MODIFIED CONNECTIVITY INDEX AND THEIR APPLICATION TO ALKYL ALCOHOL QSPR MODELS

E. Cornwell*

SUMMARY.

A modification of topological index ($^{1}{}^{h}_{corn}$) was proposed. This modification consist in the numerical evaluation of all elements a_{ij} of the matrix used in $^{1}{}^{h}$ index definition, applying a product operator ($1/D^2$) to all original matrix a_{ij} elements. This factors a_{ij} are function of the distances between the graph vertexes (i, j) (molecular atoms distances).

The models obtained using this modified index ($^{1}{}^{h}_{corn}$) present a more good statistically indices respect when using the original one. Nevertheless the means errors produced on the differences of experimental and calculated dependent variables are not significantly different between both models.

Key words: Alcohols, modified connectivity index, mathematical model

INDICE DE CONECTIVIDAD MODIFICADO Y SUS APLICACIONES A MODELOS QSPR PARA ALCOHOLES ALQUÍLICOS

RESUMEN

Se propone un índice topológico modificado ($^{1}{}^{h}_{corn}$) Esta modificación consiste en una evaluación numérica de todos los elementos a_{ij} de la matriz usada en la definición del índice ^{1h}, aplicando un operador producto ($1/D^2$) a todos los elementos originales a_{ij} . Estos factores a_{ij} son función de la distancia entre los vértices del grafo (i, j) (distancias atómicas moleculares).

El modelo obtenido usando esta modificación en el índice ($^{1}{}^{h}_{corn}$), presenta mejores parámetros estadísticos respecto cuando se usa el índice original. Sin embargo, el error promedio producido en la diferencia entre la variable dependiente experimental y la calculada no muestran diferencias significativas entre ambos modelos.

Palabras clave: Alcoholes, índice de conectividad modificado, modelo matemático.

INTRODUCTION

The most important problem of quantitative structure-property relationship (QSPR) studies remains to be mathematical representation of chemical structure, that is, a representation of molecular structure with mathematical descriptors. Significant progress has been reported during recent years in the development of various topological, geometric, electrostatic and quantum chemical indices, to be used as molecular descriptor[1-2]. Because of the simplicity of topological structural representation, these indices are sometimes preferred

[1] Departamento de Química Inorgánica y Analítica. Facultad de Ciencias Químicas y Farmacéuticas. Universidad de Chile, Olivos 1007, Santiago, Chile. E-mail ecornwel@ abello.dic.uchile.cl

to more complicated geometric, electrostatic, and quantum chemical descriptors. In general, there are causal relationships between the chemical structure of molecules and their physical properties. By constructing the quantitative structure relationship (QSPR) one would be able to predict the properties of any molecule and also obtained more profound insight into the different interaction present in any system from a theoretical view-point. A variety of QSPR models used physicochemical parameters such as dependent variables [3]; among these are boiling points and n-octanol / water partition coefficient (log P).

Topological indices (TIs) are numerical graph invariants that quantitatively characterize molecular structure. A graph G = (V, E) is an ordered pair of two sets V and E , the former representing a nonempty set and the latter representing unordered pairs of elements of the set V. When V represent the atoms of a molecule and elements of E symbolize covalent bonds between pairs of atoms, then G becomes a molecular graph (or constitutional graph, because there is no stereochemical information). Such a graph depicts the topological of the chemical species. A graph is characterized using graph invariants. An invariant may be a polynomial, a sequence of number, or a single number. A numerical graph invariant (i.e., a single number) that characterizes the molecular structure is called a topological index [4].

The original index ^{1}h to be treatise in this paper has a new idea on the adjacency matrix[5] and is the base from the modification presented in this issue by the author.

METODOLOGY

The authors C. Yang and C. Zhong [5], when used the molecular connectivity index (MCI) proposed other kind of values on the molecular grafos vertex (d_{ij}), indicating that by this way it is possible obtained a more discriminative differences between molecules. These authors defined the values o each grapho vertex (d_{ij}) by the sum of adjacency matrix $[A]_{mxm}$ raw elements. Each elements of $[A]_{mxm}$ matrix a_{ij} is obtained using the following criteria:

$$a_{ij} = \begin{cases} n + & i = j \\ 0 & i \neq j, \text{ vertexes i and j are no connected} \\ 1 + n_{Hj}/6 & i \neq j, \text{ vertexes i and j are connected} \end{cases}$$

Where n is a number of lone-pair electrons and is the number of double bond of the atoms corresponding to vertex i, n_{Hj} is the number of hydrogen atoms connecting to the non-hydrogen atoms j.

For the first order (using one bridge counting (bond) between i, j vertex (atoms)) to evaluate the present topological indexes (MCI) used in this paper, the mathematical expression to evaluated $^{1}h(5)$ is.

$$^{1}\chi^{h} = \sum_{J=1}^{N} \prod_{I=1}^{2} (d_i)_j^{-0.5} \qquad (1)$$

The proposed index ($^{1}{}^{h}_{corn}$) is evaluated by the same expression 1
From this particular order one index, a modified MCI ($^{1}{}^{h}_{corn}$) is proposed, based on $^{1}{}^{h}$ topological index but using and extended criteria to obtained matrix $[C]_{mxm}$ in which each element or grafo vertex c_{ij} value is obtained by the sum of the matrix $[C]_{mxm}$ raw elements. In this case y proposed that each elements c_{ij} were defined by the following criteria.

$$c_{ij} = \begin{cases} n + & i = j \\ (1 + n_{Hj} / 6)\, 1/D^2 & i \neq j, \text{ vertexes i and j are connected} \\ & \text{if, } i \neq j, \text{ vertexes i and j are connected, then } D = 1 \end{cases}$$

Where D is an operator, its values are indicated in Table1 in conjunction with the fractional values that indicated the distances between i an j vertexes (atoms) The distances are topological ones, using the short way between two vertexes, the fractional values are obtained in function of the modification premises concerning the definition of c_{ij}. In this case, each numerical evaluated vertex (atom, except hydrogen) contains the information of distances of all other vertexes (atoms of the molecule), that's a different situation that when used the concept used by the original authors[5] explain in the case of a_{ij} where the adjacencies distances are used.

To obtained modificated connectivity index ($^{1}{}^{h}_{corn}$) of a molecule, is used relation 1 considering c_{ij} values.
A numerical example to show the differences between both indexes, the original one and the proposed in this paper are present using n-propanol like a substrate. Matrices $[A]_{mxm}$, $[C]_{mxm}$; the sum of the matrices raw elements, the (d_{ij}) vertexes grapho values and the MCI topological indexes $^{1}{}^{h}$, $^{1}{}^{h}_{corn}$ are calculated as follow:

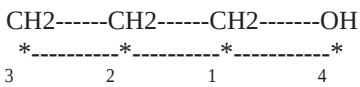
Substance: $\overset{3}{C}H_3\text{-}\overset{2}{C}H_2\text{-}\overset{1}{C}H_2\text{-}\overset{4}{O}H$ (the way to enumerate the atoms is indifferent to calculus). Original adjacency matrix to obtained $^{1}{}^{h}$ is:

$$[A]_{mxm} = \begin{vmatrix} 0 & 1{,}3333 & 0 & 1{,}1667 \\ 1{,}3333 & 0 & 1{,}5000 & 0 \\ 0 & 1{,}3333 & 0 & 0 \\ 1{,}3333 & 0 & 0 & 4 \end{vmatrix}$$

Raw sum is:

$$\sum_{j=1}^{4} a_{ij} = 2,5000 + 2,8333 + 1,3333 + 5,3333$$

Grapho:



Values of each vertex (d_{ij}) is: (1,3333; 2,8333; 2,5000; 5,3333)
The C. Yang and C. Zhong index is:

$$^{1}\chi^{h} = \sum_{j=1}^{N} \prod_{i=1}^{2} (d_i)_j^{-0.5} = (1,3333*2,8333)^{-0.5} + (2,8333*2,500)^{-0.5} +$$

(2,5000*5,3333) = 1,1641 (published element for to 1-propanol in reference 5)

Proposed index by the present author paper is:

Substance: $\underset{1}{CH_3}\text{-}\underset{2}{CH_2}\text{-}\underset{3}{CH_2}\text{-}\underset{4}{OH}$ (the way to enumerate atoms is indifferent to calculus)

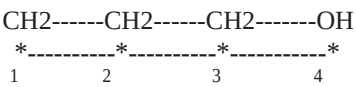
The matrix for calculated the proposed index $^{1}{}^{h}_{corn}$ is:

$$[C]_{mxm} = \begin{vmatrix} 0 & 1,3333 & 0,3333 & 0,1296 \\ 1,5000 & 0 & 1,3333 & 0,2917 \\ 0,3750 & 1,3333 & 0 & 1,1667 \\ 10,1667 & 0,3333 & 0 & 4 \end{vmatrix}$$

The raw sum is:

$$\sum_{j=1}^{4} c_{ij} = 1,7962 + 3,1250 + 2,8750 + 5,8333$$

Grapho:



Values of each vertex (d_{ij}) is: (1,7962, 3,1250, 2,8750, 5,8332)

The proposed index is:

$$ {}^{1}\chi^{h}_{corn} = \sum_{j=1}^{N} \prod_{i=1}^{2} (d_i)_j^{-0.5} = (1,7962*3,1250)^{-0.5} + (3,1250*2,8750)^{-0.5} + (2,8750*5,8333)^{-0.5} = 1,0000 $$

To facilitate the calculus, y proposed the following program using Hewlett Packard 48 GX, if not, used Excel program.

1. Wright two list X and Y with C_{ij} extracted from the grafo {1,7962 3,1250 2,8750}, {3,1250 2,8750 5,8333}

2. Wright and keep in memory the following program before its named

```
            << → X Y
  << X Y *0,5 ^ INV    LIST
>>
   >>
```

3. Put your two lists in stack 1 and 2 by ENTER function

4. Executed the program

Table 1 present 38 alcohols and two groups of physicals properties; column 2 and 3 are alcohols boiling points and logarithm values of octanol-water distributions constants (log P)[6,7] asterisk (* , **) in column 1 are calculated log P from on-line free software[8,9]. Columns 4, 5, 8, 9 are: calculated boiling points, error defined from the difference of experimental boiling point minus calculated boiling points, calculated log P and its error defined like the experimental values minus calculated one, all using C. Yang topological index (${}^{1}{}^{h}$) At the same way columns 6, 7,10,11 are obtained using the proposed ${}^{1}{}^{h}_{corn}$ by me.

Table 2 are values assigned to different molecular carbons types and operator used in function of the topological distances between vertices, these values are a_{ij} elements of $[c]_{mxm}$ matrix discussed belonging to an specific alcohol.

Table 1. Boiling points and octanol water partition constant (logP) Original ($^{1}{}^{h}$) and modificated Index ($^{1}{}^{h}_{corn}$) of alkyl alcohols (*,**) means calculated values obtained by software. All calculated (Calc) are obtained by means of the treatise indices

	Substances	Boiling Points °C	log (P)	Boiling Points Calc. °C C. Yang	error C. Yang	Boiling Points Calc. °C Corn.	error Corn.	log (P) calc. C. Yang	error C. Yang	log (P) calc. Corn.	error Corn.	$^{1}{}^{h}$	$^{1}{}^{h}_{corn}$
	Alkyl Alcohols												
1	ethanol	78,0	-0,30	65,5	12,48	69,4	8.58	-0,39	0,09	-0,23	-0,07	0,7955	0,7360
2	propanol	97,1	0,25	84,2	12,93	89,7	7,39	0,15	0,10	0,35	-0,10	1,1641	1,0000
3	butanol	117,6	0,88	103,2	14,44	109,4	8,19	0,70	0,18	0,91	-0,03	1,5395	1,2563
4	2-methyl-1-propanol	108,1	0,76	100,5	7,57	101,6	6,55	0,63	0,13	0,68	0,08	1,4875	1,1541
5	2-butanol	99,5	0,61	102,1	-2,64	104,6	-5,07	0,67	-0,06	0,77	-0,16	1,5194	1,1934
6	2-methyl-2-propanol	82,4	0,35	98,5	-16,13	93,7	-11,28	0,57	-0,22	0,46	-0,11	1,4481	1,0517
7	pentanol	138,0	1,56	122,1	15,87	129,0	8,96	1,25	0,31	1,46	0,10	1,9145	1,5118
8	3-methyl-1-butanol	131,0	1,16	119,7	11,26	116,4	14,57	1,19	-0,03	1,11	0,05	1,8672	1,3477
9	2-pentanol	119,3	1,19	120,9	-1,56	121,8	-2,48	1,22	-0,03	1,26	-0,07	1,8893	1,4173
10	3-pentanol	116,2	1,21	121,9	-5,75	124,4	-8,22	1,25	-0,04	1,33	-0,12	1,9109	1,4516
11	3-methyl-2-butanol	112,9	1,28	119,1	-6,22	116,8	-3,87	1,17	0,11	1,12	0,16	1,8549	1,3521
12	2-methyl-2-butanol	102,9	0,89	119,4	-17,12	114,0	-11,72	1,18	-0,29	1,04	-0,15	1,8609	1,3163
13	hexanol	157,6	2,03	141,1	16,50	148,1	9,50	1,80	0,23	2,01	0,02	2,2895	1,7598
14	2-ethyl-1-butanol*	147,0	1,78	140,8	6,18	140,4	6,60	1,80	-0,02	1,79	-0,01	2,2838	1,6596
15	3,3-dimethyl-1-butanol*	143,0	1,57	136,6	6,37	136,5	6,53	1,67	-0,10	1,68	-0,11	2,2011	1,6085
16	2-hexanol	140,0	1,76	139,8	0,15	144,5	-4,46	1,77	-0,01	1,90	-0,14	2,2647	1,7124
17	2,2-dimethyl-1-butanol*	136,5	1,57	138,3	-1,75	129,2	7,30	1,72	-0,15	1,47	0,10	2,2331	1,5139
18	3-hexanol	135,0	1,65	140,7	-5,68	143,5	-8,49	1,79	-0,14	1,87	-0,22	2,2811	1,6998
19	4-methyl-2-pentanol*	131,6	1,67	137,3	-5,69	135,2	-3,63	1,69	-0,02	1,64	0,03	2,2142	1,5923
20	2-methyl-2-pentanol	121,1	1,53	133,4	-12,32	133,1	-12,00	1,58	-0,05	1,58	-0,05	2,1376	1,5646
21	2,4 dimethyl-3-pentanol*	138,7	2,31	155,5	-16,75	147,2	-8,50	2,22	0,09	1,98	0,33	2,5731	1,7481
22	2,3 dimethyl-3-pentanol*	139,7	1,67	157,2	-17,54	144,4	-4,73	2,27	-0,60	1,90	-0,23	2,6084	1,7120
23	2,2 dimethyl-3-pentanol*	145,0	2,27	155,7	-10,68	143,4	1,57	2,23	0,04	1,87	0,40	2,5776	1,6990
24	3,3-dimethyl-2-butanol*	120,4	1,75	131,7	-11,29	125,0	-4,64	1,53	0,22	1,35	0,40	2,1035	1,4597
25	2,3-dimethyl-2-butanol*	118,4	1,48	136,7	-18,28	122,9	-4,48	1,68	-0,20	1,29	0,19	2,2020	1,4316
26	heptanol	176,4	2,72	160,1	16,32	166,7	9,74	2,35	0,37	2,53	0,19	2,6645	2,0013
27	2-heptanol	159,0	2,31	158,8	0,18	161,8	-2,84	2,32	-0,01	2,40	-0,09	2,6397	1,9386
28	3-heptanol	157,0	2,24	159,7	-2,67	162,2	-5,21	2,34	-0,10	2,41	-0,17	2,6564	1,9434
29	2,4-dimethyl-1-pentanol*	159,0	2,19	155,9	3,08	151,5	7,49	2,23	-0,04	2,10	0,09	2,5823	1,8041
30	4-heptanol	156,0	2,44	159,4	-3,40	162,6	-6,60	2,34	0,10	2,42	0,02	2,6512	1,9485
31	1-octanol	195,1	3,07	179,0	16,05	185,8	9,35	2,90	0,17	3,07	0,00	3,0395	2,2497
32	2-octanol	180,0	2,90	177,8	2,21	181,1	-1,06	2,87	0,03	2,94	-0,04	3,0147	2,1886
33	4-octanol	176,3	2,68	178,4	-2,10	180,8	-4,51	2,89	-0,21	2,93	-0,25	3,0266	2,1854
34	1-nonanol	213,3	3,67	198,0	15,28	203,8	9,53	3,45	0,22	3,58	0,09	3,4145	2,4841
35	3-nonanol*	195,0	3,36	199,0	-3,98	198,7	-3,74	3,48	-0,12	3,44	-0,08	3,4334	2,4187
36	4-nonanol*	192,5	3,36	197,6	-5,06	199,4	-6,87	3,44	-0,08	3,46	-0,10	3,4054	2,4269
37	5-nonanol*	193,0	3,36	197,4	-4,39	199,4	-6,36	3,44	-0,08	3,46	-0,10	3,4020	2,4267
38	decanol**	231,1	4,24	217,0	14,11	222,2	8,90	4,01	0,23	4,11	0,13	3,7895	2,7239

Table 2. Assignation values to different carbon types and alcohol group in function of topological distances between vertices.

Operator	$-CH_3$	$-CH_2$	-CH	-C-	-OH
Value	Value	Value	Value	Value	Value
1	1,5000	1,3333	1,1667	1,0000	1,1667
4	0,3750	0,3333	0,2917	,2500	0,2917
9	0,1667	0,14281	0,1296	0,1111	0,1296
16	0,0938	0,0833	0,0729	0,0625	0,0729
25	0,0600	0,0533	0,0467	0,0400	0,0467
36	0,0417	0,0370	0,0324	0,0278	0,0324
49	0,0306	0,0272	0,0238	0,0204	0,0238
64	0,0234	0,0208	0,0182	0,0156	0,0182
81	0,0185	0,0165	0,0144	0,0123	0,0144
100	0,0150	0,0133	0,0117	0,0100	0,0117
121	0,0124	0,0110	0,0096	0,0083	0,0096

The mathematical model using the original topological index ($^1\ ^h$)[5] applied to the alcohols boiling points (B.P.) and log P are represented by equations 1 and 2

$$\begin{aligned} \text{B.P.} &= 50{,}5931\,(2{,}6943) * {}^1\ ^h + 25{,}2709\,(6{,}5558) \qquad 1)\\ r &= 0{,}9525\\ \text{s.d.} &= 11{,}0443\\ F &= 352{,}5946\\ \log(P) &= -1{,}5541\,(\ 0{,}1098) * {}^1\ ^h + 1{,}4670\,(\ 0{,}04515) \qquad 2)\\ r &= 0{,}9834\\ \text{s.d} &= 0{,}1851\\ F &= 1055{,}8395 \end{aligned}$$

The modeling using the proposed modification on original topological index ($^1\ ^h_{corn}$) used by C. Yang and C. Zhong[5] ($^1\ ^h$) applied to the alcohols boiling points (B.P.) and log P are represented by equations 3 and 4.

$$B.P. = 12{,}8552(4{,}9008) + 76{,}8533(2{,}7836)*{}^{1}h_{corn} \quad 3)$$
$$r = 0{,}9772$$
$$s.d. = 7{,}7056$$
$$F = 762{,}2778$$
$$\log(P) = -1{,}8330(0{,}1032) + 2{,}1808(0{,}0586)*{}^{1}h_{corn} \quad 4)$$
$$r = 0{,}9873$$
$$s.d. = 0{,}1622$$
$$F = 1384{,}8899$$

In Table 3 are present the principal statistically relation for both topological index proposition, based on experimental and calculated linear regressions for the alcohols boiling points and log P ratio.

Table 3. Statistical parameters from boiling points and log (P)

Treatise	**r**	**s.d**	**F**
From experimental and calculated boiling points, Yang Proposition	0,9526	11,0440	352,608
From experimental and calculated boiling points, proposed in this issue	0,9772	7,7058	762,251
From experimental and calculated log (P) Yang Proposition	0,9832	0,1858	1 047,2997
From experimental and calculated log(P) proposed in this issue	0,9872	0,1623	1 382,739

The symbols on the head of Table 3 are: Correlation coefficient, correlation standard deviations and the Fisher statistically index.

The correlation type used is $y = m*x + n$

Equations 1,2,3, and 4 has a p-value in ANOVA analysis less than 0.01; this means there is a statistically significant relationship between variables (dependent-independent) at 99% confidence level. For B. P. vs. ${}^{1}h_{corn}$ the mean absolute error (MAE) is 6,882 it is the average value of the residual. Since the p-value is greater than 0,05 for Durbin- Watson statistic test

(D.W.), its means there is no indication of serial autocorrelation in the residuals. For log(P) vs. $^{1}h_{corn}$ MAE is 0,1375 that is the average of the residual; the D. W. test is the same than the preceding one, that is, there is no indicative of serial autocorrelation in the residuals. The differences of the experimental boiling points and calculated boiling points from C. Yang proposition[5] and the differences of the experimental boiling points and calculated using my proposition based on the C. Yang modificated topological index named in both cases errors, Table 1 (column 5, 7) present for independent t-Test (errors of C.Yang and errors of my proposition) a means value of $5,26E^{-4}$ and variance of 0,.034 and $-5,26E^{-4}$ and variance of 0,026 respectability with a value of Student test of -0,027 with a probability of 0,978. These means that at the 0,05 level, the two means are not significantly different; the same treatise with errors from log (P) Table 1 (column 9,11) are made, with a means values of $-5,26E^{-4}$ and variance of 118,67 and $-2,63E^{-4}$ and variance of 57,77 for case of C. Yang and my proposition on log (P) study with a value of Student test of $1,22E^{-4}$ with a probability of 0,999, these means at the 0,05 level that the two means are not significantly different.

All calculus were made based on special software and literature[10, 11,12]

The residuals for boiling points, or root-mean-square error (rms), (RMSE) [13,14] obtained for Yang and the proposed in this paper are 10,7497 and 7,5001, respectively and for log (P) 0,1801 and 0,1579, respectively. This indicated a better model when using $^{1}h_{corn}$ index than the original one.

CONCLUSSIONS

- The redefinition of c_{ij} when i and j are not adjacency elements to obtained $[C]_{mxm}$ matrix, produced minor RMSE values for calculated boiling points and log(P) treatise that when used models in which its incorporated original topological index used by C. Yang and C. Zhong (^{1}h) and superior statistically linear regressions parameters.

- The new index is easy to applied with the aid of Table 2

REFERENCES

1. A. R. Katritzky.; E. V. Gordeeva. *J. Chem Inf. Comput. Sci.* (1993) **33**, 835
2. M. Karelson., V. S. Lobanov., A. R. Katritzky. *Chem Rev.* (1996) **96**,1027
3. J. Ariens. "Drug Design" Volume I .Academic Press (1971) page 271-445.
4. S.C. Basak., A. T. Balaban G. D. Grundwald, D.D. Gute. . *J. Chem Inf. Comput Sci*. (2000) **40**, 891
5. C. Yang., C. Zhong., *J. Chem Inf. Comput. Sci.* (2003) **43**, 1998
6. L. H. Hall.; L. B. Kier. *J.Chem Inf. Comput. Sci.* (1995) **35**, 1039
7. M. H. Abraham.; H. S. Ghadha. ; G. S. Whiting. ; R. C. Mitchel. *J. Pharm. Sci.* (1994) **83**, 1085

8. // logkow.cisti.nrc.ca/logkow/search.html (*)

9. // 146.107.217.178/web/alogps/ (* *)

10. Statgraphic Plus 5.1 Copyright 1994-2001Statistically Graphic Corp.

11. Origin 73R1 V7. 0301 (B30019) Copyright © 1991-2002 Origin Lab. Corporation. One Round Plaza Northampton MA 01060 USA.

12. D. L Massart., B.G.M: Vadeginste ., S.N.N. Deming., Y. Machotte., L. Kaufman " Chemometric a textbook" Elsevier Scientific Publishng Company, Amsterdam, 1998.

13. S. Lin., S. Cai, C. Cao., Z. Li. *J.Chem Inf. Comput. Sci.* (2000) **40**, 1337 // en wikipedia.org/wiki/Root_ mean_ square_ desviation