CODEN IEJMAT *Internet Electronic Journal of Molecular Design* **2006**, *5*, 192–200 ISSN 1538–6414
BioChem Press http://www.biochempress.com

Inter*net* Electronic Journal of Molecular Design

April 2006, Volume 5, Number 4, Pages 192–200

Editor: Ovidiu Ivanciuc

Special issue dedicated to Professor Lemont B. Kier on the occasion of the 75th birthday

Boiling Points of Alcohols – A Comparative QSPR Study

Dušanka Janežič,[1] Bono Lučić,[2] Sonja Nikolić,[2] Ante Miličević,[3] and Nenad Trinajstić [2]

[1] National Institute of Chemistry, Hajdrihova 19, SI–1000 Ljubljana, Slovenia
[2] The Rugjer Bošković Institute, P.O.B. 180, HR–10002 Zagreb, Croatia
[3] The Institute of Medical Research and Occupational Health, P.O.B. 291, HR–10002 Zagreb, Croatia

Received: February 20, 2006; Accepted: March 15, 2006; Published: April 30, 2006

Citation of the article:
D. Janežič, B. Lučić, S. Nikolić, A. Miličević, and N. Trinajstić, Boiling Points of Alcohols – A Comparative QSPR Study, *Internet Electron. J. Mol. Des.* **2006**, *5*, 192–200, http://www.biochempress.com.

Inter*net* Electronic Journal of Molecular Design

BioChem Press
http://www.biochempress.com

Boiling Points of Alcohols – A Comparative QSPR Study

Dušanka Janežič,[1] Bono Lučić,[2] Sonja Nikolić,[2,]* Ante Miličević,[3] and Nenad Trinajstić [2]

[1] National Institute of Chemistry, Hajdrihova 19, SI–1000 Ljubljana, Slovenia
[2] The Rugjer Bošković Institute, P.O.B. 180, HR–10002 Zagreb, Croatia
[3] The Institute of Medical Research and Occupational Health, P.O.B. 291, HR–10002 Zagreb, Croatia

Received: February 20, 2006; Accepted: March 15, 2006; Published: April 30, 2006

Internet Electron. J. Mol. Des. **2006**, *5* (4), 192–200

Abstract

Motivation. The present report was motivated by recent papers (M. Randić, M. Pompe, D. Mills and S.C. Basak, Variable Connectivity Index as a Tool for Modeling Structure–Property Relationships, *Molecules* **2004**, *9*, 1177–1193; G. Krenkel, E. Castro and A. A. Toropov, Improved Molecular Descriptors Based on the Optimization of Correlation Weights of Local Graph Invariants, *J. Mol. Struct*. (*THEOCHEM*) **2001**, *542*, 107–113), and curiosity to see whether our CROMRsel modeling procedure leads to the QSPR model comparable to their very good models.

Method. We used the CROMRsel multivariate procedure that has been designed to select the best possible model among the set of models obtained for a given number of descriptors, the criterion being the standard error of estimate.

Results. The CROMRsel procedure reproduced the Randić–Pompe–Mills–Basak structure–boiling point model and produced a two–descriptor model close to the Krenkel–Castro–Toropov model.

Conclusions. The CROMRsel multivariate procedure is found to be competitive to the very good QSPR modeling schemes from the literature.

Keywords. Aliphatic alcohols; boiling points; CROMRsel procedure; molecular descriptors; molecular graph; topological indices; quantitative structure–property relationships; QSPR.

1 INTRODUCTION

This work was motivated by a recent paper of Randić *et al*. [1] in which the authors reported the calculation of normal boiling points for a series of 58 aliphatic alcohols using the variable vertex–connectivity index [2,3]. They have also differently weighted primary, secondary, tertiary and quarternary carbon atoms. The obtained standard error of estimate was only slightly improved (S = 3.9ºC) in comparison with the quantitative structure–property (QSPR) model based on equally

Dedicated on the occasion of the 75th birthday to Professor Lemont B. Kier.

* Correspondence author; E–mail: sonja@irb.hr.

weighted carbon atoms (S = 4.1ºC). There are also available other QSPR models using different descriptors for alcohols [*e.g.*, 4–6], the first being by Hall, Kier and Murray [7] who used the connectivity index in the original formulation [8]. Later Kier and Hall discussed in their book *Molecular Connectivity in Chemistry and Drug Research* [9] the QSPR models for boiling points of 28 aliphatic alcohols using the set of connectivity indices.

Here we decided to test our CROMRsel modeling procedure [10–17] against available QSPR models in the literature. CROMRsel is a multivariate procedure that has been designed to select the best possible model among the set of models obtained for a given number of descriptors, the criterion being the standard error of estimate. The quality of models is expressed by fitted (descriptive) statistical parameters: the correlation coefficient (R), the standard error of estimate (S, calculated using $N - I - 1$ in denominator, where N is the total number of molecules in data set, and I is the number of descriptors involved in the model) and Fisher's test (F). The models are also cross(internally)–validated by a leave–one–out procedure. Statistical parameters for the cross–validated models are symbolized by R_{cv} and S_{cv} (calculated using $N - I - 1$ in denominator), where subscript *cv* denotes the cross–validation.

2 COMPUTATIONAL DETAILS

In the present work we used 7 topological indices: ${}^1\chi$, ${}^1\chi^f(x,y)$, ${}^1\chi^f(x_1,x_2,x_3,x_4,y)$, ${}^1\chi^v$, ${}^w\varepsilon$, ${}^mM_2$ ${}^mEM_2$ and the parameter c_{OH}.

Randić *et al*. [1] used ${}^1\chi$, ${}^1\chi^f(x,y)$, ${}^1\chi^f(x_1,x_2,x_3,x_4,y)$, Hall, Kier and Hall [7] used ${}^1\chi$ and c_{OH} whilst Kier and Hall [9] used ${}^1\chi$ and ${}^1\chi^v$. So far ${}^w\varepsilon$, ${}^mM_2$ and ${}^mEM_2$ have not been used in this context though we have found them useful in QSPR modeling of different molecular properties [*e.g.*, 18–21]. Formulas we used to compute all considered descriptors are presented below. Since all these molecular descriptors have been derived using the concepts and terminology of chemical graph theory [22], we will do the same.

2.1 The Vertex–Connectivity Index ${}^1\chi$

The vertex–connectivity index ${}^1\chi$ was introduced by Randić in 1975 [8,23] and appears to be one the most used molecular descriptors in the structure–property–activity modeling [9,24–27]. It is defined as:

$$ {}^1\chi = \sum_{\text{edges}} [d(i)\, d(j)]^{-1/2} \qquad (1) $$

where $d(i)$ is the degree of a vertex i. The degree of a vertex of i of a (molecular) graph G is the number of edges incident to i [28].

2.2 The Variable Vertex–Connectivity Index $^1\chi(x,y)$

The variable vertex–connectivity index $^1\chi(x,y)$ has been introduced by Randić [2,3] and is constructed by using two variables x and y associated with atoms of different kind or different types. It is given by:

$$^1\chi(x,y) = \sum_{\text{edges}} \{[d(i) + x]\,[d(j) + y]\}^{-1/2} \tag{2}$$

where x and y are variables to be selected during the regression analysis. The numerical values of variables x and y vary from property to property. In the present study, we used the values of variables established by Randić *et al*. [1]: $x = 0.8$ for carbon atoms and $y = –0.9$ for oxygen atom.

2.3 The Variable Vertex–Connectivity Index $^1\chi(x_1, x_2, x_3, x_4\, y)$

Already Kier and Hall [9] pointed out that one is can obtain better QSPR/QSAR models by giving different weights to primary, secondary and tertiary carbon atoms. This idea was also explored by Krenkel *et al*. [6] with the inclusion of the weight for the quaternary carbon atoms, in addition to three weights for different atomic types (carbon: 0.287, hydrogen: 0.462 and oxygen: 1.000). Instead of using fixed weights, Randić *et al*. [1] used the variable weight procedure to get the weights x_1, x_2, x_3 and x_4 for primary, secondary, tertiary and quaternary carbon atoms. It this case, the variable vertex–connectivity index is given by:

$$^1\chi(x_k y) = \sum_{\text{edges}} \{[d(i) + x_k]\,[d(j) + y]\}^{-1/2} \tag{3}$$

where $k = 1, 2, 3, 4$. In this work, we used the values of variables determined by Randić *et al*. [1]: x_1, $x_2 = 0.80$ for primary and secondary carbon atoms, $x_3 = 0.96$ for tertiary carbon atom, $x_4 = 1.00$ for quaternary carbon atom and $y = –0.90$ for oxygen atom.

2.4 The Valence Vertex–Connectivity Indices $^1\chi^v$

The valence vertex–connectivity indices $^1\chi^v$ was introduced by Kier ad Hall [29,30] to account for the presence of heteroatoms in the molecule. This version of the connectivity index is defined as:

$$^1\chi^v = \sum_{\text{edges}} [\delta(i)\,\delta(j)]^{-1/2} \tag{4}$$

where $\delta(i)$ is the weight (valence–delta value) of a vertex i representing a given atom and is defined by

$$\delta(i) = [Z^v(i) - H(i)]/[Z(i) - Z^v(i) - 1] \tag{5}$$

where $Z^v(i)$ is the number of valence electrons belonging to the atom i, $Z(i)$ is its atomic number and $H(i)$ is the number of hydrogen atoms attached to the atom i. The delta values for primary, secondary, tertiary and quaternary carbon atoms are 4, 3, 2 and 1, and for oxygen in OH group is 5.

2.5 The Weighted Edge–Connectivity Index $^{w}\varepsilon$

The edge–connectivity index ε has been introduced by Estrada [31] and is defined as:

$$\varepsilon = \sum_{\text{edges}} [d(e_i)\, d(e_j)]^{-1/2} \tag{6}$$

where $d(e_i)$ is the degree of an edge e_i. The edge–connectivity index found use in the QSPR and QSAR [*e.g.*, 26,27,32–34].

The weighted edge–connectivity index $^{w}\varepsilon$ has also been introduced by Estrada [35] to account for heteroatoms in the molecule. It is defined as:

$$^{w}\varepsilon = \sum_{\text{edges}} [^{w}\delta(e_i)\, ^{w}\delta(e_j)]^{-1/2} \tag{7}$$

where $^{w}\delta(e_i)$ is the degree of an weighted edge e_i. The edge–weights of most common heteroatoms are also given by Estrada [35]. For the C–O bond the given weight is 0.8 and we have used it in our work.

2.6 The Modified the Second Zagreb Index $^{m}M_2$

The Zagreb indices have been introduced more than thirty years ago [36]. The second Zagreb index M_2 is defined as:

$$M_2 = \sum_{\text{edges}} d(i)\, d(j) \tag{8}$$

and has been modified recently [37]:

$$^{m}M_2 = \sum_{\text{edges}} [d(i)\, d(j)]^{-1} \tag{9}$$

2.7 The Modified Reformulated Second Zagreb Index $^{m}EM_2$

The Zagreb indices have been reformulated in terms of the edge–degrees [21] (the original formulation of the Zagreb indices has been based on the vertex–degrees [36]). The reformulated second Zagreb index is denoted by EM_2 and is defined as:

$$EM_2 = \sum_{\text{edges}} d(\mathrm{e}_i)\, d(\mathrm{e}_j) \tag{10}$$

The modified reformulated second Zagreb index $^{m}EM_2$ is given by [21]:

$$^{m}EM_2 = \sum_{\text{edges}} [d(\mathrm{e}_i)\, d(\mathrm{e}_j)]^{-1} \tag{11}$$

The Zagreb indices and their modifications found modest, but persistent, use in the structure–property modeling [*e.g.*, 20,24,26,27,37–40].

Table 1. The normal boiling points (in °C) of 58 aliphatic alcohols and their selected topological indices

Aliphatic alcohol	c_{OH}	$^{1}\chi$	$^{1}\chi^{f}(x,y)$	$^{1}\chi^{f}(x_k,y)$ k=1,2,3,4	$^{v}\chi$	$^{w}\varepsilon$	$^{m}M_2$	$^{m}EM_2$	bp
Methanol	1.000	1.000	2.357	2.3570	0.447000	0.000	0.200	0.000	64.7
Ethanol	0.707	1.414214	2.333	2.3353	1.023335	1.118	0.600	1.250	78.3
1–Propanol	0.707	1.914214	2.6924	2.6924	1.523335	1.491	0.8500	1.111	97.2
2–Propanol	0.577	1.732051	2.3869	2.3382	1.412899	1.161	0.73333	0.864	82.3
1–Butanol	0.707	2.414214	3.0495	3.0495	2.023335	1.980	1.100000	1.333	117.0
2–Butanol	0.577	2.270056	2.7566	2.7094	1.950904	1.993	1.066667	1.012	99.6
2–M–1–Propanol	0.707	2.270056	2.9611	2.9393	1.879177	1.943	0.93333	0.964	107.9
2–M–2–Propanol	0.500	2.000000	2.4640	2.4142	1.723607	2.106	0.80000	0.740	82.4
1–Pentanol	0.707	2.914214	3.4067	3.4067	2.523335	2.480	1.350000	1.583	137.8
2–Pentanol	0.577	2.770056	3.1137	3.0666	2.450904	2.525	1.316667	1.333	119.0
3–Pentanol	0.577	2.808060	3.1262	3.0806	2.488909	2.398	1.400000	1.199	115.3
2–M–1–Butanol	0.707	2.808060	3.3308	3.3104	2.417181	2.351	1.266667	1.155	128.7
3–M–1–Butanol	0.707	2.770056	3.3183	3.2964	2.379177	2.492	1.183333	1.324	131.2
2–M–2–Butanol	0.500	2.560660	2.8420	2.7936	2.284267	2.470	1.175000	0.904	102.0
3–M–2–Butanol	0.577	2.642735	3.0759	3.0131	2.323583	2.498	1.177778	1.069	111.5
2,2–MM–1–Propanol	0.707	2.560660	3.1832	3.1571	2.169781	2.402	0.97500	0.860	113.1
1–Hexanol	0.707	3.414214	3.7638	3.7638	3.023335	2.980	1.600000	1.833	157.0
2–Hexanol	0.577	3.270056	3.4709	3.4237	2.950904	3.025	1.566667	1.583	139.9
3–Hexanol	0.577	3.308060	3.4834	3.4377	2.988909	2.930	1.650000	1.520	135.4
2–M–1–Pentanol	0.707	3.308060	3.6879	3.6676	2.917181	2.889	1.516667	1.488	148.0
3–M–1–Pentanol	0.707	3.308060	3.6879	3.6676	2.917181	2.903	1.516667	1.519	152.4
4–M–1–Pentanol	0.707	3.270056	3.6754	3.6535	2.879177	2.997	1.433333	1.583	151.8
2–M–2–Pentanol	0.500	3.060660	3.1991	3.1507	2.784267	3.026	1.425000	1.273	121.4
3–M–2–Pentanol	0.577	3.180739	3.4021	3.3365	2.861588	2.917	1.511111	1.275	134.2
4–M–2–Pentanol	0.577	3.125898	3.3824	3.3134	2.806746	3.057	1.400000	1.357	131.7
2–M–3–Pentanol	0.577	3.180739	3.4021	3.3428	2.861588	2.915	1.511111	1.274	126.6
3–M–3–Pentanol	0.500	3.121320	3.2200	3.1729	2.844927	2.840	1.550000	1.078	122.4
2–E–1–Butanol	0.707	3.346065	3.7004	3.6816	2.955186	2.776	1.600000	1.373	146.5
2,2–MM–1–Butanol	0.707	3.121320	3.5612	3.5365	2.730441	2.772	1.350000	1.032	136.8
2,3–MM–1–Butanol	0.707	3.180739	3.6066	3.5663	2.789860	2.880	1.377778	1.25	149.0
3,3–MM–1–Butanol	0.707	3.060660	3.5404	3.5142	2.669781	2.984	1.225000	1.278	143.0
2,3–MM–2–Butanol	0.500	2.943376	3.1517	3.0825	2.666982	3.037	1.300000	1.067	118.6
3,3–MM–2–Butanol	0.577	2.943376	3.2593	3.1884	2.624224	2.981	1.233333	1.039	120.0
1–Heptanol	0.707	3.914214	4.1210	4.1210	3.523335	3.480	1.850000	2.083	176.3
3–Heptanol	0.577	3.808060	3.8405	3.7949	3.488909	3.430	1.900000	1.770	156.8
4–Heptanol	0.577	3.808060	3.8405	3.7949	3.488909	3.462	1.900000	1.842	155.0
2–M–2–Hexanol	0.500	3.560660	3.5563	3.5079	3.284267	3.526	1.675000	1.523	142.5
3–M–3–Hexanol	0.500	3.621320	3.5771	3.5301	3.344927	3.397	1.800000	1.446	142.4
3–E–3–Pentanol	0.500	3.681981	3.5980	3.5523	3.405587	3.217	1.925000	1.260	142.5
2,3–MM–2–Pentanol	0.500	3.481380	3.4923	3.4259	3.204987	3.428	1.633333	1.258	139.7
3,3–MM–2–Pentanol	0.577	3.504036	3.6373	3.5678	3.184885	3.356	1.608333	1.216	133.0
2,2–MM–3–Pentanol	0.577	3.481380	3.6290	3.5596	3.162229	3.406	1.566667	1.256	136.0
2,3–MM–3–Pentanol	0.500	3.504036	3.5007	3.4341	3.227643	3.377	1.675000	1.221	139.0
2,4–MM–3–Pentanol	0.577	3.553418	3.4148	3.3711	3.234267	3.440	1.622222	1.359	138.8
1–Octanol	0.707	4.414214	4.4781	4.4781	4.023335	3.980	2.100000	2.333	195.2
2–Octanol	0.577	4.270056	4.1852	4.1380	3.950904	4.025	2.066667	2.083	179.8
2–E–1–Hexanol	0.707	4.346065	4.4147	4.3959	3.955186	4.096	2.100000	2.238	184.6
2,2,3–MMM–3–pentanol	0.500	3.810660	3.7307	3.6686	3.534267	3.881	1.737500	1.234	152.2
1–Nonanol	0.707	4.914214	4.8353	4.8353	4.523335	4.480	2.350000	2.583	213.1
2–Nonanol	0.577	4.770056	4.5423	4.5360	4.450904	4.525	2.316667	2.333	198.5
3–Nonanol	0.577	4.808061	4.5548	4.5423	4.488909	4.430	2.400000	2.270	194.7
4–Nonanol	0.577	4.808061	4.5548	4.5423	4.488909	4.462	2.400000	2.342	193.0
5–Nonanol	0.577	4.808061	4.5548	4.5423	4.488909	4.462	2.400000	2.342	195.1
7–M–1–Octanol	0.707	4.770056	4.7468	4.7250	4.379177	4.497	2.183333	2.333	206.0
2,6–MM–4–Heptanol	0.577	4.519744	4.3779	4.3217	4.200593	4.525	2.066667	1.889	178.0
3,5–MM–4–Heptanol	0.577	4.629427	4.4173	4.4048	4.310276	4.278	2.288889	1.771	187.0
3,5,5–MMM–1–Hexanol	0.707	4.454507	4.5359	4.5097	4.063628	4.48	1.891667	1.852	193.0
1–Decanol	0.707	5.414214	5.1924	5.1924	5.023335	4.980	2.600000	2.833	230.2

2.8 Values for the Parameter c_{OH}

The numerical values for the parameter c_{OH} are obtained by formula:

$$c_{OH} = [d(i)]^{-1/2} \tag{12}$$

where $d(i)$ is the valency of the carbon atom i to which the OH group is attached – only the molecular skeleton is considered, the hydrogen atoms are neglected. The values of the parameter are: $c_{OH} = 0.70711$ for the carbon atom with valency 2, $c_{OH} = 0.57735$ for the carbon atom with valency 3 and $c_{OH} = 0.5$ for the carbon atom with valency 4. Consequently, the value of c_{OH} in methanol is 1, since the valency of its carbon atom is one.

3 RESULTS AND DISCUSSION

The values of the used topological indices, parameters c_{OH} and experimental normal boiling points of 58 aliphatic alcohols are given in Table 1. We recalculated all the topological indices employed here, although some of them are also available in the literature. This resulted in finding error in Table 3 of the paper by Randić *et al*. [1] – these authors switched the values of ${}^1\chi^f(x,y)$ and ${}^1\chi^f(x_1,x_2,x_3,x_4,y)$ indices for 3–M–1–butanol and 2–M–2–butanol, though in deriving the models they used correct values. The experimental values of boiling points are taken from Randić *et al*. [1].

We considered all models with one, two, three and four descriptors using the CROMRsel procedure. Below we give the best model in each case. Note that I is the number of descriptors.

3.1 The Best Single–Descriptor Model

$$\begin{gathered} \mathrm{bp} = -49.07\ (\pm 2.85) + 53.98\ (\pm 0.79)\ {}^1\chi^f(x_1,x_2,x_3,x_4,y) \\ N = 58 \quad R = 0.994 \quad R_{cv} = 0.994 \quad S = 3.9 \quad S_{cv} = 4.1 \quad \mathrm{F} = 4706 \end{gathered} \tag{13}$$

The model based on ${}^1\chi^f(\mathrm{x,y})$ possesses statistical characteristics that are practically identical to model (13):

$$\begin{gathered} \mathrm{bp} = -52.91\ (\pm 3.05) + 54.54\ (\pm 0.83)\ {}^1\chi^f(x,y) \\ N = 58 \quad R = 0.994 \quad R_{cv} = 0.993 \quad S = 4.1 \quad S_{cv} = 4.2 \quad \mathrm{F} = 4285 \end{gathered} \tag{14}$$

The model based on the vertex–connectivity index, that is, on the connectivity index that is not differentiating carbon and oxygen atoms, is poorer:

$$\begin{gathered} \mathrm{bp} = 16.97\ (\pm 4.34) + 37.52\ (\pm 1.24)\ {}^1\chi \\ N = 58 \quad R = 0.971 \quad R_{cv} = 0.969 \quad S = 8.7 \quad S_{cv} = 9.0 \quad \mathrm{F} = 911 \end{gathered} \tag{15}$$

All other models based on a single descriptor from the set given in Table 1 are rather very poor, the the standard errors of estimate S (S_{cv}) being in the range 10.4 (10.7) to 13.6 (14.3).

Randić *et al*. [1] reported the following regression model:

$$\begin{gathered} \mathrm{bp} = -49.003 + 53.964\ {}^1\chi^f(x_1,x_2,x_3,x_4,y) \\ N = 58 \quad R = 0.994 \quad S = 3.9 \end{gathered} \tag{16}$$

By removing outliers, these authors obtained much improved regression model:

$$\begin{gathered} \mathrm{bp} = -45.599 + 53.255\ {}^{1}\chi^{f}(x_1,x_2,x_3,x_4,y) \\ N = 53 \quad R = 0.997 \quad S = 2.6 \end{gathered} \tag{17}$$

Our models (13) and (14) reproduce results given in the paper by Randić *et al*. [1] for fitted statistical parameters – these authors did not cross–validated their models.

Krenkel *et al*. [6] have split the set of 58 alcohols (see Table 3 in their paper) into two equal subsets: a training set and a test set. We repeated their calculation for the training and test sets using linear regression and have obtained (slightly different) statistical parameters for the training set: R = 0.99758 (against 0.9953), S = 2.9752 (against 2.903) and for the test set: R = 0.99512 (against 0.9948), S = 3.05905 (against 3.025). The obtained results are very good indeed, but it should be mentioned that Krenkel *et al*. [6] used several parameters (given in Table 1 of their paper) in order to calculate descriptors used in their Eq. (11). Because of that, it is not strictly correct to say that their Eq. (11) is a single–descriptor model, though formally it looks so.

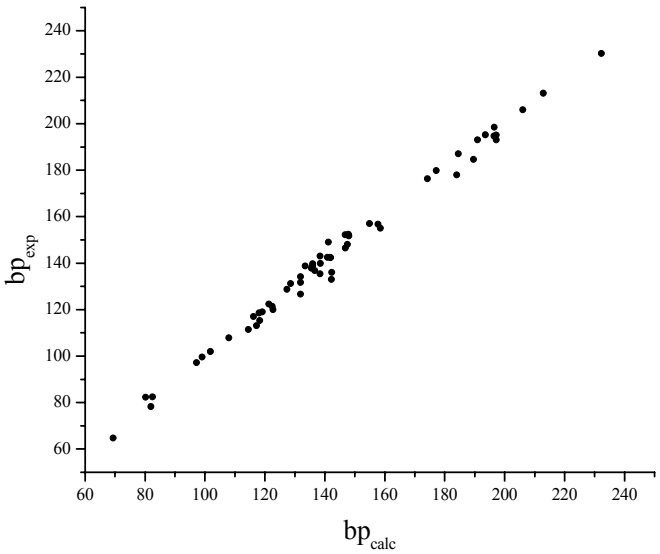

Figure 1. Scatter plot between $\mathrm{bp_{exp}}$ and $\mathrm{bp_{calc}}$ of 58 aliphatic alcohols. The values of $\mathrm{bp_{calc}}$ are obtained using Eq. (18).

3.2 The Best Two–Descriptor Model

$$\begin{gathered} \mathrm{bp} = -40.22\ (\pm 3.40) + 46.49\ (\pm 1.62)\ {}^{1}\chi^{f}(\mathrm{x,y}) + 10.97\ (\pm 2.01)\ {}^{m}EM_2 \\ N = 58 \quad R = 0.996 \quad R_{cv} = 0.995 \quad S = 3.3 \quad S_{cv} = 3.5 \quad \mathrm{F} = 3257 \end{gathered} \tag{18}$$

3.3 The Best Three–Descriptor Model

$$\begin{gathered} \mathrm{bp} = -39.09\ (\pm 4.26) + 45.66\ (\pm 2.46)\ {}^{1}\chi^{f}(\mathrm{x,y}) + 0.56\ (\pm 1.27)\ {}^{w}\varepsilon + 11.03\ (\pm 2.03)\ {}^{m}EM_2 \\ N = 58 \quad R = 0.996 \quad R_{cv} = 0.995 \quad S = 3.3 \quad S_{cv} = 3.7 \quad \mathrm{F} = 2140 \end{gathered} \tag{19}$$

This model does not present improvement over model (18), and one descriptor (${}^{w}\varepsilon$) is not statistically significant. The best four–descriptor model based on ${}^{1}\chi$, ${}^{1}\chi^{f}(x,y)$, ${}^{m}M_2$ and ${}^{m}EM_2$ posses identical statistical parameters as model (19), but the value of F (1619) which is poorer.

Since our models (18) and (19) possess practically identical statistical characteristics, following the Ockham Razor Principle (also called the Principle of Economy) [41], model (18) is our model of choice. In Figure 1, we give a scatter plot between the experimental (bp_{exp}) and calculated (bp_{calc}) boiling points for model (18).

4 CONCLUSIONS

A comparative study of several structure–boiling point models of aliphatic alcohols is carried out. The CROMRsel procedure reproduced the single–descriptor models of Randić *et al*. [1] and by cross–validating them it is shown to be the stable models. Krenkel *et al*. [6] produced very good model, but they employed a number of parameters to compute descriptor they used in building–up their model. We selected as our best model the two–descriptor model based on the variable connectivity index and the modified reformulated second Zagreb index. Judging by the standard errors of estimate, this model (S = 3.3) is comparable to the model by Krenkel *et al*. [6] (S = 3.0) and is somewhat better than the model by Randić *et al*. [1] (S = 3.9), and involves smaller number of optimized parameters.

Acknowledgment

This work was supported by Grant No. 0098034 from the Ministry of Science, Education and Sports of Croatia and by a grant from the Croatian–Slovenian joint research collaboration.

5 REFERENCES

[1] M. Randić, M. Pompe, D. Mills and S. C. Basak, Variable Connectivity Index as a Tool for Modeling Structure–Property Relationships, *Molecules* **2004**, *9*, 1177–1193.
[2] M. Randić, Novel Graph–Theoretical Approach to Heteroatoms in Quantitative Structure–Activity Relationships, *Chemom. Intel. Lab. Syst.* **1991**, *10*, 213–227.
[3] M. Randić, On Computation of Optimal Parameters for Multivariate Analysis of Structure–Property Relationship, *J. Comput. Chem.* **1991**, *12*, 970–980.
[4] M. Randić and M. Pompe, The Variable Connectivity Index $^{1}\chi^{f}$ *versus* the Traditional Molecular Descriptors: A Comparative Study of $^{1}\chi^{f}$ Against Descriptors in CODESSA, *J. Chem. Inf. Comput. Sci*. **2001**, *41*, 631–638.
[5] M. Randić and S. C. Basak, A New Descriptor for Structure–Property and Structure–Activity Correlations, *J. Chem. Inf. Comput. Sci*. **2001**, *41*, 650–656.
[6] G. Krenkel, E. Castro, and A. A. Toropov, Improved Molecular Descriptors Based on the Optimization of Correlation Weights of Local Graph Invariants, *J. Mol. Struct.* (*THEOCHEM*) **2001**, *542*, 107–113.
[7] L. H. Hall, L. B. Kier, and W.J. Murray, Molecular Connectivity. II. Relationship to Water Solubility and Boiling Point, *J. Pharm. Sci*. **1975**, *64*, 1974–1977.
[8] M. Randić, On Characterization of Molecular Branching, *J. Am. Chem. Soc*. **1975**, *97*, 6609–6615.
[9] L. B. Kier and L. H. Hall, *Molecular Connectivity in Chemistry and Drug Research*, Academic, New York, 1976, pp. 134–136.
[10] B. Lučić and N. Trinajstić, Multivariate Regression Outperforms Several Robust Architectures of Neural Networks in QSAR Modeling, *J. Chem. Inf. Comput. Sci* **1999**, *39*, 121–132.
[11] B. Lučić, N. Trinajstić, S. Sild, M. Karelson, and A. R. Katritzky, A New Efficient Approach for Variable Selection Based on Multiregression: Prediction of Gas Chromatographic Retention Times and Response Factors, *J. Chem. Inf. Comput. Sci*. **1999**, *39*, 610–621.
[12] D. Amić, D. Davidović–Amić, D. Bešlo, B. Lučić, and N. Trinajstić, Prediction of pK Values, Half–Lives and Electronic Spectra of Flavylium Salts From Molecular Structure, *J. Chem. Inf. Comput. Sci*. **1999**, *39*, 967–973.
[13] B. Lučić, D. Amić, and N. Trinajstić, Nonlinear Multivariate Regression Outperforms Several Concisely Designed

Neural Networks on Three QSPR Data Sets, *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 403–413.
[14] S. C. Basak, B. D. Gute, B. Lučić, S. Nikolić, and N. Trinajstić, A Comparative QSAR Study of Benzamidines Complement–Inhibitory Activity and Benzene Derivatives Acute Toxicity, *Comput. Chem.* **2000**, *24*, 181–191.
[15] T. Piližota, B. Lučić, and N. Trinajstić, Use of Variable Selection in Modeling the Secondary Structural Content of Proteins From Their Composition of Amino Acid Residues, *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 113–121.
[16] A. Miličević and S. Nikolić, On Variable Zagreb Indices, *Croat. Chem. Acta* **2004**, *77*, 97–101.
[17] S. Nikolić, A. Miličević and N. Trinajstić, Graphical Matrices in Chemistry, *Croat. Chem. Acta* **2005**, *78*, 241–250.
[18] S. Nikolić and N. Trinajstić, Modeling the Aqueous Solubility of Aliphatic Alcohols, *SAR QSAR Environ. Res.* **1998**, *9*, 117–126.
[19] S. Nikolić, N. Trinajstić, D. Amić, D. Bešlo and S. C. Basak, Modeling the Solubility of Aliphatic Alcohols in Water. Graph Connectivity Indices *versus* Line Graph Connectivity Indices; in: *QSPR/QSAR Studies by Molecular Descriptors*, Ed. M. V. Diudea, Nova, Huntington, N.Y., 2001, pp. 63–81.
[20] S. Nikolić, G. Kovačević, A. Miličević, and N. Trinajstić, The Zagreb Indices 30 Years After, *Croat. Chem. Acta* **2003**, *76*, 113–124.
[21] A. Miličević, S. Nikolić, N. Trinajstić, On Reformulated Zagreb Indices, *Mol. Diversity* **2004**, *8*, 393–399.
[22] N. Trinajstić, *Chemical Graph Theory*, 2nd revised ed., CRC, Boca Raton, FL, 1992.
[23] M. Randić, The Connectivity Index 25 Years After, *J. Mol. Graphics Modell.* **2001**, *20*, 19–35.
[24] A. Sabljić and N. Trinajstić, Quantitative Structure–Activity Relationships: The Role of Topological Indices, *Acta Pharm. Jugosl.* **1981**, *31*, 189–214.
[25] L. B. Kier and L. H. Hall, *Molecular Connectivity in Structure–Property Analysis*, Research Studies Press/Wiley, Letchworth, Herts, 1986.
[26] J. Devillers and A. T. Balaban, Editors, *Topological Indices and Related Descriptors in QSAR and QSPR*, Gordon & Breach, Amsterdam, 1999.
[27] R. Todeschini and V. Consonni, *Handbook of Molecular Descriptors*, Wiley–VCH, Weinheim, 2000.
[28] R. J. Wilson, *Introduction to Graph Theory*, Oliver & Boyd, Edinburgh, 1972, p. 12.
[29] L. B. Kier and L. H. Hall, Derivation and Significance of Valence Molecular Connectivity *J. Pharm. Sci.* **1981**, *70*, 583–589.
[30] L. B. Kier and L. H. Hall, General Definition of Valence Delta–Values for Molecular Connectivity, *J. Pharm. Sci.* **1983**, *72*, 1170–1173.
[31] E. Estrada, Edge Adjacency Relationships and a Novel Topological Index Related to Molecular Volume, *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 31–33.
[32] S. Nikolić, N. Trinajstić and I. Baučić, Comparison between the Vertex– and Edge–Connectivity Indices for Benzenoid Hydrocarbons, *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 42–46.
[33] D. Plavšić, N. Trinajstić, D. Amić, and M. Šoškić, Comparison between the Structure–Boiling Point Relationships with Different Descriptors for Condensed Benzenoids, *New. J. Chem.* **1998**, *22*, 1075–1078.
[34] S. Nikolić, N. Trinajstić, and S. Ivaniš, The Connectivity Indices of Regular Graphs, *Croat. Chem. Acta* **1999**, *72*, 875–883.
[35] E. Estrada, Edge Adjacency Relationships in Molecular Graphs Containing Heteroatoms: A New Topological Index Related to Molecular Volume, *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 701–707.
[36] I. Gutman and N. Trinajstić, Graph Theory and Molecular Orbitals. Total π–Electron Energy of Alternant Hydrocarbons, *Chem. Phys. Lett.* **1972**, *17*, 535–538.
[37] I. Gutman, B. Ruščić, N. Trinajstić, and C. F. Wilcox, Jr., Graph Theory and Molecular Orbitals. XII. Acyclic Polyenes, *J. Chem. Phys.* **1975**, *62*, 3399–3405.
[38] D. Bonchev and N. Trinajstić, Overall Molecular Descriptors. 3. Overall Zagreb Indices, *SAR QSAR Environ. Res.* **2001**, *12*, 213–236.
[39] E. Estrada, G. Patlewicz, and E. Uriarte, From Molecular Graphs to Drugs: A Review on the Use of Topological Indcies in Drug Design and Discovery, *Indian J. Chem.* **2003**, *42A*, 1315–1329.
[40] D. Vukičević and N. Trinajstić, On the Discriminatory Power of the Zagreb Indices for Molecular Graphs, *MATCH Commun. Math. Comput. Chem.* **2005**, *53*, 111–138.
[41] R. Hoffmann, V. I. Minkin, and B. K. Carpenter, Ockham's Razor and Chemistry, *Bull. Soc. Chim. Fr.* **1996**, *133*, 117-130.