

## Bond Additive Modeling 1. Adriatic Indices

Damir Vukičević\* and Marija Gašperov

*Faculty of Natural Sciences and Mathematics, University of Split, Nikole Tesle 12, HR-21000 Split, Croatia*

RECEIVED FEBRUARY 25, 2009, REVISED JANUARY 7, 2010, ACCEPTED FEBRUARY 12, 2010

**Abstract.** Some of the most famous molecular descriptors are bond additive, *i.e.* they are calculated as the sum of edge contributions (Randić-type indices, Balaban-type indices, Wiener index and its modifications, Szeged index...). In this paper, the methods of calculations of bond contributions of these descriptors are analyzed. The general concepts are extracted, and based on these concepts a large class of molecular descriptors is defined. These descriptors are named Adriatic indices.

An especially interesting subclass of these descriptors consists of 148 *discrete Adriatic indices*. They are analyzed on the testing sets provided by the International Academy of Mathematical Chemistry, and it has been shown that they have good predictive properties in many cases. They can be easily encoded in the computer and it may be of interest to incorporate them in the existing software packages for chemical modeling. It is possible that they could improve various QSAR and QSPR studies.

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

### 1. INTRODUCTION

The molecular descriptor is the final result of a logical and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into a useful number or the result of some standardized experiment.<sup>1</sup> Molecular descriptors have been shown to be useful in modeling many physico-chemical properties in numerous QSAR and QSPR studies.<sup>2–4</sup>

There exist several computer programs that calculate molecular descriptors. Some of the most famous are E-Dragon (Ref. 5 and references therein), Codessa,<sup>6</sup> Molgen<sup>7</sup> and references therein. For the more comprehensive list of the software for the calculation of molecular descriptors see Ref. 8.

The aim of this paper is to produce a class of easily computable molecular descriptors that can be simply encoded in any of these software packages and that have good predictive properties. We restrict our attention to bond additive descriptors, *i.e.* descriptors that can be presented as the sum of edge contributions, since some important descriptors are defined in this way: Randić index,<sup>9</sup> Zagreb index,<sup>10</sup> Balaban Index,<sup>11</sup> modified and generalized Zagreb indices,<sup>12</sup> Wiener index<sup>13</sup> and its modification,<sup>14–17</sup> Szeged index,<sup>18</sup> PI-index,<sup>19</sup>... General concepts incorporated in these descriptors are analyzed and extracted. These concepts yield a new class of descriptors that will be defined as Adriatic descriptors.

More precisely, we shall define three classes of Adriatic descriptors:

- 1) *extended Adriatic descriptors*
- 2) *variable Adriatic descriptors*
- 3) *discrete Adriatic descriptors*

The most restrictive class of these descriptors – *discrete Adriatic descriptors* consists of 148 descriptors. We have tested these descriptors on the benchmarks sets<sup>20</sup> proposed by the International Academy of Mathematical Chemistry<sup>21</sup> and we have shown that they have very good predictive properties. *E.g.*, when octane constitutional isomers are analyzed, 9 of 15 observed properties are better modeled by one-parameter linear model by these *discrete Adriatic descriptors* than by any of 102 benchmark descriptors (for more detailed analyses of these results and for other results see Section 7).

The results seem even more impressive if we restrict ourselves to discrete Adriatic indices based on vertex degree and exclude two properties with two huge outliers. In this case 72 such descriptors outperform 102 benchmark descriptors for 7 of 13 properties. This result is even more significant when we take under consideration of the fact that *discrete Adriatic descriptors* are based on a simple easily programmable procedure, while the benchmark set consists of various types of descriptors, many of which have to be independently encoded into the computer.

\* Author to whom correspondence should be addressed. (E-mail: vukicevi@pmfst.hr)

Hence, these indices can be good and simple additions to existing indices in QSAR and QSPR oriented software packages.

The present paper is organized as follows: some well known bond additive descriptors and their modifications are analyzed in Section 2; based on these analyses *extended Adriatic indices* are defined in Section 3; we demonstrate that although their definition is quite general, it is justified (*i.e.* there are descriptors outside of this class) by showing that the Wiener index is not an *extended Adriatic index* in Section 4. In the Sections 5 and 6 *variable and discrete Adriatic indices* are defined, respectively. QSAR and QSPR analyses are presented in Section 7 (octane isomers in 7.2, polychlorobiphenyls in 7.3, polyaromatic hydrocarbons and phenethylamines in 7.4). Our exposition is concluded by conclusion, acknowledgement and references.

## 2. BOND ADDITIVE DESCRIPTORS

Many descriptors are bond-additive, *i.e.* they can be presented as a sum of edge contributions. A bond additive descriptor *Des* can be written as:

$$Des(G) = \sum_{uv \in E(G)} f(G, uv),$$

where  $E(G)$  is the set of edges and  $f$  is some mapping that assigns a real value to an ordered pair consisting of a graph and its edge. It can be seen that this is a quite general definition since  $f$  can be determined in many ways.

Some of the most famous molecular descriptors are based on defining  $f(G, uv) = \alpha(d_u, d_v)$ , where  $\alpha$  is some function from  $\mathbb{N} \times \mathbb{N}$  to  $\mathbb{R}$ ; and  $d_u$  and  $d_v$  are degrees of vertices  $u$  and  $v$  respectively.

Taking  $\alpha(d_u, d_v) = \frac{1}{\sqrt{d_u d_v}}$ , we get the Randić connectivity index<sup>9</sup> defined by:

$$\chi(G) = \sum_{uv \in E(G)} \frac{1}{\sqrt{d_u d_v}}$$

Putting  $\alpha(d_u, d_v) = d_u d_v$ , we get the Zagreb index<sup>10</sup> defined by:

$$M_2(G) = \sum_{uv \in E(G)} d_u d_v.$$

Further, this index was modified to

$${}^*M_2(G) = \sum_{uv \in E(G)} \frac{1}{d_u d_v},$$

and generalized to:<sup>12</sup>

$${}^\lambda M_2(G) = \sum_{uv \in E(G)} (d_u d_v)^\lambda,$$

which corresponds to  $\alpha(d_u, d_v) = \frac{1}{d_u d_v}$  and

$\alpha(d_u, d_v) = (d_u d_v)^\lambda$  respectively. Let us observe the first Zagreb index<sup>10</sup> defined by

$$M_1(G) = \sum_{v \in V(G)} d_v^2$$

where  $V(G)$  is the set of vertices of  $G$ . Simple calculation shows that it can also be considered as bond-additive descriptor by putting  $\alpha(d_u, d_v) = d_u + d_v$ .

The other very famous class of molecular descriptors are based on defining  $f(G, uv) = \beta(n_{uv}, n_{vu})$ , where  $\beta$  is arbitrary function from  $\mathbb{N} \times \mathbb{N}$  to  $\mathbb{R}$ ;  $n_{uv}$  is the number of vertices closer to vertex  $u$  than to vertex  $v$ ; and  $n_{vu}$  is the number of vertices closer to vertex  $v$  than to vertex  $u$ . The most famous such descriptor is the Wiener index<sup>13</sup> obtained for  $\beta(n_{uv}, n_{vu}) = n_{uv} \cdot n_{vu}$ , *i.e.*

$$W(G) = \sum_{uv \in E(G)} n_{uv} \cdot n_{vu}$$

Let us remark that Wiener index was originally defined as

$$W(G) = \sum_{u, v \in V(G)} d(u, v),$$

where  $d(u, v)$  is the distance between vertices  $u$  and  $v$ . However, these two formulas coincide for the trees.

Later, formula  $W(G) = \sum_{uv \in E(G)} n_{uv} \cdot n_{vu}$  was extended to all graphs as Szeged index.<sup>18</sup> Further this index was modified<sup>14</sup> by putting  $\beta(n_{uv}, n_{vu}) = (n_{uv} \cdot n_{vu})^{-1}$  to

$${}^m W(G) = \sum_{uv \in E(G)} (n_{uv} \cdot n_{vu})^{-1}$$

and generalized<sup>15</sup> by putting  $\beta(n_{uv}, n_{vu}) = (n_{uv} \cdot n_{vu})^\lambda$  to

$${}^m W_\lambda(G) = \sum_{uv \in E(G)} (n_{uv} \cdot n_{vu})^\lambda.$$

Further modifications that depend on number of vertices

$n(G)$  have also been studied in:<sup>16,17</sup>

$$W_{\min,\lambda}(G) = \sum_{uv \in E(G)} \left( n(G)^\lambda \cdot \min\{n_{uv}, n_{vu}\}^\lambda - \min\{n_{uv}, n_{vu}\}^{2\lambda} \right);$$

$${}_\lambda W(G) = \frac{1}{2} \sum_{uv \in E(G)} \left( n(G)^\lambda - n_{uv}^\lambda - n_{vu}^\lambda \right),$$

where  $n(G)$  is the number of vertices of graph  $G$ .

Another index belonging to this class is the PI-index<sup>19</sup> defined by  $\beta(n_{uv}, n_{vu}) = n_{uv} + n_{vu}$ , i.e.

$$W(G) = \sum_{uv \in E(G)} (n_{uv} + n_{vu}).$$

Some indices are based on a somewhat more complex definition of  $f(G, uv)$ , i.e. the Balaban distance connectivity index<sup>11</sup> is defined by:

$$J(G) = \frac{e(G)}{c(G)+1} \cdot \sum_{uv \in E(G)} (D_u D_v)^{-1/2}$$

$$(i.e. f(G, uv) = \frac{e(G)}{c(G)+1} \cdot (D_u D_v)^{-1/2})$$

where  $e(G)$  is the number of edges of graph  $G$ ,  $c(G)$  is the number of cycles of graph  $G$  and  $D_u$  is the sum of distances from  $u$  to all other vertices. Namely,  $D_u = \sum_{v \in V(G)} d(u, v)$ .

It can be seen that bond-additive descriptors can be defined in variety of ways. Moreover some of the most famous molecular descriptors have been defined in this way and their modifications have been studied.

In this paper we shall present a class of such descriptors named Adriatic indices. These class of descriptors include (generalized) Randić index,<sup>9</sup> Zagreb index,<sup>10</sup> modified and variable Zagreb index.<sup>12</sup>

### 3. EXTENDED ADRIATIC INDICES

Let us consider the variable Randić index, and the modified and the variable Zagreb indices. All these three indices are created by the following procedure:

#### Procedure I

I.1) Calculate  $d_v$  for each vertex  $v$ ;  
I.2) Define some function  $\gamma: \mathbb{R} \rightarrow \mathbb{R}$  to obtain the value  $q_v = \phi(d_v)$  for every vertex  $v$ , e.g.  $q_v = 1/\sqrt{d_v}$  when the Randić index is considered,  $q_v = 1/d_v$  when the modified Zagreb index is considered, and  $q_v = d_v^\lambda$

when the generalized Zagreb index is considered;

I.3) Calculate value  $r_{uv} = q_u \cdot q_v$  for each edge  $uv \in E(G)$ ;

I.4) Calculate the sum of all contributions  $r_{uv}$ .

Let us try to generalize this procedure to:

#### Procedure II

II.1) Calculate some invariant  $p_v$  for each vertex  $v$ ;

II.2) Define some function  $\phi: \mathbb{R} \rightarrow \mathbb{R}$  to obtain value  $q_v = \phi(p_v)$ ;

II.3) Define some symmetric function  $\gamma: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  to obtain value  $r_{uv} = \gamma(q_u, q_v)$  for each edge  $uv \in E(G)$ ;

II.4) Calculate the sum of all contributions  $r_{uv}$ .

**Remark 1.** Let us remind that function  $\gamma: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  is symmetric if  $\gamma(x, y) = \gamma(y, x)$  for every  $x, y \in \mathbb{R}$ . □

**Remark 2.** Of course, steps II.2) and II.3) could be compressed in a single step namely step II.2 is surplus (because it is always possible to put a single function  $\gamma_2(x, y) = \gamma(\phi(x), \phi(y))$ ), but we use it in order be able to separately define and combine specific functions  $\phi$  and  $\gamma$  in the remainder of the text). □

It can be shown that this definition is not restrictive enough. Namely,

**Theorem 1.** Let  $\Gamma$  be any class of graphs and let  $\omega$  be any molecular descriptor  $\omega: \Gamma \rightarrow \mathbb{R}$ . Then,  $\omega$  can be calculated using Procedure II.

**Proof:** Let  $G \in \Gamma$ . Then, take:  $p_v = \frac{\omega(G)}{n(G) \cdot d_v}$ ,

$\phi(p_v) = p_v$  and  $\gamma(r_u, r_v) = r_u + r_v$ . Alternatively, even

more simply, take:  $p_v = \frac{\omega(G)}{e(G)}$   $\phi(p_v) = p_v$  and

$\gamma(r_u, r_v) = (r_u + r_v) / 2$ . ■

Hence, in order to define a meaningful family of descriptors, we need to be more restrictive in defining a calculating algorithm. We shall restrict values of  $p_v$  to those dependent only on the following two parameters: vertex degree  $d_v$  and the sum of distances from  $v$  to all other vertices  $D_v$ .

Let us elaborate our choice. Allowing  $p_v = d_v$  is inspired by the great success of Randić-type descriptors: Randić connectivity index, Zagreb index, modified and variable Zagreb indices. Further, each index obtained by *Procedure II* in which  $p_v = d_v$  can be rewritten as:

$$\sum_{1 \leq i \leq j \leq \Delta(G)} \rho(i, j) \cdot m_{ij},$$

where  $\Delta(G)$  is maximum degree of graph  $G$ ,  $\rho$  is arbitrary function and  $m_{ij}$  is the number of edges which

end-vertices have degrees  $i$  and  $j$ . It is interesting to note that these indices depend solely on the numbers  $m_{ij}$  (and not on the graph  $G$ ). Hence, in order to find potentially optimal chemical compounds one does not need to scan the set of all graphs, but only the set of all realizable  $m_{ij}$ s which is much easier using modern computers and mixed-integer programming. Realizability of  $m_{ij}$ s has been extensively studied. Trees with minimum Randić index have been determined by Caporossi *et al.*<sup>22</sup> Papers<sup>23–25</sup> furthered these results. On the other hand, such conditions have also been investigated by Vukičević and Graovac,<sup>26–28</sup> Vukičević and Trinajstić<sup>29,30</sup> and Veljan and Vukičević<sup>31</sup> and Vukičević<sup>32</sup> to analyze discriminative properties of Randić-type indices. The following families of graphs have been considered: chemical trees, *i.e.* trees with maximum degree 4,<sup>31</sup> unicyclic chemical graphs,<sup>27</sup> general chemical graphs,<sup>30</sup> and acyclic graphs.<sup>32</sup>

Choice of the value  $D_v$  is inspired by large success of Wiener index which is in Hosoya's reformulation:<sup>33</sup>

$$W(G) = \sum_{uv \in E(G)} n_{uv} \cdot n_{vu} = \frac{1}{2} \sum_{u,v \in V(G)} d(u,v)$$

based on distances. Moreover, Wiener index can be expressed in the terms of  $n$  and  $D_v$  as:

**Theorem 2.** Let  $G$  be acyclic graph. Then,

$$W(G) = \frac{1}{4} n^2 (n-1) - \frac{1}{4} \sum_{uv \in E(G)} (D_u - D_v)^2.$$

**Proof:** Denote by  $N_{uv}$  set of vertices closer to  $u$  than to  $v$ ; and  $N_{vu}$  analogously. Note that  $d(v,w) = d(u,w) + 1$  for each vertex  $w \in N_{uv}$ , and that  $d(v,w) = d(u,w) - 1$  for each vertex  $w \in N_{vu}$ . Hence,

$$\begin{aligned} D_v - D_u &= \sum_{w \in V(G)} (d(v,w) - d(u,w)) = \\ &= \sum_{w \in N_{uv}} (d(v,w) - d(u,w)) + \sum_{w \in N_{vu}} (d(v,w) - d(u,w)) = (1) \\ &= \sum_{w \in N_{uv}} 1 + \sum_{w \in N_{vu}} (-1) = n_{vu} - n_{uv}. \end{aligned}$$

On the other hand,

$$n_{uv} + n_{vu} = n. \quad (2)$$

Solving equations (1) and (2), one gets:

$$n_{uv} = \frac{n + D_u - D_v}{2};$$

$$n_{vu} = \frac{n + D_v - D_u}{2}.$$

Therefore,

$$\begin{aligned} W(G) &= \sum_{uv \in E(G)} n_{uv} \cdot n_{vu} = \sum_{uv \in E(G)} \frac{n + D_u - D_v}{2} \cdot \frac{n + D_v - D_u}{2} = \\ &= \frac{1}{4} \sum_{uv \in E(G)} [n^2 - (D_u - D_v)^2] = \\ &= \frac{1}{4} n^2 (n-1) - \frac{1}{4} \sum_{uv \in E(G)} (D_u - D_v)^2, \end{aligned}$$

which proves the Theorem. ■

Hence, if one observes the graphs with same number of vertices, the molecular descriptor:

$$W'(G) = \sum_{uv \in E(G)} (D_u - D_v)^2$$

has exactly the same predictive properties as the Wiener index in all linear and multi-linear models. An example of such set of graphs is the set of alkanes with 8 vertices which was used in one of the benchmark data sets<sup>20</sup> proposed by the International Academy of Mathematical Chemistry.<sup>21</sup>

We have explained argumentation for using  $d_v$  and  $D_v$ ; and we propose the following procedure:

### Procedure III. Extended Adriatic Indices

- III.1) Choose invariant  $p_v$  to be  $d_v$  or  $D_v$ ;
- III.2) Define some function  $\phi: \mathbb{R} \rightarrow \mathbb{R}$  to obtain value  $q_v = \phi(p_v)$ ;
- III.3) Define some function  $\gamma: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  to obtain the value  $r_{uv} = \gamma(q_u, q_v)$  for each edge  $uv \in E(G)$ ;
- III.4) Calculate the sum of all contributions  $r_{uv}$ .

**Remark 3.** As the name suggests in the remainder of the paper, we are going to further restrict this procedure. Also, note that  $\phi$  and  $\gamma$  are functions that do not depend on any particular graph  $G$ . □

## 4. THE WIENER INDEX IS NOT AN EXTENDED ADRIATIC INDEX

In order to prove that concept of extended Adriatic indices is meaningful we need to prove that there is at least one index that is not an extended Adriatic index. This will be proved for the Wiener index:

**Theorem 3.** Wiener index is not extended Adriatic index.

**Proof:** Suppose to the contrary that Wiener index is extended Adriatic index  $A$ . Distinguish two cases:



Figure 1. Graphs  $G_1$  and  $G_2$ .

CASE 1:  $A$  is based on  $d_v$ .

Let  $G_1$  and  $G_2$  be graphs presented on the Figure 1.

Note that:

$$A(G_1) = 2 \cdot \gamma(\phi(1), \phi(2)) + 2 \cdot \gamma(\phi(2), \phi(2)) + 3 \cdot \gamma(\phi(2), \phi(3));$$

$$A(G_2) = 2 \cdot \gamma(\phi(1), \phi(2)) + 2 \cdot \gamma(\phi(2), \phi(2)) + 3 \cdot \gamma(\phi(2), \phi(3));$$

$$W(G_1) = 76;$$

$$W(G_2) = 75.$$

Hence,  $A(G_1) = A(G_2)$  and  $W(G_1) \neq W(G_2)$ , but this is in contradiction with  $A = W$ .

CASE 2:  $A$  is based on  $D_v$ .

It seems as quite a difficult problem to find two non-isomorphic graphs for which all extended Adriatic indices have the same value (we propose this as an open problem). Hence, here we use a different approach. Denote by  $D_{i,j}(G)$  the number of edges  $e = uv$  such that  $\{D_u, D_v\} = \{i, j\}$  for every  $i, j \in \mathbb{N}$  and every graph  $G$ . Let us illustrate this definition by the following example on Figure 2.

Then,

$$D_{11,16}(G_5) = D_{11,16}(G_5) = 4;$$

$$D_{11,10}(G_5) = D_{10,11}(G_5) = 2;$$

$$D_{i,j}(G_5) = D_{j,i}(G_5) = 0 \text{ for every } \{i, j\} \neq \{10, 11\}, \{11, 16\}.$$

Note that for each graph one has:

$$A(G) = \sum_{1 \leq k \leq l \leq \Delta(G)} \gamma(\phi(k), \phi(l)) \cdot D_{k,l}(G).$$

Let  $k$  be any integer and let  $H_1, \dots, H_K$  be any family

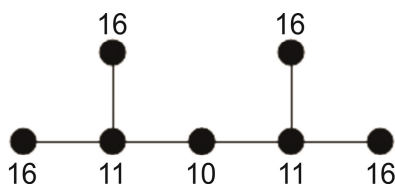


Figure 2. Graph  $G_5$  with values  $D_i(v)$ .

of graphs, and let  $q = \max\{\bar{\Delta}(H_1), \dots, \bar{\Delta}(H_k)\}$ , where  $\bar{\Delta}(H_i)$  is the diameter of graph  $H_i$ . Then,

$$A(H_i) = \sum_{1 \leq k \leq l \leq q} \gamma(\phi(k), \phi(l)) \cdot D_{k,l}(H_i). \quad (3)$$

Note that there are  $q \cdot (q+1) / 2$  pairs of numbers  $(k, l)$  such that  $1 \leq k \leq l \leq q$ . Let  $\tau$  be any bijection  $\tau: \{1, \dots, q \cdot (q+1) / 2\} \rightarrow \{(k, l) \in \mathbb{N}^2 : 1 \leq k \leq l \leq q\}$ .

Denote  $\tau(j) = (\tau_1(j), \tau_2(j))$ .

Denote by  $\Gamma$  column matrix defined by:

$$\Gamma_j = \gamma(\phi(\tau_1(j)), \phi(\tau_2(j))).$$

Let  $\Omega$  be the matrix defined by

$$\Omega_{ij} = D_{\tau_1(j), \tau_2(j)}(H_i)$$

and denote by  $\Omega_i$   $i$ -th row of  $\Omega$ . Denote by  $\Lambda$  column matrix defined by

$$\Lambda_i = A(H_i).$$

Equality (3) can be rewritten as  $\Lambda_i = A(H_i) = \Omega_i \cdot \Gamma$ . Hence,

$$\Lambda = \Omega \cdot \Gamma.$$

Let  $[\Omega \Lambda]$  be a matrix obtained by adding column  $\Lambda$  to matrix  $\Omega$  as the last column. Equality (4) implies that matrix  $\Lambda$  is a linear combination of columns of  $\Omega$ , hence matrices  $\Omega$  and  $[\Omega \Lambda]$  must have the same rank. Note that matrix  $\Omega$  depends solely on the family of graphs and that  $\Lambda$  (Wiener indices of these graphs) also depends solely on these graphs. Hence, in order to obtain a contradiction we do not need to analyze functions  $\gamma$  and  $\phi$ , but just to find a family of graphs for which

$$\text{rank } \Omega < \text{rank } [\Omega \Lambda]$$

Further note that rank does not change if we eliminate columns that consist only of zeros. Let  $\Omega'$  be a matrix obtained by elimination of all zero-columns. It is sufficient to show that:

$$\text{rank } \Omega' < \text{rank } [\Omega' \Lambda]$$

Using the computer we have analyzed the family of graphs consisting of non-isomorphic trees with 8, 9, 10, 11 and 12 vertices. There are 962 such trees. We have constructed the required matrices and obtained:

$$\begin{aligned} \text{rank } \Omega' &= 298; \\ \text{rank } [\Omega' \Lambda] &= 299. \end{aligned}$$

This proves the Theorem. ■

## 5. VARIABLE ADRIATIC INDICES

Now, let us further restrict the procedure for generating extended Zagreb indices. We specify types of functions that may be used as  $\phi$  and  $\gamma$ . Function  $\phi$  should be considered as the extent of the influence of the value  $p_v$ . For instance for the Zagreb index value  $\phi(d_v)$  increases with the increment of  $d_v$ , but for the modified Zagreb index and for the Randić index value of  $\phi(d_v)$  decreases with increment of  $d_v$ . Moreover, the same increment gives different levels of decrease in these two indices.

In mathematics we usually speak about three type of dependencies: logarithmic, polynomial and exponential. Hence, it would be reasonable to consider functions:  $\phi(x) = \log(x)$ ,  $\phi_{2,a}(x) = x^a$  and  $\phi_{3,a}(x) = a^x$ . Note that latter two dependences include variable value  $a$ . On the other hand, the first one does not. Hence, we generalize it by putting  $\phi_{1,a}(x) = \log^a(x)$ . More precisely we use the following functions:

- $\phi_{1,a}(x) = \log^a(x)$ ,  $a > 0$ ;
- $\phi_{2,a}(x) = x^a$ ,  $a \in \mathbb{R} \setminus \{0\}$
- $\phi_{3,a}(x) = a^x$ ,  $a > 0$ .

Let us note that logarithmic and exponential functions in the calculation of bond contributions are not new in the definition of molecular descriptors. For instance, mean information content on the edge magnitude<sup>34</sup> is defined by:

$${}^E \bar{I}_X^M(G) = - \sum_{uv \in E(G)} \frac{(d_u d_v)^{-1/2}}{\chi(G)} \cdot \log_2 \frac{(d_u d_v)^{-1/2}}{\chi(G)},$$

and entries of the augmented vertex-degree matrix<sup>35</sup> are given by:

$$[{}^a \Delta]_{ij} = \begin{cases} d_j / 2^{d(i,j)}, & i \neq j; \\ d_i, & i = j. \end{cases}$$

Now, let us analyze function  $\gamma$ . It should be some relatively simple symmetric function. Two immediate candidates are:

- $\gamma_1(x, y) = x \cdot y$  which is used in all Randić-type indices;
- $\gamma_2(x, y) = x + y$  which can be used to calculate the first Zagreb index.

Note that variability would not be significantly increased if we would use  $\gamma(x, y) = (xy)^z$ , because the exponent is already included in  $\phi_{1,a}$  and  $\phi_{3,a}$ . However, inspired by the modification of the Zagreb index given by the modified Zagreb index, we make analogous modification of the function  $\gamma_2$  and write  $\gamma(x, y) = \frac{1}{x+y}$ .

However, to insure that there is no 0 in the denominator, we put:

$$\bullet \quad \gamma_3(x, y) = \begin{cases} \frac{1}{x+y}, & x+y \neq 0; \\ 0, & x+y = 0. \end{cases}$$

Of course one would like to extend definitions of  $\gamma_1$  and  $\gamma_2$  to the remaining two elementary operations: difference and division. However, this is not possible to be done directly since these two operations are not symmetric. Therefore, in order to use difference, we shall apply the absolute value and put:

$$\bullet \quad \gamma_4(x, y) = |x - y|$$

Similarly as in the case of the sum, we propose the following modification

$$\bullet \quad \gamma_5(x, y) = \begin{cases} \frac{1}{|x-y|}, & x \neq y; \\ 0, & x = y. \end{cases}$$

In the case of division, we propose three possible way to symmetrize it:

$$\bullet \quad \gamma_6(x, y) = \begin{cases} \frac{\min\{x, y\}}{\max\{x, y\}}, & \max\{x, y\} \neq 0; \\ 0, & \max\{x, y\} = 0; \end{cases}$$

$$\bullet \quad \gamma_7(x, y) = \begin{cases} \frac{\max\{x, y\}}{\min\{x, y\}}, & \min\{x, y\} \neq 0; \\ 0, & \min\{x, y\} = 0; \end{cases}$$

$$\bullet \quad \gamma_8(x, y) = \begin{cases} \frac{x}{y} + \frac{y}{x}, & x, y \neq 0; \\ 0, & \text{otherwise.} \end{cases}$$

The operation applied in  $\gamma_6$  is inspired by distance/detour quotient matrix<sup>36</sup> in which (a smaller) distance is divided by (a larger) detour distance and the operation applied in  $\gamma_8$  is inspired by the extended adjacency matrix<sup>37</sup> which is defined by:

$$[E^v A]_{ij} = \begin{cases} \frac{d_i / d_j + d_j / d_i}{2}, & ij \text{ adjacent;} \\ 0, & \text{otherwise.} \end{cases}$$

Let us note that not only the sum of divisions, but also divisions of sums have been studied in chemistry.<sup>38-41</sup>

Now, let us present the procedure for calculation of variable Adriatic Indices:

#### Procedure IV. Variable Adriatic Indices

IV.1) Choose invariant  $p_v$  to be  $d_v$  or  $D_v$ ;  
 IV.2) Choose numbers  $i \in \{1, 2, 3\}$  and  $j \in \{1, 2, \dots, 8\}$ ;  
 IV.3) Choose  $a \in \mathbb{R} \setminus \{0\}$  if  $i = 2$  and  $j \in \{1, 2, \dots, 5\}$ ; otherwise if  $i = 1$  or  $2$  choose  $a \in \mathbb{R}^+$  and if  $i = 3$  choose  $a \in \mathbb{R}^+ \setminus \{1\}$ .

IV.4) Calculate  $A(G) = \sum_{uv \in E(G)} \gamma_i(\phi_{j,a}(p_u), \phi_{j,a}(p_v))$ .

**Remark 4.** Let us comment on the restriction of choices of  $a$ . Obviously,  $p_v$  regardless of the choice  $d_v$  and  $D_v$  is a positive integer. Therefore,  $p_v^a$  is well defined for all  $a \in \mathbb{R}$ . However,  $a = 0$  maps all  $p_v$  to 1, and this is not of interest. Let us observe  $\log^a p_v$ . Note that  $\log p_v$  can be zero. Hence, obviously,  $a > 0$ . Now, let us observe  $a^{p_v}$ . If  $a$  is a negative number, this function is still well defined, because  $p_v$  is positive integer. However, it is not a monotonous function in this case. Therefore, we assume that it is not the most suitable choice for the majority of QSPR studies. Further,  $a = 1$  maps all  $p_v$  to 1, and this is not of interest. It remains to explain why we do not observe negative values of  $a$  for descriptors:

$$\gamma_6(p_u^a, p_v^a) = \begin{cases} \frac{\min\{p_u^a, p_v^a\}}{\max\{p_u^a, p_v^a\}}, & \max\{p_u^a, p_v^a\} \neq 0 \\ 0, & \max\{p_u^a, p_v^a\} = 0 \end{cases} =$$

$$\left\{ \begin{array}{l} \text{because} \\ p_u^a, p_v^a > 0 \end{array} \right\} = \frac{\min\{p_u^a, p_v^a\}}{\max\{p_u^a, p_v^a\}};$$

$$\gamma_7(p_u^a, p_v^a) = \begin{cases} \frac{\max\{p_u^a, p_v^a\}}{\min\{p_u^a, p_v^a\}}, & \min\{p_u^a, p_v^a\} \neq 0 \\ 0, & \min\{p_u^a, p_v^a\} = 0 \end{cases} =$$

$$\left\{ \begin{array}{l} \text{because} \\ p_u^a, p_v^a > 0 \end{array} \right\} = \frac{\max\{p_u^a, p_v^a\}}{\min\{p_u^a, p_v^a\}};$$

$$\gamma_8(p_u^a, p_v^a) = \begin{cases} \frac{p_u^a}{p_v^a} + \frac{p_v^a}{p_u^a}, & p_u^a, p_v^a \neq 0; \\ 0, & \text{otherwise.} \end{cases} =$$

$$\left\{ \begin{array}{l} \text{because} \\ p_u^a, p_v^a > 0 \end{array} \right\} = \frac{p_u^a}{p_v^a} + \frac{p_v^a}{p_u^a}.$$

Namely, in all three cases, we obtain

$$\gamma_i(p_u^a, p_v^a) = \gamma_i(p_u^{-a}, p_v^{-a}), i = 6, 7, 8;$$

Hence only positive values cover all possibilities. Let us prove this. Assume that  $a < 0$  and without loss of generality assume that  $p_u > p_v$ . It holds:

$$\gamma_6(p_u^a, p_v^a) = \frac{\min\{p_u^a, p_v^a\}}{\max\{p_u^a, p_v^a\}} = \frac{p_u^a}{p_v^a} = \frac{p_v^{-a}}{p_u^{-a}} =$$

$$\frac{\min\{p_u^{-a}, p_v^{-a}\}}{\max\{p_u^{-a}, p_v^{-a}\}} = \gamma_6(p_u^{-a}, p_v^{-a});$$

$$\gamma_7(p_u^a, p_v^a) = \frac{\max\{p_u^a, p_v^a\}}{\min\{p_u^a, p_v^a\}} = \frac{p_v^a}{p_u^a} = \frac{p_u^{-a}}{p_v^{-a}} =$$

$$\frac{\max\{p_u^{-a}, p_v^{-a}\}}{\min\{p_u^{-a}, p_v^{-a}\}} = \gamma_7(p_u^{-a}, p_v^{-a});$$

$$\gamma_8(p_u^a, p_v^a) = \frac{p_u^a}{p_v^a} + \frac{p_v^a}{p_u^a} = \frac{p_v^{-a}}{p_u^{-a}} + \frac{p_u^{-a}}{p_v^{-a}} = \frac{p_u^{-a}}{p_v^{-a}} + \frac{p_v^{-a}}{p_u^{-a}} =$$

$$\gamma_8(p_u^{-a}, p_v^{-a});$$

which proves the claim.  $\square$

Now, let us present the nomenclature of variable Adriatic indices:

- 1) The name of each variable Adriatic index consists of five parts. The third and the fourth part form the same word.
- 2) The first part is determined by the choice of the function  $\gamma_j$ :
  - $\gamma_1$  corresponds to *Randić type*;
  - $\gamma_2$  corresponds to *sum*;
  - $\gamma_3$  corresponds to *inverse sum*;
  - $\gamma_4$  corresponds to *misbalance*;
  - $\gamma_5$  corresponds to *inverse misbalance*;
  - $\gamma_6$  corresponds to *min-max*;
  - $\gamma_7$  corresponds to *max-min*;
  - $\gamma_8$  corresponds to *symmetric division*.
- 3) The second part consists of the number  $a$  and a line, e.g. "1/3–"
- 4) The third part is determined by the choice of the function  $\phi_{i,a}$ :
  - $\phi_{1,a}$  corresponds to *lo*;
  - $\phi_{2,a}$  corresponds to no letters (the whole word is just the fourth part);
  - $\phi_{3,a}$  corresponds to *ex*.
- 5) The fourth part is determined by the choice of the invariant ( $d_v$  or  $D_v$ ):
  - $d_v$  corresponds to *deg*;
  - $D_v$  corresponds to *di*;
- 6) The fifth part is word *index*.

Let us illustrate this by several examples:

- $\sum_{uv \in E(G)} \left| \sqrt{\log d_u} - \sqrt{\log d_v} \right|$  corresponds to *misbalance 1/2-lodeg index*;
- $\sum_{uv \in E(G)} \frac{1}{D_u^{3/2} + D_v^{3/2}}$  corresponds to *inverse sum 3/2-di index*;
- $\sum_{uv \in E(G)} \frac{1}{\left| 3^{D_u} - 3^{D_v} \right|}$  corresponds to *inverse misbalance 3-exdi index*;
- $\sum_{uv \in E(G)} \left( \log^{1/4} d_u + \log^{1/4} d_v \right)$  corresponds to *sum 1/4-lodeg index*.

## 6. DISCRETE ADRIATIC INDICES

Note that in the linear model for predicting some property  $P$  by topological index  $TI$ , we optimize two variables  $P \sim k \cdot TI + l$ . On the other hand, if we search for the optimal variable index  $TI_\lambda$  for prediction of the same property using linear model, we optimize three variables:  $k$ ,  $l$  and  $\lambda$ .

Since, in benchmark sets proposed by International Academy of Mathematical Chemistry<sup>21</sup> descriptors are not variable; it would not be a fair comparison to compare variable descriptors with nonvariable ones. Therefore, we should propose discrete Adriatic indices

in which value  $a$  is not a continuous variable, but consists of several (discrete values).

Here, we choose the following values:

- $a = -1$  (inspired by modified Zagreb index)
- $a = -1/2$  (inspired by Randić connectivity index)
- $a = 1/2$  (inverse value of  $1/2$ )
- $a = 1$  (inspired by the second Zagreb index)
- $a = 2$  (inspired by the first Zagreb index)

Now, we can summarize the procedure for the calculation of discrete Zagreb indices:

### Procedure V. Discrete Adriatic Indices

- V.1) Choose invariant  $p_v$  to be  $d_v$  or  $D_v$ ;
- V.2) Choose numbers  $i \in \{1, 2, 3\}$  and  $j \in \{1, 2, \dots, 8\}$ ;
- V.3) Choose  $a \in \{-1, -1/2, 1/2, 1, 2\}$  if  $i = 2$  and  $j \in \{1, 2, \dots, 5\}$ ; otherwise if  $i = 1$  or  $2$  choose  $a \in \{1/2, 1, 2\}$  and if  $i = 3$  choose  $a \in \{1/2, 2\}$
- V.4) Calculate  $A(G) = \sum_{uv \in E(G)} \gamma_j(\phi_{i,a}(p_u), \phi_{i,a}(p_v))$ .

In this way  $2 \cdot (5 \cdot 5 + 11 \cdot 3 + 8 \cdot 2) = 148$  discrete Adriatic indices have been defined.

Now, let us present the nomenclature of discrete Adriatic indices:

- 1) The name of each Discrete Adriatic index consists of four parts. The second and the third part from the same word.
- 2) The first part is determined by the choice of the function  $\gamma_j$  completely analogously as in the case of variable Adriatic indices
- 3) The second part is determined by the value of  $\phi_{i,a}$ :
  - $\phi_{1,1/2}$  corresponds to *lor*;
  - $\phi_{1,1}$  corresponds to *lo*;
  - $\phi_{1,2}$  corresponds to *los*;
  - $\phi_{2,-1}$  corresponds to *in*;
  - $\phi_{2,-1/2}$  corresponds to *ir*;
  - $\phi_{2,1/2}$  corresponds to *ro*;
  - $\phi_{2,1}$  corresponds to no letters (the whole word is just the third part);
  - $\phi_{2,2}$  corresponds to *s*;
  - $\phi_{3,1/2}$  corresponds to *ha*;
  - $\phi_{3,2}$  corresponds to *two*;
- 4) The third part is determined by the choice of the invariant ( $d_v$  or  $D_v$ ) completely analogously as in the case of variable Adriatic indices;
- 5) The fourth part is word *index*.

Let us illustrate this by several examples:

- $\sum_{uv \in E(G)} \left| \sqrt{\log d_u} - \sqrt{\log d_v} \right|$  corresponds to *misbalanced lordeg index* (recall that the same index was named *misbalance 1/2-lodeg index*)



when it was observed as variable Adriatic index)

- $\sum_{uv \in E(G)} \frac{1}{D_u^{-1/2} + D_v^{-1/2}}$  corresponds to *inverse sum irdi index*;
- $\sum_{uv \in E(G)} \frac{1}{|2^{D_u} - 2^{D_v}|}$  corresponds to *inverse misbalance twodi index*;
- $\sum_{uv \in E(G)} (\log^{1/2} d_u + \log^{1/2} d_v)$  corresponds to *sum lordeg index*.

## 7. QSAR AND QSPR STUDIES OF DISCRETE ADRIATIC INDICES

In Ref. 20 it has been written: "In order to promote the comparison of new and old molecular descriptors, evaluate their predictive ability and better understand their meaning, the International Academy of Mathematical Chemistry<sup>21</sup> suggests the use of some benchmark data sets". We shall prove that discrete Adriatic indices demonstrate good predictive properties on these benchmark sets.

Namely, four sets of chemical compounds are proposed in Ref. 20:

- 1) set of 18 octane isomers
- 2) set of 82 polyaromatic hydrocarbons (PAH)
- 3) set of 209 polychlorobiphenyls (PCB)
- 4) set of 22 phenethylamines (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 phenethylamines.

We compare the best correlations of one-parameter linear models based on descriptors in these benchmark sets with correlations of one-parameter linear models based on discrete Adriatic indices. We exclude from our observations melting points, since graph-theoretical abstraction disregards many information that are relevant for the value of the melting point.

Therefore, in total we compare descriptive properties in 25 different cases (15 properties of octanes, 2 properties of PAHs, 7 properties of PCBs and 1 property of phenethylamines). It will be shown that Adriatic indices give better results in 13 out of these 25 cases.

### 7.1. BEST PREDICTORS (IN LINEAR MODELS) OF (CONSTITUTIONAL) OCTANE ISOMERS

We analyze 15 properties of octane isomers and for 9 properties we get better results than descriptors given in

the benchmark set. In Table 1 are our results (in the left column are predictions by the best predictor in benchmark set and in the right column are predictions by the best predictor among discrete Adriatic indices; on each of the drawing  $R^2$  is given).

Although we have obtained better correlation in these nine cases, we can see that all nine results are not equally significant. Namely, in two cases (density and molar volume) there is a single molecule (2,2,3,3-tetramethylbutane) that is significantly different than all others. This single data highly influences correlation coefficients and makes all conclusions unreliable.

### 7.2. BEST PREDICTORS (IN LINEAR MODELS) OF POLYCHLOROBIPHENYLS

We analyze 7 properties of PCBs and for 3 properties we get better results than descriptors given in the benchmark set. In Table 2 we present our results (the results are organized as in Table 1).

However, we can see that there are improvements, but they are rather small. Here we speak of comparably good descriptors. There is always a possibility that part of the improvement is the result of rounding or other imprecision in computer calculation.

### 7.3. BEST PREDICTORS (IN LINEAR MODELS) OF POLYAROMATIC HYDROCARBONS AND PHENETHYLAMINES

We have analyzed two properties of polyaromatic hydrocarbons: boiling point and octanol-water partition coefficient. In both cases all discrete Adriatic descriptors performed worse than the best descriptor in benchmark set.

Only one property (biological activity) has been analyzed for phenethylamines and we have found that there is a discrete Adriatic index that outperforms all descriptors in the benchmark set. The results are given by Table 3.

### 7.4. ADDITIONAL RESULTS

In previous sections, we have shown that in 13 cases, there is a discrete Adriatic index that provides better predictive properties than any of the descriptors in the benchmark sets. From these 13 cases, we eliminate observations of density and molar volume of octane isomers, because in these cases correlations are influenced by a single outlier that corrupts the distribution of data. In the remaining 11 cases, it may be of relevance not to analyze only discrete Adriatic index with the best predictive properties, but maybe several the best ones. Here we propose the following criteria: we assume that

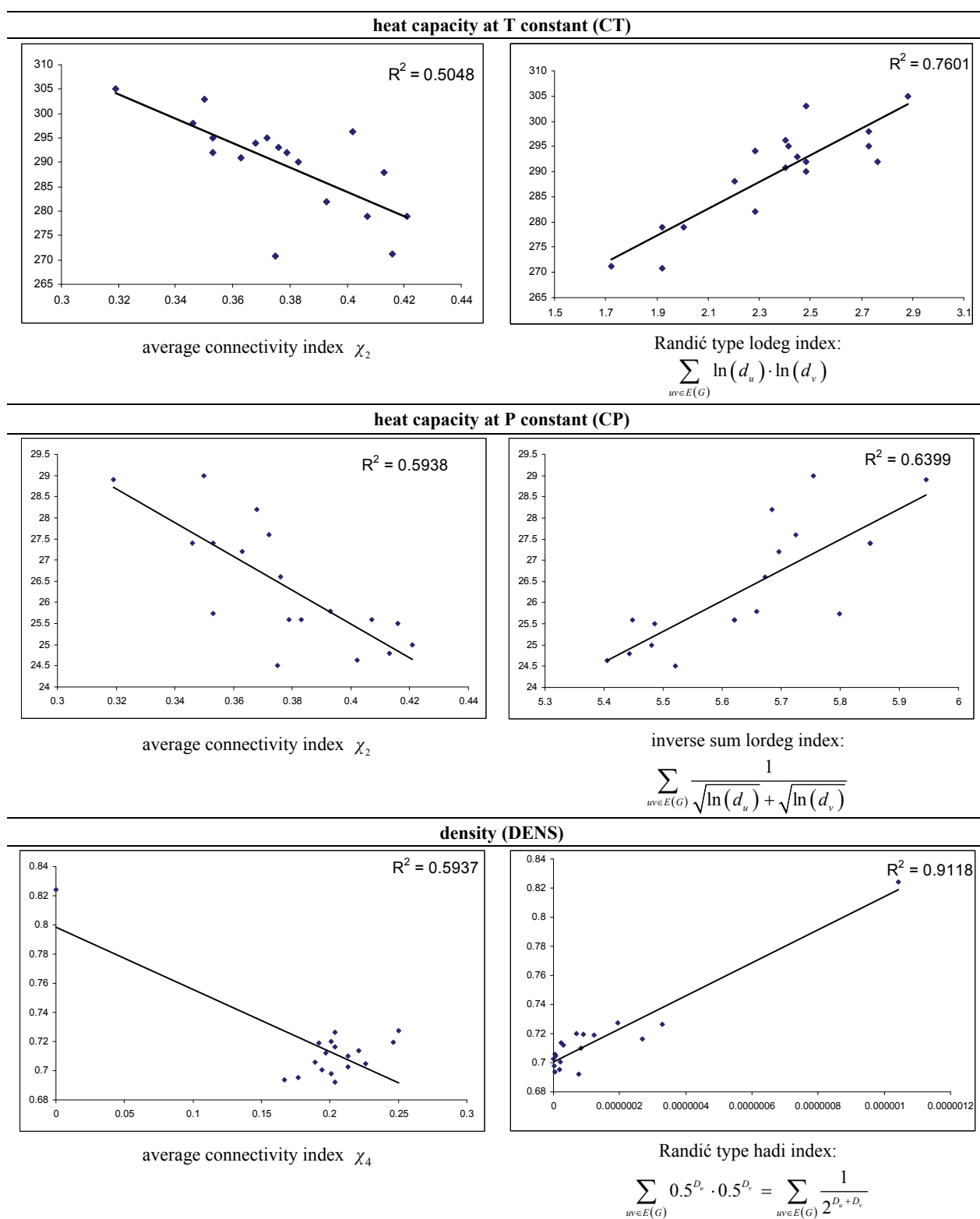
**Table 1.** Nine properties that can be predicted better (using linear fits) by one of the discrete Adriatic indices than any of descriptors in benchmark set ( $x$ -axis represents the value of the descriptor and  $y$ -axis represents the value of the property)

Table 1. (continued)

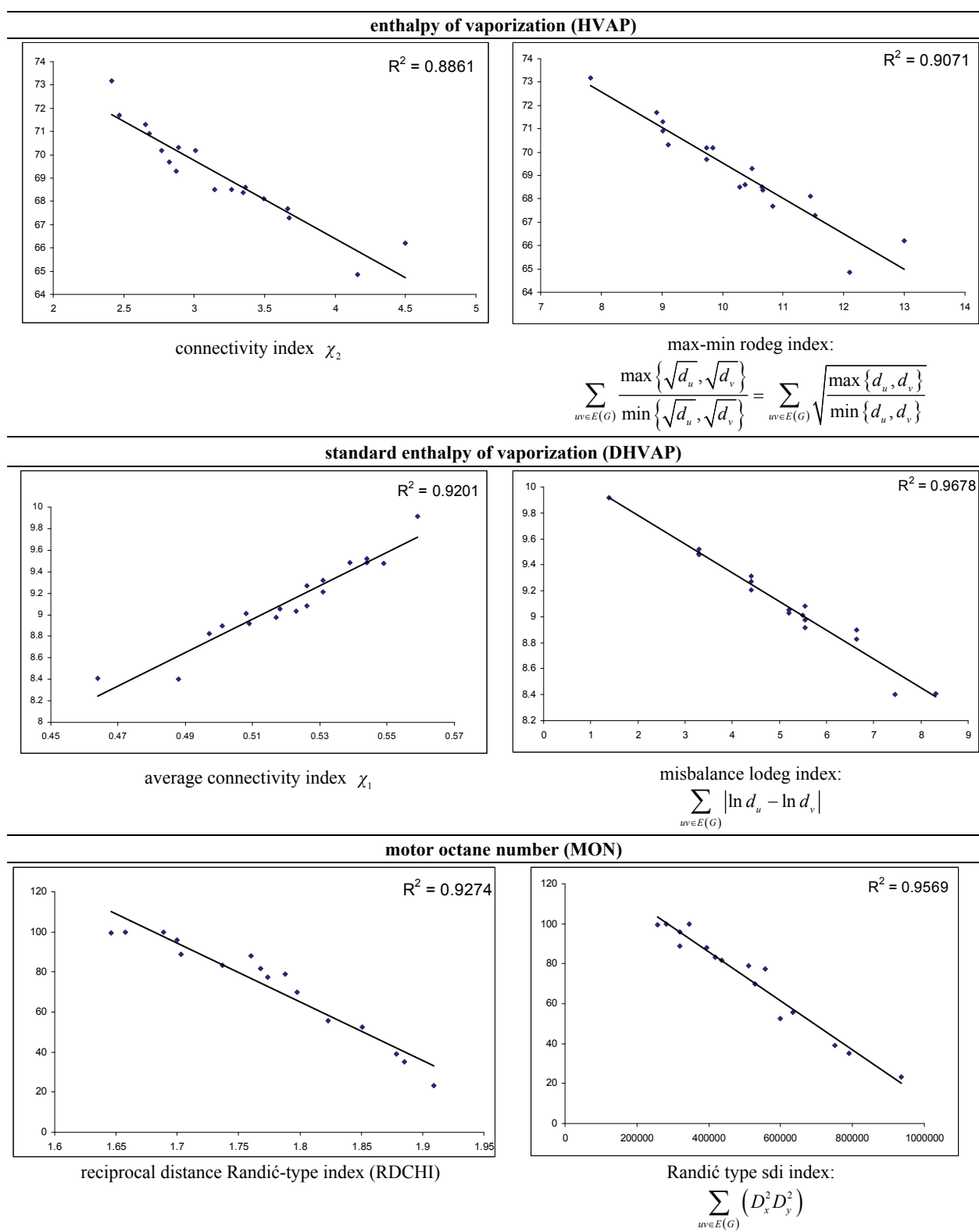
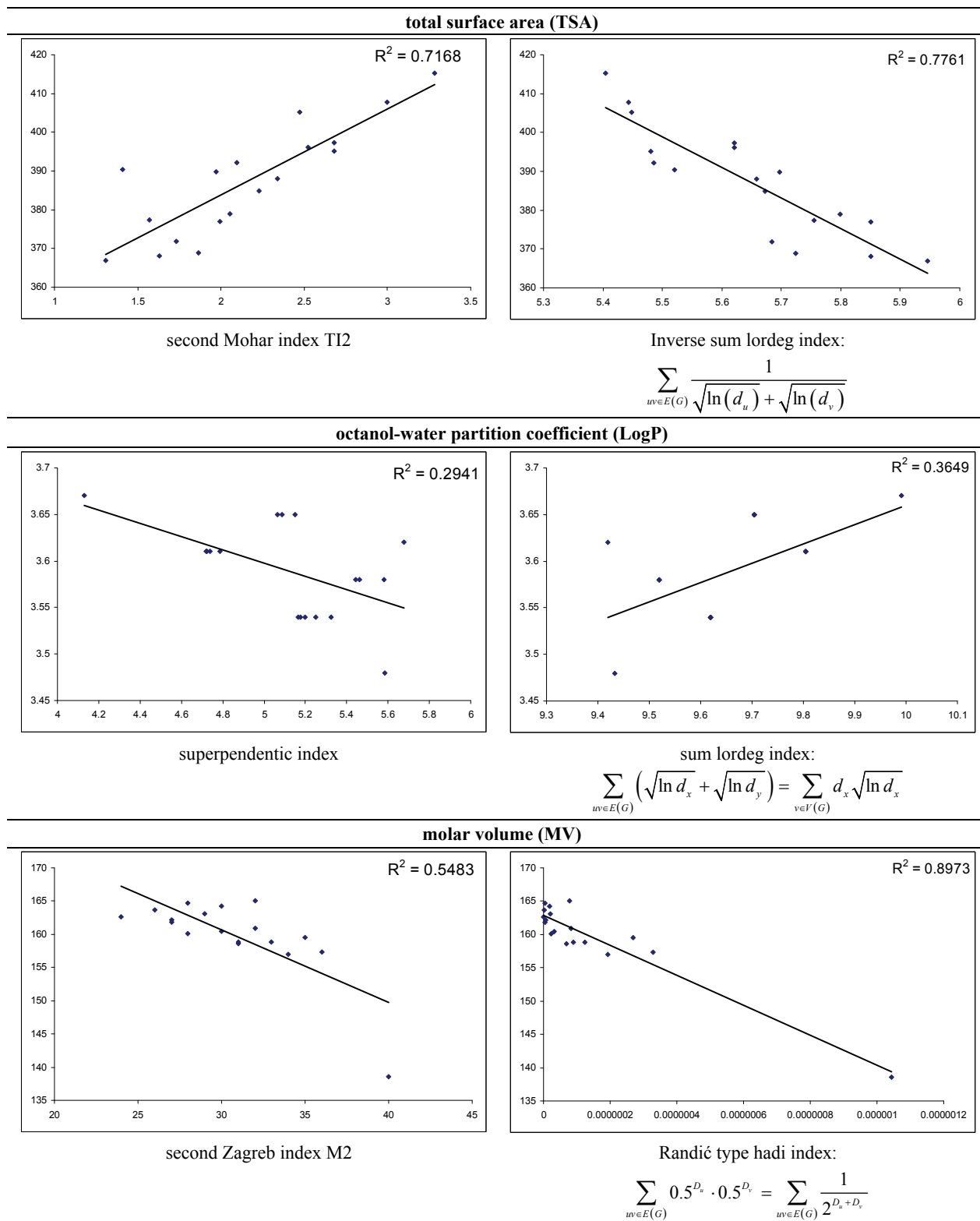
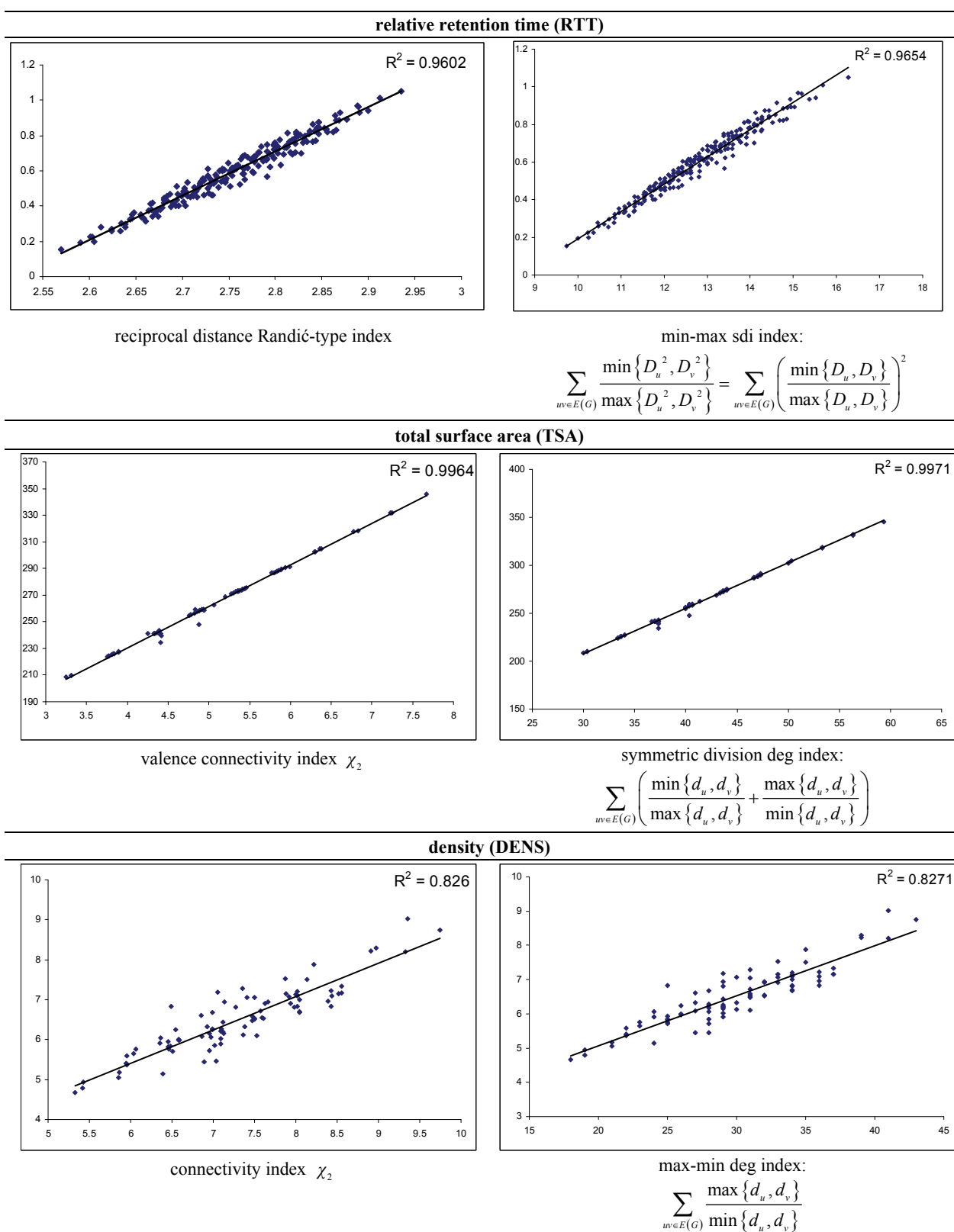
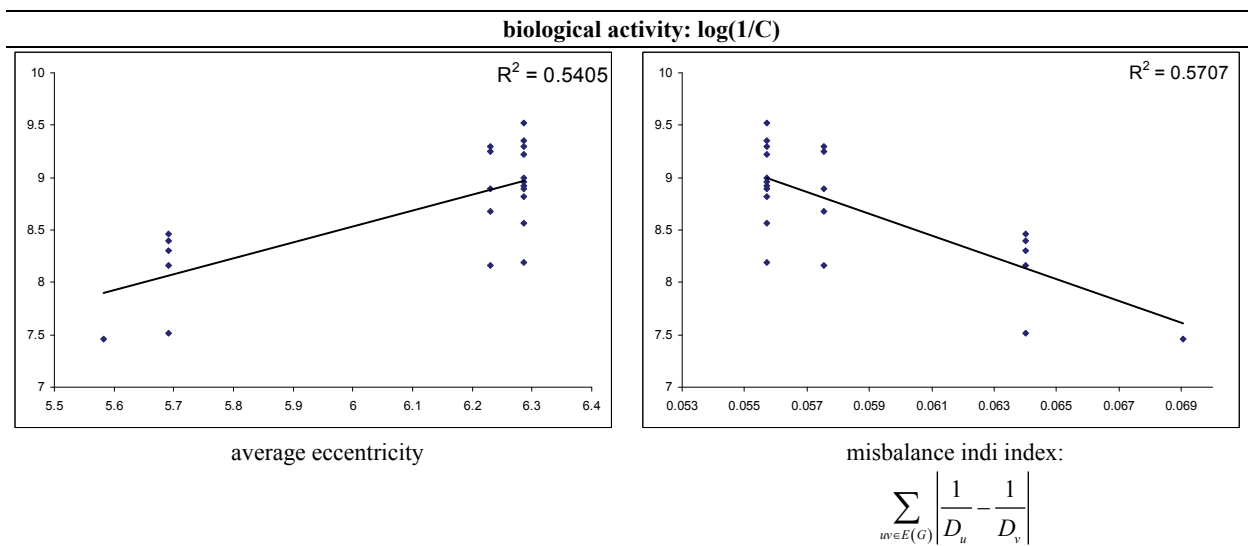


Table 1. (continued)



**Table 2.** Three properties that can be predicted better (using linear fits) by one of the discrete Adriatic indices than any of descriptors in benchmark set ( $x$ -axis represents the value of the descriptor and  $y$ -axis represents the value of the property)

**Table 3.** Predictions of biological activity by the most suitable discrete Adriatic index and the most suitable descriptor in the benchmark set ( $x$ -axis represents the value of the descriptor and  $y$ -axis represents the value of the property)**Table 4.** Analyses of discriminative properties of Adriatic descriptors.

set of molecules	property	$R_b$	number of discr. Adr. des. $x$	number of discr. Adr. des. $x$	$R_{ad}$
			such that $R_b < R_x^2 \leq \frac{R_{da}^2 + R_b^2}{2}$	such that $R_x^2 > \frac{R_{da}^2 + R_b^2}{2}$	
octane isomers	CT	0.505	0	1	0.760
	CP	0.594	0	1	0.640
	HVAP	0.886	8	<b>5</b>	0.907
	DHVAP	0.920	4	<b>10</b>	0.968
	MON	0.927	6	1	0.957
	TSA	0.717	0	<b>2</b>	0.776
	LogP	0.294	0	1	0.365
PCB	RTT	0.960	0	1	0.965
	TSA	0.996	1	1	0.997
Phenet	logYw	0.826	2	<b>3</b>	0.827
	log(1/c)	0.541	6	1	0.571

Adriatic index is a significant predictor if it improves  $R^2$  by at least half of the improvement of the best discrete Adriatic index. More precisely, let  $R_b^2$  be the correlation of the best predictor in the benchmark set, let  $R_{da}^2$  be the correlation of the best predictor among discrete Adriatic indices (in the observed cases, it is always  $R_{da}^2 > R_b^2$ ). We assume that discrete Adriatic index  $x$  is significant if  $R_x^2 \geq \frac{R_{da}^2 + R_b^2}{2}$ . In Table 4 we give the statistics of  $R^2$  values for the observed 11 properties.

Hence, in 4 cases we have found significant predictors among Adriatic descriptors different from the best predictor (bold numbers in Table 4). In three case

(HVAP, DHVAP and TSA), we can see that there is a significant improvement in  $R^2$ . On the other hand, when we consider logYw, the improvement is negligible. Nevertheless, we shall present another three significant predictors for logYw as possible candidates for the best predictor when larger set of molecules is taken under consideration. Here, the difference is so small that there is no argument in claiming that either of these descriptors is better than the best predictor in the benchmark set. We present predictors for these four properties (HVAP, DHVAP, TSA and logYw) in Tables 5, 6, 7 and 8.

**Remark 5.** It is interesting to note that connectivity index is also a discrete Adriatic index, namely:

**Table 5.** The significant discrete Adriatic descriptors and the best benchmark predictor with  $R^2$  values for HVAP in the set of octane isomers

Index	$R^2$
max-min rodeg index: $\sum_{uv \in E(G)} \frac{\max\{\sqrt{d_u}, \sqrt{d_v}\}}{\min\{\sqrt{d_u}, \sqrt{d_v}\}} = \sum_{uv \in E(G)} \sqrt{\frac{\max\{d_u, d_v\}}{\min\{d_u, d_v\}}}$	0.9071
misbalance lodeg index: $\sum_{uv \in E(G)}  \ln d_u - \ln d_v $	0.9067
misbalance hadeg index: $\sum_{uv \in E(G)} \left  \left(\frac{1}{2}\right)^{d_u} - \left(\frac{1}{2}\right)^{d_v} \right  = \sum_{uv \in E(G)}  2^{-d_u} - 2^{-d_v} $	0.9044
misbalance rodeg index: $\sum_{uv \in E(G)}  \sqrt{d_u} - \sqrt{d_v} $	0.9035
misbalance irdeg index: $\sum_{uv \in E(G)} \left  \frac{1}{\sqrt{d_u}} - \frac{1}{\sqrt{d_v}} \right $	0.9009
connectivity index $\chi_2$	0.8861

**Table 6.** The significant discrete Adriatic descriptors and the best benchmark predictor with  $R^2$  values for DHVAP in the set of octane isomers.

Index	$R^2$
misbalance lodeg index: $\sum_{uv \in E(G)}  \ln d_u - \ln d_v $	0.9678
misbalance rodeg index: $\sum_{uv \in E(G)}  \sqrt{d_u} - \sqrt{d_v} $	0.9653
max-min rodeg index: $\sum_{uv \in E(G)} \frac{\max\{\sqrt{d_u}, \sqrt{d_v}\}}{\min\{\sqrt{d_u}, \sqrt{d_v}\}} = \sum_{uv \in E(G)} \sqrt{\frac{\max\{d_u, d_v\}}{\min\{d_u, d_v\}}}$	0.9642
misbalance hadeg index: $\sum_{uv \in E(G)} \left  \left(\frac{1}{2}\right)^{d_u} - \left(\frac{1}{2}\right)^{d_v} \right  = \sum_{uv \in E(G)}  2^{-d_u} - 2^{-d_v} $	0.9623
misbalance irdeg index: $\sum_{uv \in E(G)} \left  \frac{1}{\sqrt{d_u}} - \frac{1}{\sqrt{d_v}} \right $	0.9617
misbalance deg index: $\sum_{uv \in E(G)}  d_u - d_v $	0.9534
min-max rodeg index: $\sum_{uv \in E(G)} \frac{\min\{\sqrt{d_u}, \sqrt{d_v}\}}{\max\{\sqrt{d_u}, \sqrt{d_v}\}} = \sum_{uv \in E(G)} \sqrt{\frac{\min\{d_u, d_v\}}{\max\{d_u, d_v\}}}$	0.9526
misbalance losdeg index: $\sum_{uv \in E(G)}  \ln^2 d_u - \ln^2 d_v $	0.9496
misbalance indeg index: $\sum_{uv \in E(G)} \left  \frac{1}{d_u} - \frac{1}{d_v} \right $	0.9493
max-min deg index: $\sum_{uv \in E(G)} \frac{\max\{d_u, d_v\}}{\min\{d_u, d_v\}}$	0.9467
average connectivity index $\chi_1$	0.9201

**Table 7.** The significant discrete Adriatic descriptors and the best benchmark predictor with  $R^2$  values for TSA in the set of octane isomers

Index	$R^2$
inverse sum losdeg index: $\sum_{uv \in E(G)} \frac{1}{\sqrt{\ln(d_u)} + \sqrt{\ln(d_v)}}$	0.7761
inverse sum indeg: $\sum_{uv \in E(G)} \frac{1}{\frac{1}{d_u} + \frac{1}{d_v}} = \sum_{uv \in E(G)} \frac{d_u d_v}{d_u + d_v}$	0.7532
second Mohar index TI2	0.7169

**Table 8.** The significant discrete Adriatic descriptors and the best benchmark predictor with  $R^2$  values for logYw in the set of PCB

Index	$R^2$
max-min deg index: $\sum_{uv \in E(G)} \frac{\max\{d_u, d_v\}}{\min\{d_u, d_v\}}$	0.8271
max-min sdeg index: $\sum_{uv \in E(G)} \frac{\max\{d_u^2, d_v^2\}}{\min\{d_u^2, d_v^2\}} = \sum_{uv \in E(G)} \left( \frac{\max\{d_u, d_v\}}{\min\{d_u, d_v\}} \right)^2$	0.8270
max-rodeg index: $\sum_{uv \in E(G)} \frac{\max\{\sqrt{d_u}, \sqrt{d_v}\}}{\min\{\sqrt{d_u}, \sqrt{d_v}\}} = \sum_{uv \in E(G)} \sqrt{\frac{\max\{d_u, d_v\}}{\min\{d_u, d_v\}}}$	0.8267
connectivity index $\chi_2$	0.8260

$$\text{Randić type indeg index: } \sum_{uv \in E(G)} \frac{1}{\sqrt{d_u d_v}}.$$

**Remark 6.** As mentioned in the remark 5 the connectivity index is also a discrete Adriatic index. Hence, here we have 6 discrete Adriatic indices of comparable quality (6 presented here plus two more mentioned in Table 5).

## CONCLUSIONS

The aim of this paper was to define variable and discrete Adriatic indices and to show the motivation for the definition of these indices. We have started by the analyses of two very popular types of molecular descriptors: Randić-type indices and Wiener-type indices. We have extracted properties of definitions of these indices and used it to define variable Adriatic indices (by *Procedure IV*) and discrete Adriatic indices (by *Procedure V*). Note that discrete Adriatic indices include (up to linear transformation) Randić index, the Zagreb index and the modified Zagreb index. They also include (up to linear transformation) the Wiener index when the number of atoms is fixed. Moreover, variable Adriatic indices include variable Zagreb indices. We expect that these indices will show good predictive properties in many QSAR and QSPR studies. Using, just benchmark sets proposed by International Academy of Mathematical

Chemistry, we have found that following descriptors are useful:

- **Randić type lodeg index:**  $\sum_{uv \in E(G)} \ln(d_u) \cdot \ln(d_v)$

This is the best predictor of heat capacity at constant  $T$  for octane isomers.

- **Randić type sdi index:**  $\sum_{uv \in E(G)} (D_x^2 D_y^2)$

This is the best predictor of motor octane number for octane isomers.

- **Randić type hadi index:**

$$\sum_{uv \in E(G)} 0.5^{D_u} \cdot 0.5^{D_v} = \sum_{uv \in E(G)} \frac{1}{2^{D_u + D_v}}$$

This is the best predictor of density and of molar volume for octane isomers. However, because of the outlier, this result may not be correct.

- **sum lordeg index:**

$$\sum_{uv \in E(G)} (\sqrt{\ln d_x} + \sqrt{\ln d_y}) = \sum_{v \in V(G)} d_x \sqrt{\ln d_x}$$

This is the best predictor of octanol-water partition coefficient for octane isomers.

- **inverse sum lordeg index:**

$$\sum_{uv \in E(G)} \frac{1}{\sqrt{\ln(d_u)} + \sqrt{\ln(d_v)}}$$

This is the best predictor of heat capacity at constant  $P$  and of total surface area for octane isomers.



- **inverse sum indeg:** 
$$\sum_{uv \in E(G)} \frac{1}{\frac{1}{d_u} + \frac{1}{d_v}} = \sum_{uv \in E(G)} \frac{d_u d_v}{d_u + d_v}$$

This is a significant predictor of total surface area for octane isomers.

- **misbalance lodeg index:** 
$$\sum_{uv \in E(G)} |\ln d_u - \ln d_v|$$

This is the best predictor of standard enthalpy of vaporisation for octane isomers. Also, this is a significant predictor of enthalpy of vaporization for octane isomers.

- **misbalance losdeg index:** 
$$\sum_{uv \in E(G)} |\ln^2 d_u - \ln^2 d_v|$$

This is a significant predictor of standard enthalpy of vaporisation for octane isomers.

- **misbalance indeg index:** 
$$\sum_{uv \in E(G)} \left| \frac{1}{d_u} - \frac{1}{d_v} \right|$$

This is a significant predictor of standard enthalpy of vaporisation for octane isomers.

- **misbalance irdeg index:** 
$$\sum_{uv \in E(G)} \left| \frac{1}{\sqrt{d_u}} - \frac{1}{\sqrt{d_v}} \right|$$

This is a significant predictor of enthalpy of vaporisation and of standard enthalpy of vaporisation for octane isomers.

- **misbalance rodeg index:** 
$$\sum_{uv \in E(G)} |\sqrt{d_u} - \sqrt{d_v}|$$

This is a significant predictor of enthalpy of vaporisation and of standard enthalpy of vaporisation for octane isomers.

- **misbalance deg index:** 
$$\sum_{uv \in E(G)} |d_u - d_v|$$

This is a significant predictor of standard enthalpy of vaporisation for octane isomers.

- **misbalance hadeg index:**

$$\sum_{uv \in E(G)} \left| \left( \frac{1}{2} \right)^{d_u} - \left( \frac{1}{2} \right)^{d_v} \right| = \sum_{uv \in E(G)} |2^{-d_u} - 2^{-d_v}|$$

This is a significant predictor of enthalpy of vaporisation and of standard enthalpy of vaporisation for octane isomers.

- **misbalance indi index:** 
$$\sum_{uv \in E(G)} \left| \frac{1}{D_u} - \frac{1}{D_v} \right|$$

This is the best predictor of biological activity for phenethylamines.

- **min-max rodeg index:**

$$\sum_{uv \in E(G)} \frac{\min\{\sqrt{d_u}, \sqrt{d_v}\}}{\max\{\sqrt{d_u}, \sqrt{d_v}\}} = \sum_{uv \in E(G)} \sqrt{\frac{\min\{d_u, d_v\}}{\max\{d_u, d_v\}}}$$

This is a significant predictor of standard enthalpy of vaporisation for octane isomers.

- **min-max sdi index:**

$$\sum_{uv \in E(G)} \frac{\min\{D_u^2, D_v^2\}}{\max\{D_u^2, D_v^2\}} = \sum_{uv \in E(G)} \left( \frac{\min\{D_u, D_v\}}{\max\{D_u, D_v\}} \right)^2$$

This is the best predictor of relative retention time for polychlorobiphenyls.

- **max-min rodeg index:**

$$\sum_{uv \in E(G)} \frac{\max\{\sqrt{d_u}, \sqrt{d_v}\}}{\min\{\sqrt{d_u}, \sqrt{d_v}\}} = \sum_{uv \in E(G)} \sqrt{\frac{\max\{d_u, d_v\}}{\min\{d_u, d_v\}}}$$

This is the best predictor of enthalpy of vaporisation for octane isomers. Also, this is a significant predictor of standard enthalpy of vaporisation for octane isomers and a significant predictor of log water activity coefficient for polychlorobiphenyls, too.

- **max-min deg index:** 
$$\sum_{uv \in E(G)} \frac{\max\{d_u, d_v\}}{\min\{d_u, d_v\}}$$

This is the best predictor of log water activity coefficient for polychlorobiphenyls.

- **max-min sdeg index:**

$$\sum_{uv \in E(G)} \frac{\max\{d_u^2, d_v^2\}}{\min\{d_u^2, d_v^2\}} = \sum_{uv \in E(G)} \left( \frac{\max\{d_u, d_v\}}{\min\{d_u, d_v\}} \right)^2$$

This is a significant predictor of log water activity coefficient for polychlorobiphenyls.

- **symmetric division deg index:**

$$\sum_{uv \in E(G)} \left( \frac{\min\{d_u, d_v\}}{\max\{d_u, d_v\}} + \frac{\max\{d_u, d_v\}}{\min\{d_u, d_v\}} \right)$$

This is the best predictor of total surface area for polychlorobiphenyls.

In this paper, we have just checked predictive properties of one parameter linear models in four benchmark sets proposed by the International Academy of Mathematical Chemistry. However, study of discrete Adriatic indices can be extended much further. Hence, we propose the following open problems - it would be interesting to:

- 1) check usability in multilinear models of discrete Adriatic indices for benchmark sets proposed by the International Academy of Mathematical Chemistry;
- 2) check usability in non-linear one-parameter models of discrete Adriatic indices for benchmark sets proposed by the International Academy of Mathematical Chemistry;
- 3) check usability in non-linear multi-parametric models of discrete Adriatic indices for benchmark sets proposed by the International Academy of Mathematical Chemistry;
- 4) to extend the study of these descriptors beyond benchmark sets proposed by the International Academy of Mathematical Chemistry.

9. *Acknowledgements.* The partial support of Croatian Ministry of Science, Education and Sport (grants no. 177-0000000-0884 and 037-0000000-2779) is gratefully acknowledged.

## REFERENCES

- R. Todeschini and V. Consonni, *Handbook of Molecular Descriptors*, Wiley-VCH, Weinheim, 2000.
- N. Trinajstić, *Chemical Graph Theory*, CRC Press, Boca Raton, 1992.
- J. Devillers and A. T. Balaban (Eds.), *Topological Indices and Related Descriptors in QSAR and QSPR*, Gordon and Breach, Amsterdam, 1999.
- M. Karelson, *Molecular Descriptors in QSAR/QSPR*, Wiley-Interscience, New York 2000.
- <http://www.vcclab.org/lab/edragon/>
- <http://www.codessa-pro.com/>
- <http://www.molgen.de/?src=documents/molgenqspr.html>
- <http://www.molecularDescriptors.eu/software/software.htm>
- M. Randić, *J. Am. Chem. Soc.*, **97** (1975) 6609–6615.
- I. Gutman, B. Ruščić, N. Trinajstić, and C. F. J. Wilcox Jr, *Chem. Phys.*, **62** (1975) 3399–3405.
- A. T. Balaban, *Chem. Phys. Lett.*, **89** (1982) 399–404.
- A. Miličević and S. Nikolić, *Croat. Chem. Acta*, **77** (2004) 97–101.
- H. Wiener, *J. Am. Chem. Soc.*, **69** (1947) 17–20.
- S. Nikolić, N. Trinajstić, and M. Randić, *Chem. Phys. Lett.*, **333** (2001) 319–321.
- I. Gutman, D. Vukičević, and J. Žerovnik, *Croat. Chem. Acta*, **77** (2004) 103–109.
- D. Vukičević and J. Žerovnik, *Acta Chim. Slov.*, **52** (2005) 272–281.
- D. Vukičević and J. Žerovnik, *MATCH Commun. Math. Comput. Chem.*, **53** (2005) 385–402.
- P. V. Khadikar, N. V. Deshpande, P. P. Kale, A. A. Dobrynin, I. Gutman, and G. Domotor, *J. Chem. Inf. Comput. Sci.*, **35** (1995) 547–550.
- P. V. Khadikar, P. P. Kale, N. V. Deshpande, S. Karmarkar, and V. K. Agrawal, *J. Math. Chem.*, **29** (2001) 143–150.
- <http://www.molecularDescriptors.eu/dataset/dataset.htm>
- <http://www.iamc-online.org/>
- G. Caporossi, I. Gutman, and P. Hansen: *Comput. Chem.*, **23** (1999) 469–477.
- M. Fischermann, A. Hoffman, D. Rautenbach, and L. Volkman, *Discrete Appl. Mat.*, **128** (2003) 375–385.
- Lj. Pavlović, *Discrete Appl. Math.*, **127** (2003) 615–626.
- G. Caporossi, I. Gutman, P. Hansen, and Lj. Pavlović, *Comput. Biol. Chem.*, **27** (2003) 85–90.
- D. Vukičević and A. Graovac, *Croat. Chem. Acta*, **77** (2004) 313–319.
- D. Vukičević and A. Graovac, *Croat. Chem. Acta*, **77** (2004) 481–490.
- D. Vukičević and A. Graovac, *Croat. Chem. Acta*, **77** (2004) 501–508.
- D. Vukičević and N. Trinajstić, *Croat. Chem. Acta*, **76** (2003) 183–187.
- D. Vukičević and N. Trinajstić, *MATCH Commun. Math. Comput. Chem.*, **53** (2005) 111–138.
- D. Veljan and D. Vukičević, *J. Math. Chem.*, **40** (2006) 155–178.
- D. Vukičević: On the Edge Degrees of Trees, *Glasnik Matematički*, to appear.
- H. Hosoya, *Bull. Chem. Soc. Jap.*, **44** (1971) 2332–2339.
- D. Bonchev, O. Mekenyan, and N. Trinajstić, *J. Comput. Chem.*, **2** (1981) 127–148.
- M. Randić, *Croat. Chem. Acta*, **74** (2001) 683–705.
- M. Randić, *J. Chem. Inf. Comput. Sci.*, **37** (1997) 1063–1071.
- Y-Q. Yang, L. Xu, and C.-Y. Hu, *J. Chem. Inf. Comput. Sci.*, **34** (1994) 1140–1145.
- P. Hansen and D. Vukičević, *Croat. Chem. Acta*, **80** (2007) 165–168.
- D. Vukičević and A. Graovac, *MATCH Commun. Math. Comput. Chem.*, **57** (2007) 587–590.
- D. Vukičević, *MATCH Commun. Math. Comput. Chem.*, **57** (2007) 633–641.
- D. Vukičević and A. Graovac, *MATCH Commun. Math. Comput. Chem.*, **60** (2008) 37–44.

## SAŽETAK

### Vezno-aditivno modeliranje 1. Jadranski indeksi

Damir Vukičević i Marija Gašperov

*Prirodoslovno-matematički fakultet, Sveučilište u Splitu, Nikole Tesle 12, HR-21000 Split, Hrvatska*

Neki od najpoznatijih molekularnih deskriptora se dobivaju sumiranjem bridnih doprinosa. Primjerice, Randićev indeks i njemu slični deskriptori, Balabanov indeks i njemu slični deskriptori, Wienerov indeks i njegove modifikacije, Szeged indeks, ... U ovom radu su analizirane metode računanja doprinosa birdova. Uočeni su glavni principi po kojima se ovi doprinosi računaju i pomoću njih je definirana klasa molekularnih deskriptora nazvanih *Jadranski indeksi (Adriatic Indices)*.

Posebno je zanimljiva podklasa koja se sastoji od 148 *diskretnih Jadranskih indeksa*. Oni su analizirani na testnim skupovima predloženim od Međunarodne akademije matematičke kemije (International Academy of Mathematical Chemistry) i pokazali su se dobrim prediktorima u mnogim slučajevima. Ovi indeksi se lako kodiraju, te bi mogli naći svoje mjesto u softwarima za kemijsko modeliranje. Ovi indeksi bi mogli unaprijediti razna QSAR i QSPR istraživanja.