improvement to R^2.

Taking all of this into account, we are not interested in variable Adriatic indices that make modest improvements of R^2, but only in descriptors that make significant improvements to R^2. It will be shown that there are three cases in which a large improvement of R^2 occur.

MAIN RESULTS

Note that in the Adriatic descriptors, parameter a is chosen from an infinite set of values (moreover from the set of values of cardinality c, *i.e.* the cardinality of the set of real numbers). Hence, it is not possible to calculate the correlation for each of these values. Mathematical optimizations of R^2 would be quite involved and the solutions of obtained equations would not be exactly solvable for most of these descriptors (since they involve logarithms and exponential functions). Hence, we use the following strategy. We restrict ourselves to some (sufficiently large) discrete set of values. In our case, we use the following set

$$\{-5.01, -4.99, -4.97, ..., -0.03, -0.01, 0.01, 0.03, ..., 4.97, 4.99, 5.01\}$$

rather than entire set of real numbers.

In the following four tables (Tables 1–4) we summarize the results (obtained using C++ program) of the

Table 1. Analyses of properties in the set of the octane isomers

property	highest R^2 for one-parametric linear models based onbenchmark set of descriptors	highest R^2 for one-parametric linear models based on discrete Adriatic indices	highest R^2 for one-parametric linear models based on variable Adriatic indices
boling point	0.78	0.73	0.77
heat capacity at *V* constant	0.50	0.76	0.76
heat capacity at *P* constant	0.59	0.64	0.69
entropy	0.92	0.91	0.93
density	0.59	0.91	0.93
enthalpy of vaporization	0.89	0.91	0.91
standard enthalpy of vaporisation	0.92	0.97	0.97
enthalpy of formation	0.83	0.79	0.83
standard enthalpy of formation	**0.67**	**0.60**	**0.75**
motor octane number	0.93	0.96	0.97
molar refraction	0.98	0.93	0.99
acentric factor	0.99	0.99	0.99
total surface area	**0.72**	**0.78**	**0.92**
octanol-water partition coefficient	**0.29**	**0.36**	**0.99**
molar volume	0.55	0.90	0.91

Table 2. Analyses of properties in the set of the polyaromatic hydrocarbons

property	highest R^2 for one-parametric linear models based on-benchmark set of descriptors	highest R^2 for one-parametric linear models based on discrete Adriatic indices	highest R^2 for one-parametric linear models based on variable Adriatic indices
boling point	0.98	0.98	0.98
octanol-water partition coefficient	0.94	0.92	0.94

Table 3. Analyses of properties in the set of the polychlorobiphenyls

property	highest R^2 for one-parametric linear models based on-benchmark set of descriptors	highest R^2 for one-parametric linear models based on discrete Adriatic indices	highest R^2 for one-parametric linear models based on variable Adriatic indices
relative retention time	0.96	0.97	0.97
octanol-water partition coefficient	0.93	0.92	0.93
total surface area	>0.995	>0.995	>0.995
log Henry constant	0.71	0.40	0.46
log water solubility	0.94	0.94	0.94
log water activity coefficient	0.83	0.83	0.83
relative enthalpy of formation	0.67	0.55	0.62

Table 4. Analyses of properties in the set of biological activity in the phenetylamines

property	highest R^2 for one-parametric linear models based on-benchmark set of descriptors	highest R^2 for one-parametric linear models based on discrete Adriatic indices	highest R^2 for one-parametric linear models based on variable Adriatic indices
biological activity	0.54	0.57	0.58

comparison of the best correlations of the one-parameter linear models. In the second column, the highest R^2 value for one-parametric linear models based on benchmark set of descriptors is given. The highest R^2 for one-parametric linear models based on discrete Adriatic indices is given in the third column. In the last column the highest R^2 for one-parametric linear models based on variable Adriatic indices is given. Detailed tables with the names and values of these descriptors can be found in the supplementary materials.

The analyses of these four tables show that there are significant improvements only in the first table (*i.e.* when octane isomers are considered). These improvements correspond to the following three properties: standard enthalpy of formation, total surface area and octanol-water partition coefficient.

The result for octanol-water partition coefficient is especially interesting. Note that there was a very low correlation between this property and each of the indices in the benchmark set. Also, the discrete Adriatic indices made some progress, but the correlation coefficient was still very low. Contrary to this, an almost perfect correlation has been obtained for the *sum 0.37-exdeg index.*

We present these correlations in Table 5 (in the left column we present predictions by the best predictor in the benchmark set and in the right column we present predictions by the best predictor among the variable Adriatic indices; on each of the drawings R^2 is given).

From Table 5, it is obvious that the inverse sum -1.950-deg index, inverse sum 0.43-lodeg index and sum 0.37-exdeg index strongly correlate properties of molecules with their structure, and therefore, they may be a step forward in QSPR studies.

Note that each of these indices can be reformulated as:

$$\sum_{uv\in E(G)} \frac{1}{d_u^{-1.95}+d_v^{-1.95}} = \sum_{1\le i\le j\le\Delta}\left(\frac{1}{i^{-1.95}+j^{-1.95}}\right)\cdot m_{ij}(G);$$

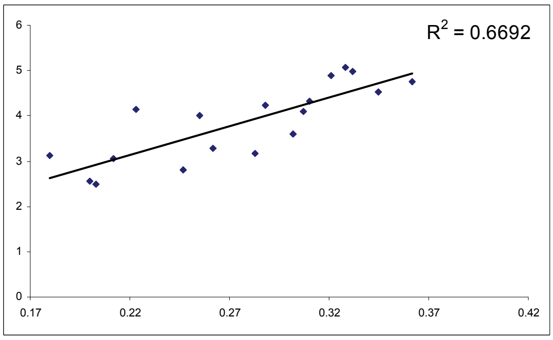
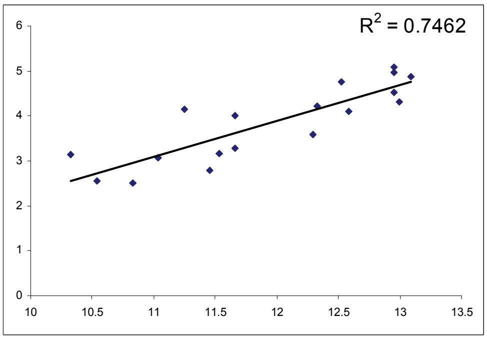
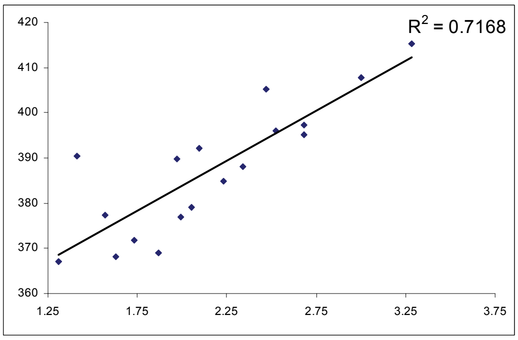
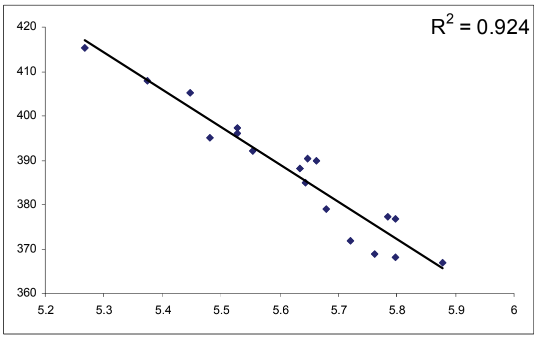
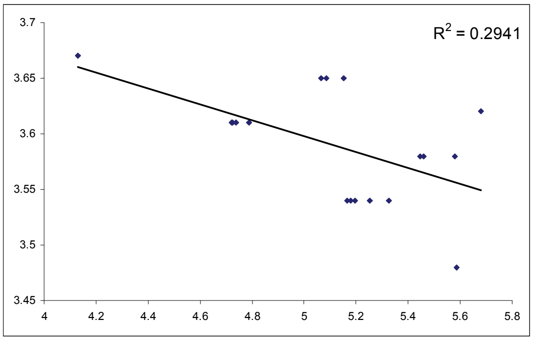
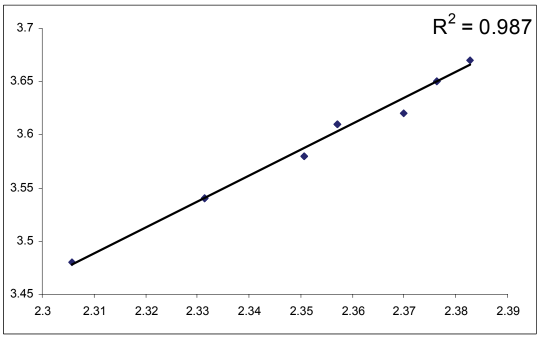
$$\sum_{uv\in E(G)} \frac{1}{\log^{0.43} d_u+\log^{0.43} d_v} = \sum_{1\le i\le j\le\Delta}\left(\frac{1}{\log^{0.43} i+\log^{0.43} j}\right)\cdot m_{ij};$$

$$\sum_{uv\in E(G)}\left(0.37^{d_u}+0.37^{d_v}\right) = \sum_{1\le i\le j\le\Delta}\left(0.37^{i}+0.37^{j}\right)\cdot m_{ij}(G)$$

where Δ is the maximal degree of graph G and $m_{ij}=m_{ij}(G)$ is the number of edges incident to vertices of degrees i and j. This is interesting, because numbers m_{ij} s have been extensively studied.[7–17] Further, the *sum 0.37-exdeg index* can be reformulated as:

$$\sum_{uv\in E(G)}\left(0.37^{d_u}+0.37^{d_v}\right) =$$
$$\sum_{u\in V(G)}\ \sum_{v\in V(G):uv\in E(G)} 0.37^{d_u} = \sum_{u\in V(G)} d_u\cdot 0.37^{d_u}.$$

Table 5. Predictions made by the best descriptor in the benchmark set and by the best predictor in the set of variable Adriatic indices for octane isomers.

standard enthalpy of formation	
 path/walk 3 - Randic shape index	 inverse sum −1.950-deg index: $\sum_{uv\in E(G)} \frac{1}{d_u^{-1.95}+d_v^{-1.95}}$
total surface area	
 second Mohar index TI2	 inverse sum 0.43-lodeg index: $\sum_{uv\in E(G)} \frac{1}{\log^{0.43} d_u+\log^{0.43} d_v}$
octanol-water partition coefficient	
 superpendentic index	 sum 0.37-exdeg index: $\sum_{uv\in E(G)} \left(0.37^{d_u}+0.37^{d_v}\right)$

Hence, this index can be observed not only as a bond additive index, but also as a vertex additive index which is much more simple. Further, if we denote by n_i the number of vertices of degree i, this index can be re-formulated as:

$$\sum_{u\in V(G)} d_u \cdot 0.37^{d_u} = \sum_{i=1}^{\Delta} i \cdot 0.37^{i} \cdot n_i .$$

In the case of chemical graphs, this reduces to:

$$\sum_{i=1}^{4} i\cdot 0.37^{i}\cdot n_i \approx 0.37\cdot n_1 + 0.274\cdot n_2 + 0.152\cdot n_3 + 0.075\cdot n_4 .$$

Hence, this is a very simple and efficient predictor for the octanol-water partition coefficient.

Supplementary Materials. – Supporting informations to the paper are enclosed to the electronic version of the article. These data can be found on the website of *Croatica Chemica Acta* (http://public.carnet.hr/ccacaa).

Acknowledgements. The partial support of Croatian Ministry of Science, Education and Sport (grants no. 177-0000000-0884 and 037-0000000-2779) is gratefully acknowledged. Sarah Michele Rajtmajer's help considering English language is also gratefully acknowledged.

REFERENCES

1. D. Vukičević and M. Gašperov, Bond Additive Modeling 1. Adriatic Indices, *Croat. Chem. Acta* **83** (3) (2010) 243–260.
2. M. Randić, *J. Am. Chem. Soc.* **97** (1975) 6609−6615.
3. I. Gutman, B. Ruščić, N. Trinajstić, and C. F. J. Wilcox Jr, *Chem. Phys.* **62** (1975) 3399−3405.
4. http://www.moleculardescriptors.eu/dataset/dataset.htm
5. http://www.iamc-online.org/
6. D. Vukičević, Bond Additive Modeling 2. Mathematical properties of Max-min rodeg indeks, *Croat. Chem. Acta* **83** (3) (2010) 261–273.
7. G. Caporossi, I. Gutman, and P. Hansen, *Comput. Chem.* **23** (1999) 469−477.
8. M. Fischermann, A. Hoffman, D. Rautenbach, and L. Volkmann, *Discrete Appl. Mat.* **128** (2003) 375−385.
9. Lj. Pavlović, *Discrete Appl. Math* **127** (2003) 615−626.
10. G. Caporossi, I. Gutman, P. Hansen, and Lj. Pavlović, *Comput. Biol. Chem.* **27** (2003) 85−90.
11. D. Vukičević and A. Graovac, *Croat. Chem. Acta* **77** (2004) 313−319.
12. D. Vukičević and A. Graovac, *Croat. Chem. Acta* **77** (2004) 481−490.
13. D. Vukičević and A. Graovac, *Croat. Chem. Acta* **77** (2004) 501−508.
14. D. Vukičević and N. Trinajstić, *Croat. Chem. Acta* **76** (2) (2003) 183−187.
15. D. Vukičević and N. Trinajstić, *MATCH Commun. Math. Comput. Chem.* **53** (2005) 111−138.
16. D. Veljan and D. Vukičević, *J. Math. Chem.* 40 (2006) 155−178.
17. D. Vukičević, *Glas. Mat. Ser. III* **44** (2) (2009) 259−266.