MATCH
Communications in Mathematical and in Computer Chemistry

MATCH Commun. Math. Comput. Chem. **64** (2010) 359-372

ISSN 0340 - 6253

New Local Vertex Invariants and Molecular Descriptors Based on Functions of the Vertex Degrees

Roberto Todeschini and Viviana Consonni

Milano Chemometrics and QSAR Research Group, Department of Environmental Science, University of Milano-Bicocca, P.za della Scienza 1, I-20126 Milano, Italy
roberto.todeschini@unimib.it, viviana.consonni@unimib.it

(Received September 25, 2009)

Abstract

Different kinds of Local Vertex Invariants (LOVIs) derived from vertex degrees were reviewed and investigated by comparison. Novel LOVIs, based on a multiplicative form of some vertex degrees defined in literature were also proposed. A data set comprised of 730 diverse molecules with 383 different atom types was used to study similarity/diversity of 17 different LOVIs by Principal Component Analysis.

Finally, 12 different kinds of molecular descriptors were derived from each of the LOVIs in analysis, obtaining a total of 204 molecular descriptors. These descriptors were used to calculate QSPR models for the series of 18 hydrocarbons with 8 carbon atoms (C8).

The molecular descriptors derived from the new LOVIs gave quite surprising results in correlating the physico-chemical properties of octane isomers.

1. Introduction

For several decades molecular graph theory has been playing a very important role in theoretical chemistry, quantitative structure-activity relationships (QSAR) and structure-property relationships (QSPR), and similarity/diversity analysis of molecular libraries. In effect, several molecular descriptors used in these research fields are derived from molecular graphs, which represent chemicals, by the use of well defined algorithms to obtain numbers from graphs and, then, using these numbers to describe molecules [1].

Local vertex invariants (LOVIs) are among the most important quantities derived from molecular graphs, often they being the starting point to calculate molecular descriptors. LOVIs are numerical quantities associated with graph vertices regardless of any arbitrary vertex numbering; they are used to characterize local properties in a molecule. They can be either purely topological, if heteroatoms are not distinguished from carbon atoms, or chemical, if the heteroatoms are assigned distinct values from carbon atoms, even when these are topologically equivalent [2-5].

Vertex degree is the most common LOVI used to describe an atom of a molecule, it being related to atom connectivity. Several different kinds of formula to calculate vertex degrees were proposed in literature so far [1]. Since the graph theory was applied to describe molecules, the concept of vertex degree was largely exploited and several molecular descriptors were derived from it. Nowadays, a lot of different formulations of vertex degree exist and, accordingly, by means of them new molecular descriptors were defined and may be further defined in the future.

From a topological point of view, there are two distinct ways to calculate vertex degree: a) vertex degree takes into account the vertex and its incident edges and b) vertex degree takes into account the vertex, the incident edges and the first neighbour vertices (Figure 1 and Scheme 1).

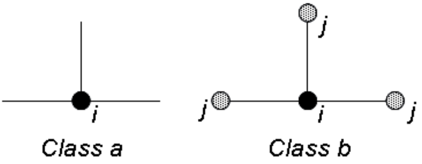

Figure 1. In structure a) only the *i*-th vertex and its incident edges are considered; in structure b) the *j*-th first neighbour vertices of the *i*-th vertex are also considered in the vertex degree calculation.

The *vertex degree* δ of an atom (or *simple vertex degree*) is the count of its σ electrons in the H-depleted molecular graph, i.e., the number of incident edges. This quantity is a local vertex invariant which is easily calculated from the adjacency matrix **A** as the sum of the entries a_{ij} in a matrix row:

$$\delta_i \equiv VS_i(\mathbf{A}) = \sum_{j=1}^{A} a_{ij}$$

where *A* is the number of graph vertices and *VS* is the general symbol for matrix row sum.

The *bond vertex degree*, δ_i^b, is another local invariant, which accounts for atom connectedness and also for bond multiplicity. It is calculated from the atom connectivity matrix **C** as the sum of row entries [6]:

$$\delta_i^b \equiv VS_i(\mathbf{C}) = \sum_{j=1}^{A} a_{ij} \cdot \pi_{ij}^* \qquad \pi_{ij}^* = 0 \text{ if } (i,j) \notin \mathcal{E}(\mathrm{G})$$

where π^* is the conventional bond order, equal to one for single bonds, two for double bonds, three for triple bonds, 1.5 for conjugated bonds, and zero otherwise; a_{ij} are the adjacency matrix elements equal to one only for vertices adjacent to the *i-th* vertex. $\mathcal{E}(G)$ is the set of graph edges.

In order to take into account all valence electrons of an atom, the simple vertex degree can be replaced by the *valence vertex degree*, δ_i^v, (also called *vertex valence*) defined as:

$$\delta_i^v = Z_i^v - h_i = \sigma_i + \pi_i + n_i - h_i$$

where Z_i^v is the number of valence electrons (σ electrons, π electrons and lone pair electrons n) of the *i-th* atom and h_i is the number of hydrogen atoms bonded to it [6]. This definition holds for atoms of the second principal quantum level (C, N, O, F). For atoms of higher principal quantum levels (P, S, Cl, Br, I), Kier and Hall proposed to account for both valence and non-valence electrons, as the following:

$$\delta_i^v = \frac{\left(Z_i^v - h_i\right)}{\left(Z_i - Z_i^v - 1\right)}$$

where Z_i is the total number of electrons of the *i-th* atom, i.e., its atomic number. δ_i^v encodes the electronic identity of the atom in terms of both valence electron and core electron counts; it is a valence electron descriptor.

A modification of the valence vertex degree, still proposed by Kier and Hall [7;8] and used in electrotopological state indices, is the *intrinsic state* I_i defined as:

$$I_i = \frac{(2/L_i)^2 \cdot \delta_i^v + 1}{\delta_i}$$

where L_i is the principal quantum number, δ_i^v is the valence vertex degree and δ_i is the simple vertex degree of the *i-th* atom; the term $(2/L_i)^2$ is equal to one for the elements of the second principal quantum level. For elements of higher levels, L is used to account for the increase in the screening effect of the inner electrons.

The vertex degree, defined by Kupchik [9;10] and denoted as δ^{het}, is the following:

$$\delta_i^{het} = \frac{R_C}{R_i} \cdot \left(Z_i^v - h_i\right) = \frac{R_C}{R_i} \cdot \delta_i^v$$

where R_i and R_C are the covalent radius of the *i-th* atom and the carbon atom, respectively; Z_i^v is the number of valence electrons and h_i the number of hydrogen atoms bonded to the *i-th* atom. As the valence vertex degree, Kupchik's vertex degree accounts both for heteroatoms and bond multiplicity. The second equality holds only for the atoms of the second principal quantum number.

The vertex degree defined by Hu and Xu [11] is calculated according to the following:

$$\delta_i' = \delta_i \cdot \sqrt{Z_i}$$

where Z_i is the atomic number of the considered atom and δ_i the simple vertex degree.

In order to prevent possible degenerative cases where vertex degrees are summed (e.g. $\sqrt{1} + \sqrt{4} = \sqrt{9}$), Alikhanidi *et al*. [12], modified Hu-Xu vertex degree as the following:

$$\delta_i^A = \delta_i \cdot \sqrt{Z_i'}$$

where Z' is a function of the atomic numbers Z_j of the atoms adjacent to the *i-th* atom; this function was called consecutive AT number and defined as:

$$Z_i' = \left[\sum_{j=1}^{A} a_{ij} \cdot \sqrt{\left(\sqrt{2} + Z_j\right)} \right]^2$$

a_{ij} being the elements of the adjacency matrix, which take value of one only for pairs of bonded atoms.

The vertex degree, defined by Ren [13] and denoted as δ^m, is a modification of Kier-Hall intrinsic state I_i defined above :

$$\delta_i^m = \delta_i + \left[\left(\frac{2}{L_i} \right)^2 \cdot \delta_i^v + 1 \right]^{-1} = \delta_i + \left(I_i \cdot \delta_i \right)^{-1}$$

where δ_i is the simple vertex degree of the *i-th* atom, L_i the principal quantum number, and δ_i^v the valence vertex degree. This formula is applied only to heteroatoms or carbon atoms with multiple bonds and/or bonded to heteroatoms; otherwise, the Ren vertex degree coincides with the simple vertex degree δ_i.

The vertex degree defined by Li *et al*. [14], denoted as δ^{Li}, is the following:

$$\delta_i^{Li} = \frac{Z_i^v \cdot \left(Z_i^v - h_i\right)}{L_i^2}$$

A further modification of Kier-Hall vertex degree definition is the vertex degree conceived by Yang in order to better describe atoms also in complex organic compounds by using the Yang's Electronegative Force Gauge Y [15]. It is denoted as δ^Y and defined as [16]:

$$\delta_i^Y = \frac{\left(Z_i^v - h_i\right) \cdot b_i}{L_i^2 \cdot Y_i} = \frac{\delta_i^v \cdot b_i}{L_i^2 \cdot Y}$$

where b_i is the bonding electron number (i.e., the vertex degree in the H-filled molecular graph) and Y_i is the Yang's Electronegative Force Gauge, which reflects the atomic ability of attracting charge in the formed bond.

The *perturbation delta value* was defined in terms of the valence vertex degree δ^v modified by the atomic environment as [17]:

$$\delta_i^p = \delta_i^v + \sum_{j=1}^{A} a_{ij} \cdot \gamma_{ij} \cdot \delta_j^v$$

where the perturbation term of the *i-th* atom is the sum of the valence vertex degrees of its first neighbours (a_{ij} being the elements of the adjacency matrix equal to one for vertices adjacent to the *i-th* vertex), each weighted by parameter γ_{ij} accounting for the type of the bond *i-j*. γ values should be functions of the properties of the connected atoms *i* and *j* (e.g. between –0.30 and +0.30). For $\gamma = 0$, perturbation delta values coincide with the corresponding valence vertex degrees.

The *chemical degree*, defined by Madan *et al*. [18-20] and denoted as δ^c, is calculated by summing up relative atomic weights of all the vertices *j* adjacent to *i*:

$$\delta_i^c = \sum_{j=1}^{A} a_{ij} \cdot \frac{\mathrm{m}_j}{\mathrm{m}_\mathrm{C}}$$

where m_j and m_C are the atomic weights of the *j-th* atom and the carbon atom, respectively; a_{ij} are the elements of the adjacency matrix, which are equal to one only for vertices adjacent to vertex *i*.

A vertex degree, called *CT vertex degree*, accounting for the number of adjacent atoms, heteroatoms, multiple bonds and conjugation was defined [1] as:

$$\delta_i^{CT} = \sqrt{\delta_i \cdot \left(\prod_{j=1}^{A} \left(\pi_{ij}^*\right)^{a_{ij}} \right)^{1/\delta_i} \cdot \sum_{j=1}^{A} a_{ij} \cdot \left(\frac{\mathrm{Z}_i + \mathrm{Z}_j}{2 \cdot \mathrm{Z}_\mathrm{C}} \right)} = \sqrt{\delta_i \cdot \Pi \cdot \mathbb{Z}}$$

where a_{ij} are the elements of the adjacency matrix, π^* the conventional bond order, Z the atomic number. The term Π accounts for multiple bonds and conjugation, while the term $\mathbb{Z}$ for heteroatoms. For saturated hydrocarbons, $\delta_i^{CT} = \delta_i$, Π being equal to 1 and $\mathbb{Z}$ equal to δ_i. Unlike the other vertex degree definitions, this vertex degree distinguishes conjugated atoms from non-conjugated atoms.

In Figure 2, the main features of the 13 different vertex degrees defined above are collected.

Note that for hydrocarbons not containing multiple bonds, the values of valence vertex degree, bond vertex degree, Kupchik's vertex degree, Li's vertex degree, Ren's vertex degree, Madan's vertex degree, and CT vertex degree coincide with the simple vertex degree, that is:

$$\delta_i = \delta_i^v = \delta_i^b = \delta_i^K = \delta_i^{Li} = \delta_i^m = \delta_i^c = \delta_i^{CT}$$

2. New local vertex invariants

A new LOVI that accounts for the degree of a vertex is here defined based on the same expression as the perturbation delta value, where the valence vertex degree is replaced with the intrinsic state as:

$$\delta_i^I = I_i + \frac{\gamma}{\delta_i} \cdot \sum_{j=1}^{A} a_{ij} \cdot I_j$$

This LOVI is here termed *perturbation intrinsic state*. Note that, unlike the perturbation delta value definition, the γ parameter here is a constant not accounting for the type of bond (i-j).

Moreover, substituting the summation with the product into the formula of perturbation delta value, Alikhanidi vertex degree, and perturbation intrinsic state above defined, novel LOVIs can be calculated. These are summarized in Table 1, together with the original additive versions. They are called the *multiplicative perturbation delta value*, the *multiplicative Alikhanidi vertex degree*, and the *multiplicative perturbation intrinsic state.* In functions of type 1 and 3, γ is a real user-defined parameter.

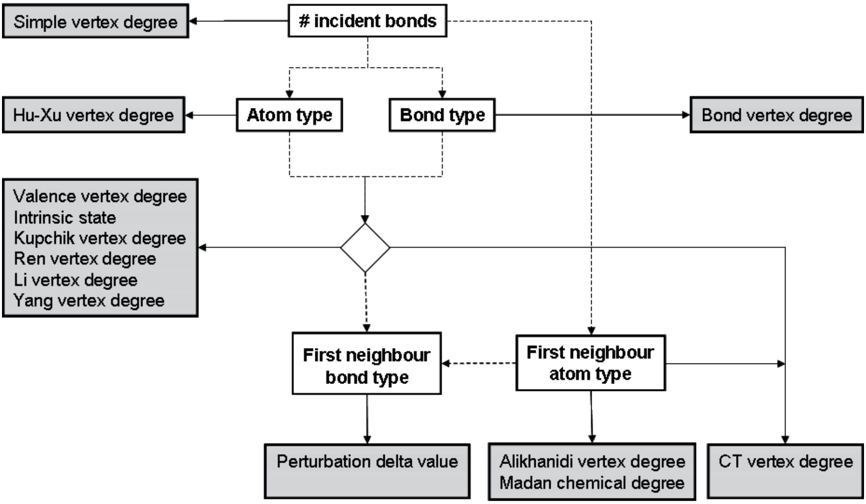

Figure 2. Partition of the considered 13 different vertex degrees, based on the algorithm for their calculations. On the top of the figure, # incident bonds, atom type and bond type refer to the focused i-th atom; on the bottom, first neighbour atom and bond types refer to the j-th atoms bonded to the i-th atom.

Table 1. LOVIs based on the perturbation scheme. In the third column LOVIs based on an additive scheme, are collected, while in the last column those based on a multiplicative scheme. a_{ij} are the elements of the adjacency matrix equal to one for vertices adjacent to the *i-th* vertex; γ is areal user-defined parameter.

Type	*Vertex degree*	*Additive version*	*Multiplicative version*
1	perturbation delta value	$\delta_i^p = \delta_i^v + \sum_{j=1}^{A} a_{ij} \cdot \gamma_{ij} \cdot \delta_j^v$	${}^{\pi}\delta_i^p = \delta_i^v + \prod_{j=1}^{A} \left(\gamma_{ij} \cdot \delta_j^v\right)^{a_{ij}}$
2	Alikhanidi vertex degree	$\delta_i^A = \delta_i \cdot \sqrt{\sum_{j=1}^{A} a_{ij} \cdot \sqrt{\left(\sqrt{2} + Z_j\right)}}$	${}^{\pi}\delta_i^A = \delta_i \cdot \sqrt{\prod_{j=1}^{A} \left[\sqrt{\left(\sqrt{2} + Z_j\right)}\right]^{a_{ij}}}$

3	perturbation intrinsic state	$\delta_i^I = I_i + \frac{\gamma}{\delta_i} \cdot \sum_{j=1}^{A} a_{ij} \cdot I_j$	$^{\pi}\delta_i^I = I_i + \gamma \cdot \left(\prod_{j=1}^{A} I_j^{a_{ij}} \right)^{1/\delta_i}$

3. Molecular descriptors

From the LOVIs early defined, 12 molecular descriptors were calculated. The former group (Table 2, ID 1-5) includes molecular descriptors derived from the classical algorithms applied to local vertex invariants of a molecular graph. The latter group (Table 2, ID 6-12) consists of spectral indices derived from a diagonal matrix **V** where LOVIs are located on the main diagonal:

$$[\mathbf{V}]_{ij} = \begin{cases} \mathcal{L}_i & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

$\mathcal{L}_\mathrm{i}$ being one of the different vertex degrees of the *i-th* atom.

12 molecular descriptors were calculated for each of the 17 LOVIs considered in this study (i.e., 13 classical LOVIs defined in the introduction plus the 4 new LOVIs), thus obtaining a total of 204 descriptors. For both perturbation delta value and its multiplicative version, the parameter γ_{ij} was selected equal to 0.1, regardless the bond type, whereas for perturbation intrinsic state and its multiplicative version, the parameter γ has was fixed to 0.5.

Table 2. Molecular descriptors defined in terms of LOVIs ($\mathcal{L}$) (ID 1-5) and derived from the eigenvalues (λ) obtained from the **V** matrix (ID 6-12). **V** is a diagonal matrix where LOVIs are located on the main diagonal.

ID	*Formula*	*generalized molecular descriptors*
1	$S = \sum_{i=1}^{A} \mathcal{L}_i$	sum-like indices
2	$Ssqinv = \sum_{i=1}^{A} \mathcal{L}_i^{-2}$	reciprocal squared sum-like indices
3	$M_1 = \sum_{i=1}^{A} \mathcal{L}_i^2$	first Zagreb-like indices
4	$M_2 = \sum_{i=1}^{A-1} \sum_{j=i+1}^{A} a_{ij} \cdot \left(\mathcal{L}_i \cdot \mathcal{L}_j \right)$	second Zagreb-like indices
5	$^{1}\chi = \sum_{i=1}^{A-1} \sum_{j=i+1}^{A} a_{ij} \cdot \left(\mathcal{L}_i \cdot \mathcal{L}_j \right)^{-1/2}$	Randic-like indices
6	$SpSum_{+}^{-1}(\mathbf{V},1) = \sum_{i=1}^{A} \frac{1}{\lambda_i} \qquad \lambda_i > 0$	reciprocal eigenvalue sum
7	$SpMax(\mathbf{V},1) = \max_i \{\lambda_i\}$	leading eigenvalue

8	$SpDiam(\mathbf{V},1)=\max(\lambda_i)-\min(\lambda_i)$	spectral diameter
9	$SpAD(\mathbf{V},1)=\sum_{i=1}^{A}\left\|\lambda_i-\bar{\lambda}\right\|$	eigenvalue absolute deviation
10	$SpMAD(\mathbf{V},1)=\sum_{i=1}^{A}\left\|\lambda_i-\bar{\lambda}\right\|/A$	eigenvalue mean absolute deviation
11	$EE(\mathbf{V},1)=\sum_{i=1}^{A}e^{\lambda_i}$	Estrada-like indices
12	$\mathrm{Ho}(\mathbf{V};1)=\sum_{i=0}^{A}\left\|c_i\right\|$	Hosoya-type indices

4. Similarity/diversity analysis

In order to evaluate relationships among the different vertex degrees here considered, a data set of 730 diverse chemicals was selected and for each atom of each molecule the whole set of vertex degrees was calculated. A total of 7268 row vectors each containing the values of the 17 LOVIs were derived from all the atoms of all the data set molecules. Then, only the diverse row vectors were selected, thus finally obtaining 483 atom types. Some information about the data set used for this analysis are given in Table 3.

Table 3. Basic descriptive statistics of the data set comprised of 730 diverse compounds.

	Mean	*Min*	*Max*	*Std. Dev.*
Molecular Weight	156.10	26.04	578.65	87.46
No. heavy atoms	9.97	2	40	5.21
No. cycles	1.11	0	11	1.41
No. circuits	3.43	0	552	23.83
No. benzene rings	0.91	0	11	1.30
Rotatable Bond Number	1.34	0	21	2.41
No. double bonds	0.41	0	6	0.74
No. triple bolds	0.03	0	2	0.17
No. aromatic bonds	5.45	0	50	6.87
No. nitrogen atoms	0.23	0	4	0.57
No. oxygen atoms	0.86	0	8	1.15
No. phosphorus atoms	0.01	0	1	0.08
No. sulphur atoms	0.04	0	2	0.21
No. fluorine atoms	0.21	0	7	0.88
No. chlorine atoms	0.62	0	8	1.43
No. bromine atoms	0.07	0	5	0.45
No. iodine atoms	0.01	0	2	0.13

In order to denote the defined LOVIs in the most simple way, the following notations were adopted in subsequent analysis: simple vertex degree (d), valence vertex degree (vd), bond vertex degree (bd), Hu-Xu vertex degree (HX), Kupchik vertex degree (Kup), Alikhanidi vertex degree (Alik), Li vertex degree (Li), Ren vertex degree (Ren), Madan chemical degree (Mad), Yang vertex degree (Yang), CT vertex degree (CT), perturbation delta value (p), intrinsic state (I), multiplicative perturbation delta value (mp), perturbation intrinsic state (pI), multiplicative perturbation intrinsic state (mI), and multiplicative Alikhanidi index (mAlik). Moreover, symbols for molecular descriptors were composed by two distinct pairs, the former indicating the type of descriptor operator and the latter the type of LOVI (e.g., SpAD-mp refers to the spectral absolute deviation of the matrix collecting the multiplicative perturbation delta values of non-H atoms).

A Principal Component Analysis (PCA) was then performed on the resulting matrix of dimension 483×17. The loading plots of PC1 vs PC2, PC3 vs PC4, and PC5 vs PC6 are reported in Figures 3, 4, and 5, respectively. The first six principal components explain 96.8% of the total variance.

The first component (E.V. 49.4%), which represents the maximum variance direction, is determined by the majority of the vertex degrees, i.e., simple, bond, Alikhanidi in both additive and multiplicative versions, Ren, Hu-Xu, Yang, CT vertex degrees, all partially in contrast to the vertex degrees based on the instrinsic state, e.g. intrinsic state together with its additive and multiplicative versions); the second component (E.V 27.6%) characterizes the remaining group of the vertex degrees, i.e. valence vertex degree, perturbation delta value, Li vertex degree, and Kupchik vertex degree) still together with the vertex degree based on the intrinsic state. The third component (E.V. 7.9%) is strongly characterized by the multiplicative version of the perturbation delta value (mp) together with some contributions of the Madan chenical degree (Mad), the multiplicative version of the Alikhanidi vertex degree (mAlik), and the additive (pI) and multiplicative (mI) versions of the intrinsic states. The fourth component (E.V. 6.0%) reveals the different behaviour between perturbation (p) and Yang vertex degree (Yang) on one side and the valence vertex degree (vd) on the other side. The fifth component (E.V. 3.4%) specifically characterizes the difference between the Madan chemical degree (Mad) and the multiplicative perturbation delta value (mp); finally, the sixth component (E.V. 2.6%) collects some residual specific information of the bond vertex degree (bd), both additive (pI) and multplicative perturbation intrinsic state (mI), and the CT vertex degree (CT).

In conclusion, it is possible to say that the first PC highlights a group of largely correlated LOVIs which explain almost the same information as the simple vertex degree, whereas the second PC identifies those LOVIs that behave similarly to the valence vertex degree.

LOVIs based on the intrinsic states seem to constitute a third group that gives a quite different information. Moreover, looking at the remaining components, both multiplicative and additive versions of the perturbation delta value, Madan chemical degree, multiplicative version of Alikhanidi degree, and Yang vertex degree, show some peculiarities not accounted for by the first two principal components.

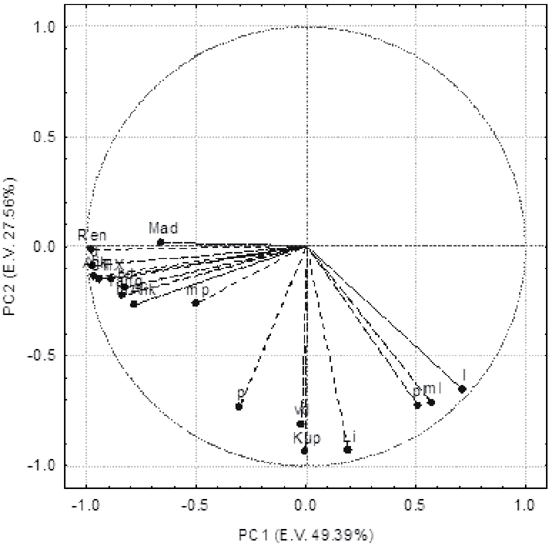

Figure 3. Loading plot of the first and second principal components for the 17 considered LOVIs. Symbols are explained in the text.

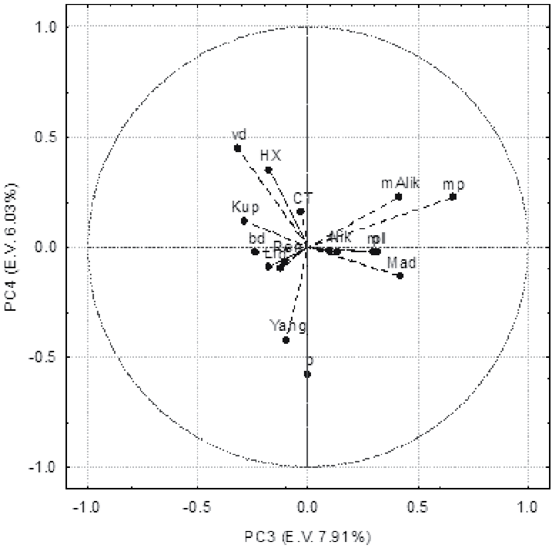

Figure 4. Loading plot of the third and fourth principal components for the 17 considered LOVIs. Symbols are explained in the text.

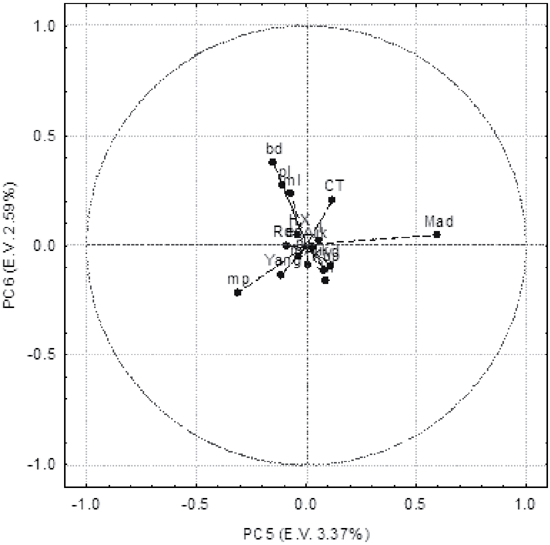

Figure 5. Loading plot of the fifth and sixth principal components for the 17 considered LOVIs. Symbols are explained in the text.

5. QSAR/QSPR modeling

In order to check the modeling ability of the novel LOVIs, a simple congeneric data set was chosen constituted by the 18 isomers of the C8 hydrocarbons. Five physico-chemical properties were studied: boiling point (bp), heat of formation (Hf), motor octane number (MON), total surface area (TSA), and molar refractivity (MR). The molecular descriptors used in this analysis are the 12 molecular descriptors defined above calculated by means of the 17 vertex degrees, for a total of 204 molecular descriptors. For each property, the Q^2 leave-one-out technique was used to select the best models. Only univariate regression models were considered in this analysis.

Before discussing results, it should be noted that the molecular descriptors *SpAD* and *SpMAD* are correlated one in the case of a congeneric series of compounds, thus SpAD was arbitrary chosen in the subsequent analysis. Moreover, remember that the following relationship

$$\delta_i = \delta_i^v = \delta_i^b = \delta_i^K = \delta_i^{Li} = \delta_i^m = \delta_i^c = \delta_i^{CT}$$

holds for this data set and, accordingly, the molecular descriptors derived from them all coincide. Therefore, only descriptors based on the simple vertex degree (d) were considered.

As it can be easily observed, for all the properties, except for the heat of formation (Hf), the new descriptors based on the multiplicative approach to generate LOVIs gave the best results. In particular, for the boiling point (bp), the eigenvalue absolute deviation calculated by the multiplicative version of the perturbation delta value (SpAD-mp) showed a prediction ability about 21% larger than that of the most predictive old descriptor, i.e., the second Zagreb-like index calculated using the intrinsic state as the local vertex invariant (M$_2$-I).

Table 4. QSPR models for the 5 studied properties of C8 compound series. For each property, the 3 best models obtained by using both the old and new LOVIs (13 + 4) are collected, while the remaining models are the 3 best models constituted by the 13 old LOVIs. Descriptors based on the new LOVIs are highlighted in boldface.

bp (all)	Q^2 (R^2)	MR (all)	Q^2 (R^2)	TSA (all)	Q^2 (R^2)
SpAD-mp	80.4 (85.1)	**$^1\chi$-mI**	89.0 (92.3)	**$^1\chi$-mI**	83.5 (86.9)
EE-mp	67.5 (74.3)	**M$_2$-mp**	87.8 (90.8)	**M$_2$-mp**	80.9 (85.6)
SpDiam-mp	66.7 (72.5)	**$^1\chi$-mp**	86.7 (89.8)	**$^1\chi$-mp**	79.9 (84.7)
M$_2$-I	59.6 (73.3)	Ho-p	60.6 (70.8)	Ho-p	76.0 (80.8)
$^1\chi$-Alik	59.5 (73.3)	M$_2$-Alik	49.5 (58.4)	SpSum^{-1}-p	55.8 (65.2)
$^1\chi$-HX	53.1 (67.4)	Ssqinv-p	35.8 (49.7)	Ssqinv-p	52.8 (63.4)

MON (all)	Q^2 (R^2)	Hf (all)	Q^2 (R^2)
SpSum^{-1}-mp	92.8 (94.6)	M$_2$-I	74.6 (79.3)
Ssqinv-mp	92.7 (94.5)	$^1\chi$-Alik	74.5 (79.2)
Ho-mI	89.8 (92.6)	**$^1\chi$-mAlik**	73.7 (78.1)
$^1\chi$-I	88.7 (92.0)	$^1\chi$-HX	67.0 (72.4)
SpAD-p	83.0 (87.1)	$^1\chi$-d	66.9 (72.3)
M$_2$-p	79.9 (85.7)	$^1\chi$-p	61.6 (67.6)

For molar refractivity (MR), an even more surprising result was obtained: the predictive ability of the Randic-like index calculated by the multiplicative version of the intrinsic state ($^1\chi$-mI) resulted 28% larger than the predictive ability of any other old molecular descriptor. Similar results were also obtained for total surface area (TSA) and motor octane number (MON), even if with less significant difference in predictive ability than the other descriptors.

In the case of the heat of formation (Hf), the second Zagreb index calculated by the intrinsic state (M$_2$-I) and the Randic-like index calculated using the Alikhanidi vertex degree ($^1\chi$-Alik) gave a little bit larger predictive ability than the first multiplicative descriptors, i.e., the first order Randic-like index based on the multiplicative version of Alikhanidi vertex degree ($^1\chi$-mAlik**).**

For what it is concerned with the descriptors based on the perturbation intrinsic state (pI), we can say that, even if it does not appear within the 3 best models, it gave interesting predictive results for molar refractivity ($^1\chi$-pI: Q^2 = 73.1), total surface area ($^1\chi$-pI: Q^2 = 77.8), and motor octane number (Ssqinv-pI: Q^2 = 87.7).

6. Conclusions

Due to the basic role of the vertex degree in graph theory and in deriving well-known molecular descriptors, such as connectivity and Zagreb indices, some new vertex degrees were defined, derived from the perturbation delta value.

Comparison of 17 vertex degrees performed on a data set of diverse molecules showed the existence of different classes of vertex degrees, for which the prototypes are simple vertex degree, valence vertex degree, perturbation delta value, Madan chemical degree and intrinsic state. Moreover, in modeling some physico-chemical properties of a set of isomeric compounds, namely the hydrocarbons with eight carbon atoms, quite surprising results were obtained. Indeed, molecular descriptors based on the newly defined vertex degrees showed higher prediction ability than that obtained by the classical vertex degrees. In particular, the multiplicative version of the perturbation delta value and the multiplicative version of the perturbation delta value where the valence vertex degree was replaced by the intrinsic state showed high predictive ability in modeling physico-chemical properties of C8 data set.

It should be expected that, for non-congeneric data sets and/or for modeling biological activities, other global descriptors may be taken into account in multivariate models, in order to better explain holistic behaviour of molecules; however, interesting contributions of the descriptors derived from the new vertex degrees may be also expected to be useful.

Acknowledgements

This study has been financed by funds PRIN 2007 (National Ministry of University and Research and University of Milano-Bicocca, code 2007R57KT7).

References

[1] R. Todeschini, V. Consonni, *Molecular Descriptors for Chemoinformatics* (*2 volumes*), Wiley-VCH Verlag GmbH, Weinheim, Germany, 2009, p. 1257.

[2] A.T. Balaban, *Numerical Modelling of Chemical Structures: Local Graph Invariants and Topological Indices*, in: R.B. King, D.H. Rouvray (Eds.), Graph Theory and Topology in Chemistry, Elsevier, Amsterdam (The Netherlands), 1987, pp. 159-176.

[3] P.A. Filip, T.-S. Balaban, A.T. Balaban, A New Approach for Devising Local Graph Invariants: Derived Topological Indices with Low Degeneracy and Good Correlation Ability, *J.Math.Chem*. **1** (1987) 61-83.

[4] O. Ivanciuc, T.-S. Balaban, A.T. Balaban, Design of Topological Indices. Part 4. Reciprocal Distance Matrix, Related Local Vertex Invariants and Topological Indices, *J.Math.Chem*. **12** (1993) 309-318.

[5] A.T. Balaban, Local vs. Global (*i.e.* Atomic versus Molecular) Numerical Modeling of Molecular Graphs, *J.Chem.Inf.Comput.Sci*. **34** (1994) 398-402.

[6] L.B. Kier, L.H. Hall, *Molecular Connectivity in Structure-Activity Analysis*, Research Studies Press - Wiley, Chichester, UK, 1986, p. 262.

[7] L.B. KierL.H. Hall, An Electrotopological-State Index for Atoms in Molecules, *Pharm.Res*. **7** (1990) 801-807.

[8] L.B. Kier, L.H. Hall, *Molecular Structure Description. The Electrotopological State*, Academic Press, London, UK, 1999, p. 246.

[9] E.J. Kupchik, Structure-Molar Refraction Relationships of Alkylsilanes Using Empirically-Modified First Order Molecular Connectivity Indices, *Quant.Struct.-Act.Relat*. **5** (1986) 95-98.

[10] E.J. Kupchik, Structure-Molar Refraction Relationships of Alkylgermanes Using Molecular Connectivity, *Quant.Struct.-Act.Relat*. **7** (1988) 57-59.

[11] C.-Y. HuL. Xu, Developing Molecular Identification Numbers by an All-Paths Method, *J.Chem.Inf.Comput.Sci.* **37** (1997) 311-315.

[12] S. AlikhanidiY. Takahashi, New Molecular Fragmental Descriptors and Their Application to the Prediction of Fish Toxicity, *MATCH Commun.Math.Comput.Chem*. **55** (2006) 205-232.

[13] B. Ren, Novel atom-type AI indices for QSPR studies of alcohols, *Computers Chem*. **26** (2002) 223-235.

[14] X. Li, A.F. Jalbout, M. Solimannejad, Definition and application of a novel valence molecular connectivity index, *J.Mol.Struct.(Theochem)* **663** (2003) 81-85.

[15] P. Yang, *Distribution and Physical Property in Molecule*, Union Press of Shanxi University, Tai Yuan, China, 1992

[16] Y.-R. Jiang, J.-Y. Liu, Y.-H. Hu, T. Fujita, Novel Topological Index for Research on Structure-Property Relationships of Complex Organic Compounds, *J.Comput.Chem*. **24** (2003) 842-849.

[17] V.K. Gombar, A. Kumar, M.S. Murthy, Quantitative Structure-Activity Relationships: Part IX - A Modified Connectivity Index as Structure Quantifier, *Indian J.Chem.* **26B** (1987) 1168-1170.

[18] A. GoelA.K. Madan, Structure-Activity Study on Antiinflammatory Pyrazole Carboxylic Acid Hydrazide Analogs Using Molecular Connectivity Indices, *J.Chem.Inf.Comput.Sci*. **35** (1995) 510-514.

[19] S. Gupta, M. Singh, A.K. Madan, Applications of graph theory: Relationships of molecular connectivity index and atomic molecular connectivity index with anti-HSV activity, *J.Mol.Struct.(Theochem)* **571** (2001) 147-152.

[20] S. Bajaj, S.S. Sambi, A.K. Madan, Prediction of Anti-Inflammatory Activity of N-Arylanthranilic Acids: Computational Approach Using Refined Zagreb Indices, *Croat.Chem.Acta* **78** (2005) 165-174.