MATCH
Communications in Mathematical and in Computer Chemistry

MATCH Commun. Math. Comput. Chem. **56** (2006) 217-230

ISSN 0340 - 6253

STUDY ON QSPR OF ALCOHOLS WITH A NOVEL EDGE CONNECTIVITY INDEX ^{m}F

Lailong Mu[*], *Changjun Feng*

Department of Chemistry, Xuzhou Normal University, Xuzhou 221116,P.R.China

Lu Xu

Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Changchun, 130022,P.R.China

(Received March 18,2005)

ABSTRACT

A novel edge degree f_i for heteroatom and multiple bonds in molecular graph is derived on the basis of the edge degree $\delta(e_r)$. A novel edge connectivity index ^{m}F is introduced. The multiple linear regression by using the edge connectivity index ^{m}F and alcohol-type parameter δ, alcohol-distance parameter L can provide high-quality QSPR models for the normal boiling points (BPs), molar volumes (MVs), molar refraction (MRs), water solubility(log(1/S)) and octanol/water partition (logP) of alcohols with up to 17 non-hydrogen atoms. The results imply that these physical properties may be expressed as a liner combination of the edge connectivity index and alcohol-type parameter, δ, alcohol-distance parameter, L. For the models of the five properties, the correlation coefficient r and the standard errors are 0.9969,3.022; 0.9993, 1.504; 0.9992, 0.446; 0.9924,0.129 and 0.9973,0.123 for BPs, MVs, MRs, log(1/S) and logP, respectively. The cross-validation by using the leave-one-out method demonstrates the models to be highly reliable from the point of view of statistics.

1. INTRODUCTION

The quantitative structure-property/activity relationship (QSPR/QSAR) studies of organic compounds have been a focus of great attention by the scientists for a long time[1-3]. Large number of QSPR/QSAR models have been developed by using various model parameters to describe and predict the physical properties and biological activities of organic compounds from their molecular structures. Among these model parameters, the molecular topological indices are particularly interested because they can be derived

[*] Tel:+86-516-3403163;Fax: +86-516-3403164

E-mail address: mull@xznu.edu.cn

directly from the molecular structures without experimental efforts.

There is a great number of topological indices, in which we can find molecular connectivity index(χ)[4], Hosoya Z index[5], Balaban J index[6], Bonchev I_D index[7], Schluze MTI index[8],and Wiener's index W[9], Electrotopological state(E-state) index[10], Ren's AI index[11-15], Fernandez's orthogonal descriptors[16], Li's m-connectivity index[17]etc.. However, most of the existing topological indices largely limit their field of application for lack of information on multiple bonds and heteroatoms in molecular graphs. In order to differentiate the multiple bonds and heteroatom in molecular graphs, several different empirical and unempirical approaches have been introduced in the past few years, such as Kier and Hall's concept of valence molecular connectivity[4,18], Bogdanov's topographic distance matrices[19], Randic's three-dimensional molecular descriptors[20], Estrada's approach of edge weights using quantum-chemical parameters[21], and the 3D topological indices[22,23], etc.

Recently a parameter of revised bond intensity (S_j) and edge valence(f_i) were proposed, on the basis of adjacency matrix and edge valence (f_i) , a novel connectivity index (${}^mF$) was developed. The connectivity index ${}^mF$ has good correlations for the boiling points of chain hydrocarbons, aldehydes and alkanones[24]. In this paper we are going to verify the high potential of these indices for applications to five physical properties of alcohols such as the normal boiling points (BPs), molar Volumes (MVs), molar refractions (MRs), water solubility (log(1/S)) and Octanol/water partition (logP).

2. METHODS

Estrada's edge-adjacency matrix of simple graph G={V,E}, where V is the vertex set and E is the edge set, is a square and symmetric matrix $E=[g_{i,j}]_{m\times n}$, where m is the number of edges in a graph. If $v_r \in V/v_r \sim e_i, e_j, g_{i,j}=1$, otherwise $g_{i,j}=0$. Edge degree, $\delta(e_r)$ is defined as the sum of element of r*th* row or column in E matrix[25]:

$$\delta(e_r) = \sum_i g_{i,r} = \sum_j g_{r,j} \qquad (1)$$

For heterographs, i.e., graphs containing heteroatom or multiple bonds, it is necessary to use weight set W. Set W can be selected from different ways. Estrada selected Kc-x parameters as the W set[21], successfully applied this approach to a QSPR study for a dataset compounds having different heteroatoms. On the basic of this, Estrada extended the concept of edge connectivity index to a series of indices[26,27]. In this paper we propose to use the S parameters as set W. The S parameters are related to the electronegativity χ_p [28]of the bonding atomic orbit:

$$S = (\chi_{p_A} + \chi_{p_B}) / 4.96 \qquad (2)$$

where χ_{p_A}, χ_{p_B} are the orbit electronegativities of atom A and atom B, respectively.

Table 1. Electronegativity of Atomic orbit

Atomic rbit	C(sp^3)	C(sp^2)	C(sp)	O(sp^3)	O(sp^2)
χ_p	2.48	2.75	3.29	4.93	5.54

For C-C (single bond),C=C(double bond) ,C-O (single bond),C=O(double bond), the S parameter are 1.000, 1.1089, 1.4940, 1.6714, respectively.

Elements $g_{i,j}$ of the edge of the weighted adjacency matrix E are defined as follows: if If $v_r \in V(v_r \sim e_i, e_j)$, e_j have weight S, $g_{i,j}$=S, otherwise $g_{i,j}$=0. Hence, E for heterographs is a square but nonsymmetric matrix, and, as a consequence, a novel edge degree f_i is calculated only as the sum of elements of i*th* row

$$f_i = \sum_j g_{i,j} \qquad (3)$$

Novel edge connectivity index ^mF can be calculated with the same expression used by using the graphic invariant of Randic[4]:

$$^mF = \sum_{j=1}^{n_m} \prod_{1}^{m+1} (f_i)_j^{-0.5} \qquad (4)$$

where, m is the order of the molecular connectivity index and n_m is the number of the relevant paths.

The alcohol-distance parameter, L, is defined as: $L=\Sigma d_i$, where d_i is the distance (the number of bonds) between the carbon atom C_i and -OH group. The alcohol-type parameter, δ, is vertex degree of the carbon atom connecting –OH group.

3. DATA SETS

Alcohols represent an attractive class of polar compounds for QSPR/QSAR studies considering the influence of the hydrogen-bonding interaction. Five physical properties of alcohols are as: the normal BPs are taken from ref. 11. MVs are calculated as MW/d, where MW is the molecular weight and d is the density (g/cm^3) at 20 °C. MRs are calculated according to the Lorentz–Lorenz expression:

$$MR = \frac{n_0^2 - 1}{n_0^2 + 2} MV \qquad (5)$$

where, n_0 is the index of refraction at 20 °C, and taken from refs. 29 and 30. The experimental BPs are given in Table 2. The experimental values of MVs and MRs are shown in Table 3. The experimental water solubility as log(1/S), where S is the solubility in moles per liter, are taken from ref.11, and listed in Table 4. The experimental data of logP are taken from ref. 11, and shown in Table 5.

4. RESULTS AND DISCUSSION

Only multiple regression analysis (MRA) is used to construct the prediction models in this study, because the results obtained by using approach MRA are good enough.

4. 1. *Correlations to BP*

If only the index ^{0}F is used, a quite good correlation for the BPs of the 138 can be obtained. The correlation coefficient *r* is 0.9847. When the distance-parameter *L* is included, the model can be significantly improved (r = 0.9919); whereas the results by using three-parameters, ^{0}F , $L^{0.5}$, *and L,* as well as four-parameters, ^{0}F , $L^{0.5}$, *L* , and *δ* are more better (*r* = 0.9954, and 0.9969, respectively). The improvements of the results may be as the introduction of *δ* and *L* indices contain information about the hydrogen-bonding interaction, because it is the important factor to influent the BPs of the alcohols.

The model by using four parameters is shown as follows:

$$BP=-6.3810(\pm2.6283)+15.5021(\pm0.6444)\,{}^{0}F+25.9277(\pm3.1851)\delta^{-1}$$
$$+20.4945(\pm1.2887)L^{0.5}-0.8447(\pm0.0690)L$$
$$r = 0.9969,\ s = 3.022,\ F = 5425,\ N = 138 \qquad (6)$$

The results of t-test show that these variables in this model are significant. This model explains more than 99.3% of the variances in the experimental values of BPs for these compounds.

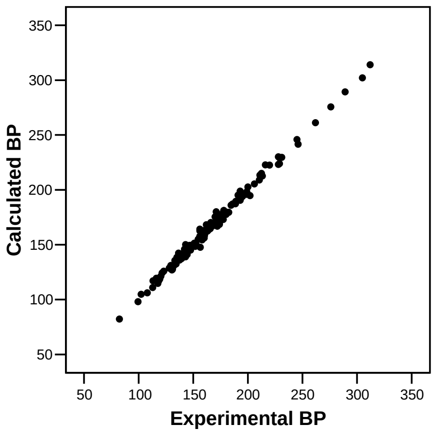

Figure 1. A plot of calculated versus experimental BPs for 138 alcohols

Mihalic and Trinajstic[31] once pointed out that a good QSPR model for BPs should be that r>0.99 and s<5.0 °C. Obviously, our results satisfy the criterion. The calculated results from Eq. (6) for 138 compounds are shown in Table 2. The average absolute deviation is

2.265 °C. The calculated BPs versus experimental data is shown in Figure 1. From this figure again we can see that the model is quite excellent.

The results obtained in this study seem better than those in the literatures[11,15,32-36].

Table 2. The calculated and Experimental Boiling Points(BPs) of 138 Alcohols

No	Compound	BP(ºC) exp	calcd	res	No	Compound	BP(ºC) exp	calcd	res
1	1-Butanol	117.6	114.7	2.9	70	3,4-dimethyl-2-hexanol	165.5	164.6	0.9
2	2-methyl-1-propanol	107.9	106.2	1.7	71	2,5-dimethyl-2-hexanol	154.5	154.9	-0.4
3	2-butanol	99.5	98.0	1.5	72	4-methyl-4-heptanol	161.0	161.1	-0.1
4	2-methyl-2-propanol	82.4	82.2	0.2	73	2,4,4-trimethyl-1-pentanol	168.5	169.7	-1.2
5	1-Pentanol	137.5	136.0	1.5	74	3-ethyl-3-hexanol	160.5	163.2	-2.7
6	3-methyl-1-butanol	131.0	127.6	3.4	75	2,3-dimethyl-2-hexanol	160.0	156.2	3.8
7	2-pentanol	119.3	118.7	0.6	76	3,5-dimethyl-3-hexanol	158.0	154.5	3.5
8	2-methyl-1-butanol	128.0	128.7	-0.7	77	2,3-dimethyl-3-hexanol	158.1	154.4	3.7
9	3-pentanol	116.2	119.5	-3.3	78	2-methyl-3-ethyl-2-pentanol	156.0	157.3	-1.3
10	3-methyl-2-butanol	112.9	111.0	1.9	79	2,4,4-trimethyl-2-pentanol	147.5	145.1	2.4
11	2,2-dimethyl-1-propanol	113.1	117.1	-4.0	80	2,2,4-trimethyl-3-pentanol	150.5	150.7	-0.2
12	2-methyl-2-butanol	102.3	104.8	-2.5	81	2,2-dimethyl-3-hexanol	156.0	157.6	-1.6
13	1-Hexanol	157.0	156.5	0.5	82	2,5-dimethyl-3–hexanol	157.5	159.2	-1.7
14	4-methyl-1-pentanol	151.9	148.5	3.4	83	4,4-dimethyl-3-hexanol	160.4	161.0	-0.6
15	2-hexanol	140.0	138.9	1.1	84	3,4-dimethyl-2-hexanol	165.5	164.6	0.9
16	3-methyl-1-pentanol	153.0	149.6	3.5	85	6-methyl-2-heptanol	174.0	170.0	4.0
17	2-methyl-1-pentanol	148.0	148.5	-0.5	86	3-methyl-1-heptanol	186.0	187.1	-1.1
18	3-hexanol	135.0	138.5	-3.5	87	2-methyl-3-ethyl-3-pentanol	158.0	156.5	1.5
19	2-ethyl-1-butanol	146.5	149.4	-2.9	88	2,3,4-trimethyl-3-pentanol	156.5	147.7	8.8
20	4-methyl-2-pentanol	132.0	130.7	1.3	89	1-Nonanol	213.3	212.6	0.7
21	3,3-dimethyl-1-butanol	143.0	139.0	4.0	90	7-methyl-1-octanol	206.0	205.4	0.6
22	2,3-dimethyl-1-butanol	144.5	141.2	3.3	91	2-nonanol	198.5	195.9	2.6
23	2-methyl-2-pentanol	121.5	124.3	-2.8	92	3-nonanol	195.0	194.5	0.5
24	3-methyl-2-pentanol	134.3	132.5	1.8	93	4-nonanol	192.5	191.6	0.9
25	2-methyl-3-pentanol	129.5	131.1	-1.6	94	5-nonanol	193.0	190.5	2.5
26	2,2-dimethyl-1-butanol	136.5	142.3	-5.8	95	2-methyl-2-octanol	178.0	181.2	-3.2
27	3-methyl-3-pentanol	123.0	125.9	-2.9	96	2,6-dimethyl-2-heptanol	173.0	173.7	-0.7
28	3,3-dimethyl-2-butanol	120.4	121.4	-1.0	97	2,6-dimethyl-3-heptanol	175.0	177.4	-2.4
29	2,3-dimethyl-2-butanol	118.4	117.1	1.3	98	2,6-dimethyl-4-heptanol	174.5	175.2	-0.7
30	1-heptanol	176.4	176.0	0.4	99	3,6-dimethyl-3-heptanol	173.0	173.0	0.0
31	5-methyl-1-hexanol	170.0	168.4	1.6	100	2,2,3-trimethyl-3-hexanol	156.0	164.2	-8.2
32	2-heptanol	160.4	158.6	1.8	101	3,5-dimethyl-4-heptanol	171.0	180.0	-9.0
33	4-methyl-1-hexanol	173.0	169.8	3.2	102	2,3-dimethyl-3-heptanol	173.0	172.0	1.0
34	2-methyl-1-hexanol	164.0	163.5	0.5	103	2,4-dimethyl-4-heptanol	171.0	170.7	0.3
35	3-heptanol	157.0	157.5	-0.5	104	2-methyl-3-ethyl-3-hexanol	177.5	173.0	4.5
36	3-methyl-1-hexanol	169.0	168.6	0.4	105	2-methyl-3-ethyl-1-hexanol	193.0	198.8	-5.8
37	4-heptanol	156.0	156.0	0.0	106	5-methyl-3-ethyl-3-hexanol	172.0	173.0	-1.0
38	5-methyl-2-hexanol	151.0	150.7	0.3	107	2,4,4-trimethyl-3-hexanol	170.0	170.4	-0.4
39	2-methyl-3-hexanol	145.5	149.0	-3.5	108	3,4,4-trimethyl-3-hexanol	165.5	166.3	-0.8
40	2-methyl-2-hexanol	143.0	150.1	-7.1	109	4-methyl-4-octanol	180.0	178.4	1.6
41	2,4-dimethyl-1-pentanol	159.0	155.7	3.3	110	4-ethyl-4-heptanol	182.0	179.6	2.5
42	5-methyl-3-hexanol	148.0	149.5	-1.5	111	2-methyl-2-octanol	178.0	181.2	-3.2
43	3-methyl-3-hexanol	143.0	144.1	-1.1	112	1-decanol	231.1	229.7	1.4
44	2,4-dimethyl-2-pentanol	133.1	135.6	-2.5	113	8-methyl-1-nonanol	219.9	222.6	-2.7
45	2,4-dimethyl-3-pentanol	140.0	141.8	-1.8	114	2-decanol	211.0	213.4	-2.4
46	3-ethyl-3–pentanol	142.0	146.0	-4.0	115	4-decanol	210.5	209.0	1.5
47	2,3-dimethyl-2-pentanol	139.7	137.8	1.9	116	3,7-dimethyl-1-octanol	212.5	215.0	-2.5

Table 2. Continued.

No	Compound	BP(ºC) exp	calcd	res	No	Compound	BP(ºC) exp	calcd	res
48	2,3-dimethyl-3-pentanol	139.0	137.2	1.8	117	2,7-dimethyl-3-octanol	193.5	195.2	-1.7
49	2,3,3-trimethyl-2-butanol	130.5	127.0	3.5	118	2,6-dimethyl-4-octanol	195.0	194.0	1.1
50	3-methyl-2-hexanol	151.0	151.5	-0.5	119	2,3-dimethyl-3-octanol	189.0	189.5	-0.5
51	1-Octanol	195.2	194.8	0.4	120	5-methyl-5-nonanol	202.0	194.7	7.3
52	6-methyl-1-heptanol	188.6	187.3	1.3	121	4-methyl-1-nonanol	216.0	222.8	-6.8
53	2-octanol	180.0	177.6	2.4	122	2-methyl-3-nonanol	200.0	202.7	-2.7
54	3-octanol	175.0	176.2	-1.2	123	2,2,5,5-tetramethyl-3-hexanol	170.0	175.5	-5.5
55	4-methyl-1-heptanol	188.0	188.1	-0.1	124	4-propyl-4-heptanol	191.0	195.2	-4.2
56	4-octanol	176.3	175.1	1.3	125	2,4,6-trimethyl-4-heptanol	181.0	179.6	1.4
57	2-ethyl-1-hexanol	184.6	186.0	-1.4	126	3-ethyl-3-octanol	199.0	198.3	0.7
58	2-methyl-2-heptanol	156.0	162.7	-6.7	127	3-ethyl-2-methyl-3-heptanol	193.0	190.9	2.1
59	2,5-dimethyl-1-hexanol	179.5	179.0	0.5	128	1-undecanol	245.0	245.8	-0.8
60	5-methyl-2-heptanol	172.0	166.9	5.1	129	2-undecanol	228.0	230.2	-2.2
61	6-methyl-3-heptanol	174.0	168.5	5.5	130	3-undecanol	229.0	229.3	-0.3
62	3,5-dimethyl-1-hexanol	182.5	179.6	2.9	131	5-undecanol	229.0	223.9	5.1
63	3-methyl-2-heptanol	166.1	170.2	-4.1	132	6-undecanol	228.0	223.2	4.8
64	2-methyl-3-heptanol	167.5	167.0	0.5	133	1-dodecanol	261.9	261.2	0.7
65	2-methyl-4-heptanol	164.0	166.0	-2.0	134	2-dodecanol	246.0	241.6	4.4
66	5-methyl-3-heptanol	172.0	169.8	2.2	135	1-tridecanol	276.0	275.7	0.3
67	3-methyl-3-heptanol	163.0	162.4	0.6	136	1-tetradecanol	289.0	289.3	-0.3
68	4-methyl-3-heptanol	170.0	169.6	0.4	137	1-pentadecanol	304.9	302.1	2.8
69	3-methyl-4-heptanol	162.0	168.3	-6.3	138	1-hexadecanol	312.0	314.0	-2.0

4.2 *Correlations to MV*

The experimental values of MVs for 42 alcohols are listed in Table 3. The edge connectivity index ^1F produces a good regression for MVs, and the correlation coefficient r is 0.9985. If introducing the alcohol-type parameter δ as another variable, the result is improved greatly (r = 0.9992). Whereas, if ^0F is included, the model can not be improved obviously. The three-parameter model is shown as follows:

$$MV=41.7022(\pm1.1831)+31.1873(\pm0.9217)\,^1F$$
$$-18.1628(\pm2.8923)\delta^{-1}+1.2016(\pm0.6451)\,^0F$$
$$r = 0.9993,\ s = 1.504,\ F = 9125,\ N = 42 \qquad (7)$$

Similarly, the results of the t-test show that the variables contained in the equation are significant. The model explains more than 99.8% of the variances in experimental values of MVs for these compounds with a mean absolute deviation 1.149. Evidently, this model is also quite excellent.

The calculated results from the above models are shown in Table 3. A comparison of calculated and experimental data for MVs is shown in Figure 2.

4.3 *Correlations to MR*

The MR is also a useful parameter for chemical, biological and pharmaceutical sciences. The experimental values of MRs for 41 alcohols are shown in Table 3. The edge connectivity index ^1F produces a good regression for MRs and the correlation coefficient r

is 0.9987. Inclusion of the alcohol-type parameter δ as an additional parameter, the result is slightly improved (r = 0.9990). But, when the edge connectivity index ${}^0F$ is included, the model can be improved significantly. The three-parameter model is shown as follows:

$$MR=7.6429(\pm0.3541)+8.1996(\pm0.2839)\,{}^1F-4.3272(\pm0.8981)\,\delta^{-1}+0.6938(\pm0.2005)\,{}^0F$$

$$r = 0.9992,\ s = 0.446,\ F = 8016,\ N = 41 \qquad (8)$$

As above, the t-test has been made, and the results show that the parameters contained in equation (8) are significant. The model accounts for 99.8% of the variances for the experimental values of MRs for 41 compounds with a mean absolute deviation 0.323. The calculated results by using equation (8) are shown in Table 3. A comparison of calculated and experimental data for MRs is shown in Figure 3. The agreement between the calculated and experimental data is quite good.

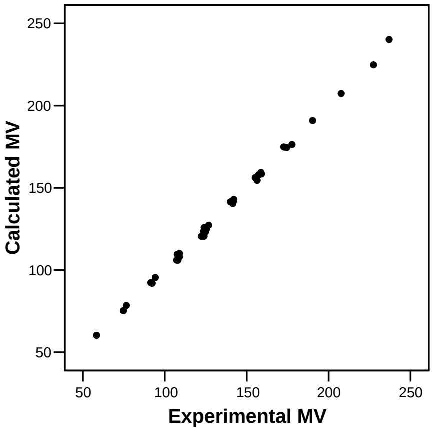

Figure 2. A plot of calculated versus experimental MVs for 42 alcohols.

Table 3. Molar volumes (MVs) and molar refraction (MRs) of alcohols

No.	Compound	MV(cm^3/mol)			MR(cm^3/mol)		
		Exp	Calcd	Res	Exp	Calcd	Res
1	1-Ethanol	58.368	60.296	-1.928	12.927	13.449	-0.522
2	1-Propanol	74.798	75.243	-0.445	17.565	17.690	-0.125
3	2-propanol	76.561	78.414	-1.853	17.613	18.200	-0.587
4	1-Butanol	91.529	92.346	-0.817	22.145	22.458	-0.313
5	2-methyl-1-propanol	92.338	91.985	0.353	22.182	22.215	-0.033
6	2-butanol	91.903	92.063	-0.160	22.144	22.133	0.011
7	2-methyl-2-propanol	94.216	95.407	-1.191	22.033	22.713	-0.680
8	1-Pentanol	108.160	108.774	-0.614	26.798	27.048	-0.250
9	3-methyl-1-butanol	108.559	109.115	-0.556	26.770	26.978	-0.208
10	2-pentanol	108.962	110.063	-1.101	26.724	27.136	-0.412
11	2-methyl-1-butanol	108.027	105.997	2.030	26.753	26.232	0.521

Table 3. Continued.

No.	Compound	MV(cm^3/mol)			MR(cm^3/mol)		
		Exp	Calcd	Res	Exp	Calcd	Res
12	3-pentanol	107.265	106.014	1.251	26.565	26.144	0.421
13	3-methyl-2-butanol	107.631	109.539	-1.908	26.638	26.864	-0.226
14	2-methyl-2-butanol	108.962	107.983	0.979	26.718	26.376	0.342
15	2,2-dimethyl-1-propanol	108.559	107.782	0.777			
16	1-Hexanol	125.590	125.203	0.387	31.636	31.639	-0.003
17	2-methyl-1-pentanol	123.795	123.610	0.185	31.262	31.134	0.128
18	2-ethyl-1-butanol	122.401	120.535	1.866	31.130	30.386	0.744
19	4-methyl-2-pentanol	126.774	127.228	-0.454	31.497	31.761	-0.264
20	2,3-dimethyl-2-butanol	124.065	125.792	-1.727	31.239	31.203	0.036
21	3,3-dimethyl-1-butanol	124.005	120.550	3.455	31.224	30.177	1.047
22	3,3-dimethyl-2-butanol	124.838	123.055	1.783	31.268	30.524	0.744
23	3-hexanol	124.716	124.013	0.703	31.297	31.148	0.149
24	3-methyl-3-pentanol	123.391	120.684	2.707	31.134	30.073	1.061
25	1-heptanol	141.345	141.631	-0.286	36.015	36.229	-0.214
26	2-heptanol	142.176	142.919	-0.743	36.077	36.317	-0.240
27	3-heptanol	141.535	140.441	1.094	35.981	35.739	0.242
28	4-heptanol	142.002	142.012	-0.010	35.928	36.152	-0.224
29	2,4-dimethyl-3-pentanol	140.101	141.476	-1.375	35.794	35.742	0.052
30	1-Octanol	157.473	158.059	-0.586	40.679	40.819	-0.140
31	2-octanol	158.720	159.347	-0.627	40.668	40.907	-0.239
32	4-octanol	158.972	158.440	0.532	40.649	40.742	-0.093
33	2-ethyl-1-hexanol	156.357	154.576	1.781	40.514	39.878	0.636
34	2,2,4-trimethyl-1-pentanol	155.221	156.175	-0.954	40.097	39.927	0.170
35	3,5-dimethyl-1-hexanol	156.960	157.563	-0.603	40.135	40.431	-0.296
36	1-Nonanol	174.417	174.488	-0.071	45.266	45.410	-0.144
37	2,6-dimethyl-4-heptanol	177.638	176.344	1.294	45.244	45.402	-0.158
38	5-nonanol	172.642	174.869	-2.227	44.589	45.333	-0.744
39	1-decanol	190.252	190.916	-0.664	49.734	50.000	-0.266
40	1-undecanol	207.652	207.344	0.308	54.640	54.591	0.049
41	2,6,8-trimethyl-4-nonanol	227.438	224.792	2.646	59.289	58.855	0.434
42	1-tridecanol	236.965	240.201	-3.236	63.375	63.771	-0.396

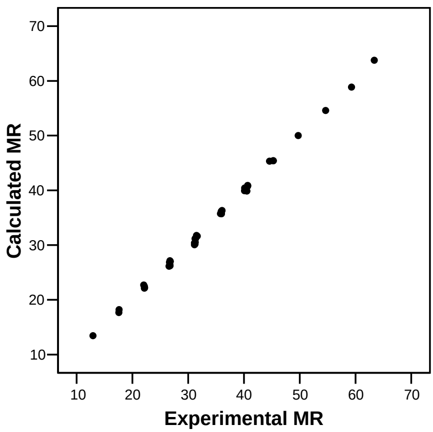

Figure 3. A plot of calculated versus experimental MRs for 41 alcohols.

4.4 *Correlations to log(1/S)*

Aqueous solubility is another important property of organic compounds, and it is widely applied in the fields of pharmaceutical chemistry, biological chemistry, and environmental science. It is also valuable to understand drug transport in organism and the influences for environment. The experimental values as log(1/S), where S is the solubility in moles per liter water, are listed in Table 4 for 63 alcohols. A model for these compounds is generated using the edge connectivity index ${}^{m}F$ and the class parameter of alcohol δ. When the edge connectivity index ${}^{0}F$ and ${}^{1}F$ together are included, the correlation coefficient r is 0.9876. If the alcohol-type parameter δ as the third parameter is introduced, the results can be improved significantly (r=0.9924, s=0.1294). The model is:

$$\log(1/S) = -2.0724(\pm 0.1031) + 0.3665(\pm 0.0489)\,{}^{0}F + 0.6515(\pm 0.0717)\,{}^{1}F - 0.1601(\pm 0.0261)\delta$$

$$r = 0.9924,\ s = 0.1294,\ F = 1284,\ N = 63 \qquad (9)$$

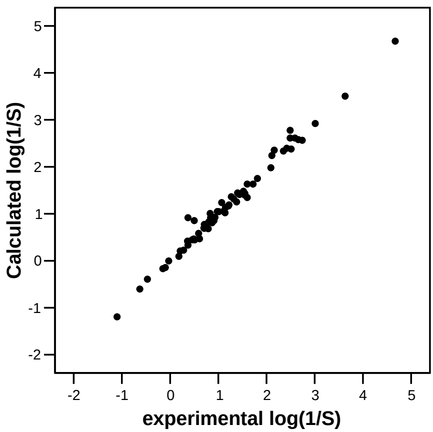

Figure 4. A plot of calculated versus experimental log(1/S) for 63 alcohols

The t-test is also used, and the results reveal that the variables in Eq.9 are all significant. This model explains more than 98.4% of variances in the experimental data of log(1/S) for these compounds. The calculated values and residuals for 63 alcohols are shown in Table 4. A plot of calculated versus experimental data is show in Fig. 4. From this figure we can see that the calculated values are very close to the experimental data. This model is also excellent.

4.5 ***Correlations to logP***

The octanol/water partition, logP, is very important for pharmaceutical compounds, and it has often been used to represent molecular lipophilicity, which seems to be a key factor related to many other biological events[37]. Thus, logP is a crucial parameter in QSAR/QSPR studies and drug design. The experiment data of logP for 62 alcohols are shown in Table 5. The model for these compounds is derived by using the parameters, ^{m}F and the δ. Only using the index ^{1}F, a good model can be obtained for logP, and the correlation coefficient r is 0.9934. If index ^{0}F is also included, the model can be obviously improved (r = 0.9967). If we take δ as the third parameter, the result is improved further (r=0.9973, s=0.1231). The model is:

$$\log P=-1.0418(\pm 0.0934)+0.2544(\pm 0.0492)\,^{0}F+0.6996(\pm 0.0724)\,^{1}F-0.0956(\pm 0.0262)\,\delta$$

$$r = 0.9973,\ s = 0.1231,\ F = 3551,\ N = 62 \qquad (10)$$

Table 4. Calculated and Experimental Water Solubility log(1/S) of 63 Alcohols

No	Compound	Log(1/S)			No	Compound	Log(1/S)		
		Exp	Calcd	Res			Exp	Calcd	Res
1	1-Ethanol	-1.10	-1.19	0.09	33	5-methyl-2-hexanol	1.38	1.25	0.13
2	1-Propanol	-0.63	-0.60	-0.03	34	2-methyl-3-hexanol	1.32	1.32	0.00
3	1-Butanol	-0.03	0.00	-0.03	35	2-methyl-2-hexanol	1.07	1.24	-0.17
4	2-methyl-1-propanol	-0.10	-0.14	0.04	36	2,2-dimethyl-1-pentanol	1.52	1.48	0.04
5	2-butanol	-0.47	-0.39	-0.08	37	4,4-dimethyl-1-pentanol	1.55	1.43	0.12
6	1-Pentanol	0.59	0.58	0.01	38	2,4-dimethyl-1-pentanol	1.60	1.35	0.25
7	3-methyl-1-butanol	0.51	0.44	0.07	39	3-methyl-3-hexanol	0.98	1.05	-0.07
8	2-pentanol	0.28	0.23	0.05	40	2,4-dimethyl-2-pentanol	0.93	0.92	0.01
9	2-methyl-1-butanol	0.46	0.45	0.01	41	2,4-dimethyl-3-pentanol	1.22	1.19	0.03
10	3-pentanol	0.21	0.21	0.00	42	3-ethyl-3-pentanol	0.83	1.01	-0.18
11	3-methyl-2-butanol	0.18	0.09	0.09	43	2,3-dimethyl-2-pentanol	0.87	0.93	-0.06
12	2-methyl-2-butanol	-0.15	-0.17	0.02	44	2,3-dimethyl-3-pentanol	0.84	0.92	-0.08
13	1-Hexanol	1.21	1.17	0.04	45	2,2-dimethyl-3-pentanol	1.15	1.13	0.02
14	4-methyl-1-pentanol	1.14	1.02	0.12	46	2,2,3-trimethyl-3-pentanol	1.27	1.36	-0.09
15	2-hexanol	0.87	0.81	0.06	47	2,3,3-trimethyl-2-butanol	0.71	0.77	-0.06
16	2-methyl-1-pentanol	1.11	1.06	0.05	48	1-Octanol	2.35	2.34	0.01
17	3-hexanol	0.80	0.83	-0.03	49	2-octanol	2.09	1.98	0.11
18	2-ethyl-1-butanol	1.01	1.05	-0.04	50	2-ethyl-1-hexanol	2.11	2.24	-0.13
19	4-methyl-2-pentanol	0.79	0.68	0.11	51	2-methyl-2-heptanol	1.72	1.63	0.09
20	3,3-dimethyl-1-butanol	0.50	0.85	-0.35	52	3-methyl-3-heptanol	1.60	1.63	-0.03
21	2,3-dimethyl-1-butanol	0.37	0.92	-0.55	53	1-Nonanol	3.01	2.92	0.09
22	2-methyl-2-pentanol	0.49	0.46	0.03	54	7-methyl-1-octanol	2.49	2.78	-0.29
23	3-methyl-2-pentanol	0.71	0.69	0.02	55	2-nonanol	2.74	2.57	0.17
24	2-methyl-3-pentanol	0.70	0.70	0.00	56	3-nonanol	2.66	2.58	0.08
25	2,2-dimethyl-1-butanol	0.91	0.85	0.06	57	4-nonanol	2.59	2.61	-0.02
26	3-methyl-3-pentanol	0.36	0.42	-0.06	58	5-nonanol	2.49	2.61	-0.12
27	3,3-dimethyl-2-butanol	0.61	0.47	0.14	59	2,6-dimethyl-4-heptanol	2.16	2.36	-0.20
28	2,3-dimethyl-2-butanol	0.37	0.33	0.04	60	3,5-dimethyl-4-heptanol	2.51	2.38	0.13
29	1-heptanol	1.81	1.75	0.06	61	2,2-diethyl-1-pentanol	2.42	2.40	0.02
30	2-heptanol	1.55	1.40	0.15	62	1-decanol	3.63	3.51	0.12
31	3-heptanol	1.44	1.41	0.03	63	1-dodecanol	4.67	4.68	-0.01
32	4-heptanol	1.40	1.44	-0.04					

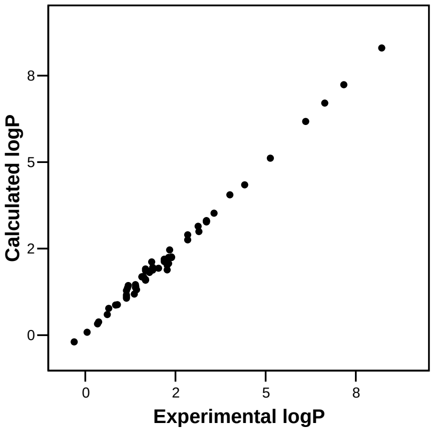

Figure 5. A plot of calculated versus experimental logP for 62 alcohols

Table 5. Calculated and Experimental Octanol/Water Partition (logP) of 62 Alcohols

No	Compound	logP exp	logP calcd	logP res	No	Compound	logP exp	logP calcd	logP res
1	1-Ethanol	-0.31	-0.20	-0.11	32	3-heptanol	2.31	2.21	0.10
2	1-Propanol	0.34	0.33	0.01	33	4-heptanol	2.31	2.25	0.06
3	2-propanol	0.05	0.08	-0.03	34	5-methyl-2-hexanol	2.19	2.13	0.06
4	1-Butanol	0.84	0.87	-0.03	35	2-methyl-3-hexanol	2.19	2.16	0.03
5	2-methyl-1-propanol	0.65	0.77	-0.12	36	2-methyl-2-hexanol	1.84	2.12	-0.28
6	2-butanol	0.61	0.59	0.02	37	2,2-dimethyl-1-pentanol	2.39	2.26	0.13
7	2-methyl-2-propanol	0.37	0.38	-0.01	38	4,4-dimethyl-1-pentanol	2.39	2.24	0.15
8	1-Pentanol	1.40	1.40	0.00	39	2,4-dimethyl-1-pentanol	2.19	2.19	0.00
9	3-methyl-1-butanol	1.42	1.31	0.11	40	3-methyl-3-hexanol	1.87	1.96	-0.09
10	2-pentanol	1.14	1.16	-0.02	41	2,4-dimethyl-2-pentanol	1.67	1.91	-0.24
11	2-methyl-1-butanol	1.14	1.29	-0.15	42	2,4-dimethyl-3-pentanol	2.31	2.07	0.24
12	3-pentanol	1.14	1.11	0.03	43	3-ethyl-3-pentanol	1.87	1.88	-0.01
13	3-methyl-2-butanol	1.14	1.07	0.07	44	2,3-dimethyl-2-pentanol	2.27	1.89	0.38
14	2,2-dimethyl-1-propanol	1.36	1.19	0.17	45	2,3-dimethyl-3-pentanol	1.67	1.86	-0.19
15	2-methyl-2-butanol	0.89	0.88	0.01	46	2,2-dimethyl-3-pentanol	2.27	2.02	0.25
16	1-Hexanol	2.03	1.93	0.10	47	1-Octanol	3.15	2.99	0.16
17	4-methyl-1-pentanol	1.78	1.84	-0.06	48	2-octanol	2.84	2.75	0.09
18	2-hexanol	1.61	1.69	-0.08	49	2-ethyl-1-hexanol	2.84	2.90	-0.06
19	2-methyl-1-pentanol	1.78	1.84	-0.06	50	1-Nonanol	3.57	3.52	0.05
20	3-hexanol	1.61	1.68	-0.07	51	2-nonanol	3.36	3.28	0.08
21	2-ethyl-1-butanol	1.78	1.81	-0.03	52	3-nonanol	3.36	3.27	0.09
22	4-methyl-2-pentanol	1.67	1.61	0.06	53	4-nonanol	3.36	3.31	0.05
23	3,3-dimethyl-1-butanol	1.57	1.68	-0.11	54	5-nonanol	3.36	3.31	0.05
24	2-methyl-2-pentanol	1.39	1.46	-0.07	55	2,6-dimethyl-4-heptanol	3.13	3.15	-0.02
25	3-methyl-2-pentanol	1.67	1.59	0.08	56	1-decanol	4.01	4.05	-0.04
26	2-methyl-3-pentanol	1.67	1.59	0.08	57	2-undecanol	4.42	4.34	0.08
27	2,2-dimethyl-1-butanol	1.57	1.68	-0.11	58	1-dodecanol	5.13	5.11	0.02
28	3-methyl-3-pentanol	1.39	1.38	0.01	59	1-tetradecanol	6.11	6.18	-0.07
29	3,3-dimethyl-2-butanol	1.19	1.43	-0.24	60	1-pentadecanol	6.64	6.71	-0.07
30	2,3-dimethyl-2-butanol	1.17	1.36	-0.19	61	1-hexadecanol	7.17	7.24	-0.07
31	1-heptanol	2.34	2.46	-0.12	62	1-octadecanol	8.22	8.30	-0.08

In equation (10), the variables are also significant verified by t-test. This model explains more than 99.4% of variances in the experimental data of logP for these compounds. Predicted results for 62 compounds are shown in Table 5. A plot of calculated versus experimental data is shown in Figure 5. Obviously, that the calculated values are very close to the experimental data. As yet, this model seems the one of the best models had been published.

4.6 *Model validation*

Finally, the above models generated for the five properties of the alcohols are verified by the cross-validation using leave-one-out method, and the correlation coefficients r_{cv} and standard deviations s_{cv} together with the normal r and s are shown in Table 6. This table reveals that the results of the cross-validations for each property are very close to the normal results of the models. This means that the models constructed in this work are stable.

Table 6. Statistics of MLR and Leave-One-Out Cross-Validation for the Five Final Models

properties	r	s	SEP	r_{cv}	s_{cv}	SEPcv
BP	0.9969	3.022	2.966	0.9967	3.101	3.078
MR	0.9993	1.504	1.431	0.9991	1.691	1.650
MV	0.9992	0.446	0.424	0.9990	0.487	0.475
log(1/S)	0.9924	0.129	0.125	0.9914	0.135	0.133
logP	0.9973	0.123	0.119	0.9969	0.130	0.127

5. CONCLUSIONS

A novel edge connectivity index ^{m}F based on the Estrada index ε is proposed for the predictions of the compounds containing heteroatom and/or multiple bonds in a molecular graph. The excellent QSPR models for the normal boiling point, molar volumes, molar refractions, water solubility and octanol/water partition can be constructed by using the ^{m}F index, the alcohol- type parameter δ, and the alcohol-distance parameter L of alcohols with up to 17 non-hydrogen atoms. For each of the five properties, all the correlation coefficients are great than 0.99. The results of the cross-validation verify that the models are statistically significant. It can be expected that the results for other applications by using these parameters will be good.

REFERENCE AND NOTES

(1) Trinajstic, N. *Chemical Graph Theory*, 2nd ed., CRC Press, Boca Raron,1992.

(2) Balaban, A. T. Applications of Graph in Chemistry. *J. Chem. Inf. Comput. Sci.* **1985**,25,334~343.

(3) Balaban, A. T. Chemical Graphs: looking Back and Glimpsing Ahead. *J. Chem. Inf. Comput. Sci*. **1995**,35,339~350.

(4) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: New York, **1976**.

(5) Hosoya, H. Topological Index. A Proposed Quantity Characterizing the Topological Nature of Structural Isomers of Saturated Hydrocarbons.*Bull. Chem. Soc. Jpn*. **1971**, 44, 2332-2339

(6) Balaban, A. T. Highly Discriminating Distance-Based Topological Index. *Chem. Phys. Lett*. **1982**, 89, 399-404.

(7) Bonchev, D.; Trinajstic, N. Information Theory, Distance Matrix, and Molecular Branching. *J. Chem. Phys*. **1977**, 67, 4517-4533.

(8) Schultz, H. P. Topological Organic Chemistry, 1. Graph Theory and Topological Indices. *J. Chem. Inf. Comput. Sci.* **1989**, 29, 227-228.

(9) Wiener, H. Structural Determination of Paraffin Boiling Points. *J. Am.Chem. Soc*. **1947**, 69, 17-20.

(10)Hall, L. H.; Mohney, B.;Kier, L. B. The Electrotopological state: Structure information at atomic Level for Molecular Graphs. *J. Chem. Inf. Comput. Sci.***1991**, 31,76-82.

(11)Ren, B. Novel Atomic-Level-Based AI Topological Descriptors: Application to QSPR/QSAR Modeling, *J. Chem. Inf. Comput. Sci.* **2002**, 42, 858-868

(12)Ren, B. Application of Novel Atom-Type AI Topological Indices to QSPR Studies of Alkanes. *Comput. Chem*. **2002**, 26, 357.

(13)Ren, B. Application of Novel Atom-Type AI Topological Indices in the Structure-Property Correlations. *J. Mol. Struct.* (THEOCHEM) **2002,** 586, 137-148.

(14)Ren, B. A New Topological Index for QSPR of Alkanes. *J. Chem. Inf. Comput. Sci.* **1999**, 39, 139-143.

(15)Ren. B. Novel Atom-type AI Indices for QSPR Studies of Alcohols *Comput. Chem*. **2002**, 26, 223-235

(16) Fernandez, F.M. Duchowitz, P.R. Castro E.A., About orthogonal descriptors in QSPR/QSAR, *MATCH Commun. Math. Comput. Chem*. **2004**, 51, 39-57

(17) Li, H., Lu, M. The m-connectivity index of graphs, *MATCH Commun. Math. Comput. Chem*. **2005**,54, 417-423

(18)Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure-Activity Studies*; Research Studies Press: Letchworth, **1986**.

(19)Bogdanov, B.; Nikolic, S.; Trinajstic, N. On the Three-Dimensional Wiener Number. *J. Math. Chem.* **1989**, 3, 299-309.

(20)Randic, M.; Jerman-Blazic, B.; Trinajstic, N. Development of 3-Dimensional Molecular Descriptors. *Comput. Chem.* **1990**, 14, 237-246.

(21)Estrada, E. Edge Adjacency Relationship in Molecular Graphs Containing Heteroatoms: A New Topological Index Related to Molar Volume, *J. Chem. Inf. Comput. Sci.*, **1995**, 35, 701-707.

(22)Toropov, A.; Toropova, A.; Ismailov, T.; Bonchev, D. 3D Weighting of Molecular Descriptors for QSPR/QSAR by the Method of Ideal Symmetry (MIS): 1. Application to Boiling Points of Alkanes. *J. Mol.Struct.* (THEOCHEM) **1998**, 424, 237-247.

(23)Diudea, M. V.; Horvath, D.; Graovac, A. Molecular Topology. 3D Distance Matrices and Related Topological Indices. *J. Chem. Inf.Comput. Sci.* **1995**, 35, 129-135.

(24)Mu L L.; Feng C. J.. Novel Connectivity Index of Edge Valence and its Application. *J. Chem. Ind. Eng.(*China), **2004**, 55(4), 531~540

(25)Estrada, E. Edge Adjacency Relationship and A novel Topological Index Related toMolecular Volume, *J. Chem. Inf. Comput. Sci.*, **1995,** 35, 31-33.

(26)Estrada, E., Rodriguez, L. Edge-connetivity indices in QSPR/QSAR studies. 1. comparison to other topological indices in QSPR studies. *J. Chem. Inf. Comut. Sci.***1999**, 39, 1037-1041.

(27)Estrada, E.,Guevara, N., Gutman, I. Extension of edge connectivity index. Relationships to line graph indices and QSPR applications. J. *Chem. Inf. Comput. Sci*.1998, 38, 428-431.

(28)Feng, C. Z., *Inorganic chemistry Teaching Reference Book,* Higher Education Press, Beijing, **1995**, P.95(in Chinese)

(29)Huang,F.; Liu X., *Alcohols, Encyclopedia of Chemical Industry*, Vol. 2, Chemical Industry Press, Beijing,**1991** in chinese.

(30)Yaws C. L., *Chemical Properties Handbook*, McGraw-Hill Beijing,**1999**.

(31)Mihalic Z.; Trinajstic N. A Graph-theoretical Approach to Structure-Property Relationships. *J. Chem. Edu.*, **1992**,69,701-712.

(32)Yang, Y.-Q.; Xu, L.; Hu, C.-Y. Extended Adjacency Matrix Indices and Their Applications. *J. Chem. Inf. Comput. Sci.* **1994**, 34, 1140-1145.

(33)Yao, Y.-Y.; Xu, L.; Yang, Y.-Q.; Yuan, X. S. Study on Structure-Activity Relationships of Organic Compounds. 3. New topological Indices and Their Applications. *J. Chem. Inf. Comput. Sci.* **1993**, 33, 590-595.

(34)Galvez, J.; Garcia, R.; Salabert, M. T.; Soler, R. Charge Indexes. New Topological Descriptors. *J. Chem. Inf. Comput. Sci.* **1994**, 34, 520-525.

(35)Kier, L. B.; Hall, L. H. *Molecular Structure Description*. *The Electrotopological State*; Academic Press: New York, **1999.**

(36)Estrada, E.; Molina, E. Novel Local (fragment-based) Topological Molecular Descriptors for QSPR/QSAR and Molecular Design. *J. Mol.Grap. Mod*. **2001**, 20, 54-64

(37)Klopman, G.; Li, J.-Y.; Wang, S.; Dimayuga, M. Computer Automated logP Calculations Based on an Extended Group Contribution Approach. J. *Chem. Inf. Comput. Sci.* **1994**, 34, 752-781.