ACADEMIA ROMÂNĂ
Revue Roumaine de Chimie
http://web.icf.ro/rrch/

Rev. Roum. Chim.,
2019, *64*(2), 183-189
Doi: 10.33224/rrch.2019.64.2.08

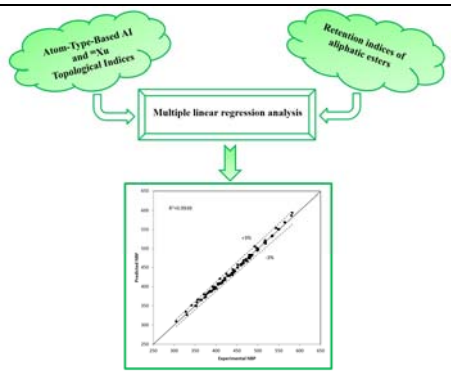
QSPR STUDY ON THE BOILING POINTS OF ALIPHATIC ESTERS USING THE ATOM-TYPE-BASED AI TOPOLOGICAL INDICES

Bahare OSAGHI and Fariba SAFA*

Department of Chemistry, Rasht Branch, Islamic Azad University, Rasht, I. R. Iran

Received April 1, 2018

In this work, normal boiling points (NBPs) for a group of aliphatic esters were modeled using a combination of the modified Xu (^{m}Xu) and the atom-type-based AI topological indice. The multiple linear regression model consisting of ^{m}Xu, AI(–CH_3) and AI(–O–) showed the squared correlation coefficient, Fisher ratio and standard error values of 0.994, 6705.7 and 4.54, respectively. Statistical validity of the model was verified by the external validation technique. Based on the results, fraction contribution of the topological indices entered the model decreased in the order of ^{m}Xu > AI(–CH_3) > AI(–O–) showing that NBPs of aliphatic esters are mainly dominated by molecular size, and degree of branching and polarity of the molecules have smaller contributions to the normal boiling points.



INTRODUCTION

Quantitative structure–property relationship (QSPR) modeling is one of the most effective approaches for estimation of a variety of physicochemical[1-6] and toxicological properties[7-9] which can provide significant information on the molecular features determining the properties of chemical compounds. The first step in a QSPR study is quantifying chemical structure of the molecules by numerical codes named descriptors that can show structural similarity and diversity of the molecules. Among the descriptors, topological indices have found major popularity in QSPR studies since they can be directly derived from the structure of the molecules on the basis of graph theory without any experimental effort. However, the conventional topological indices such as the well-known molecular connectivity index[10] characterize a molecule as a whole (*i.e.*, molecular size or shape). Additionally, the indices give no information on compounds with multiple bonds and heteroatoms. Unlike the conventional topological indices, the atomic level topological indices describe the structural information of a molecule at the atomic leve and code the structural environment of each atom type in a molecule. Kier *et al.* first introduced the concept of atom-type-based topological index, *i.e.* the so-called electrotopological state index.[11] Ren proposed a set of atomic-based AI topological indices that along with mXu index have been successfully used in developing high quality models for estimation of the physicochemical properties such as normal boiling points, molar refractions, van der Waals' constants, Pitzer's acentric factors, etc.[12-16] In a previous paper, our group reported QSPR study of standard formation enthalpies of acyclic alkanes using the AI topological indices.[17] Nevertheless, the use of AI indices in prediction of NBPs of esters has been largely neglected.

* Corresponding author: Safa@iaurasht.ac.ir; Safa_f@yahoo.com

The main goal of the present work was to further verify the potential of the atom-type-based AI topological indices as effective descriptors in prediction of NBPs of aliphatic esters. Moreover, the influence of different structural features and atomic groups on the normal boiling points of the model compounds is illustrated. As far as we are aware, this is the first report on QSPR study of aliphatic esters in which the atom-type-based AI topological indices are employed.

EXPERIMENTAL

Data set and regression analysis

Experimental data of NBPs for some aliphatic esters including C2-C16 linear and branched molecules were taken from the report of Brown and Stein[18] One compound was not modeled in this study, as its structure prohibited accurate modeling of structure-boiling point relationship. Table 1 presents model compounds and the experimental NBP values in the range of 304.70 to 582.00 K.

Table 1

Experimental and predicted normal boiling points of aliphatic esters

No.	Compound	NBP_{exp} (K)	NBP_{pred} (K)	No.	Compound	NBP_{exp} (K)	NBP_{pred} (K)
1	Methyl formate	304.70	309.85	44	pentyl propionate	441.80	437.15
2	Methyl acetate	330.00	326.94	45	propyl pentanoate	440.70	435.56
3	Ethyl formate	327.50	334.83	46	propyl 3-methylbutanoate	429.00	428.63
4	n-propyl formate	354.10	359.08	47	methyl 4-methylhexanoate	439.00	430.79
5	Isopropyl formate	341.40	351.62	48	tert-butyl butanoate	409.40	420.32
6	Methyl propionate	353.00	349.48	49	2-ethylbutyl acetate	432.00	427.78
7	Ethyl propionate	372.20	371.67	50	1,3-dimethylbutyl acetate	420.70	422.83
8	Isopropyl acetate	362.00	366.31	51	methyl n-octanoate	466.10	460.93
9	Methyl 2-methylpropionate	365.00	365.06	52	ethyl heptanoate	461.70	457.88
10	Butyl formate	379.70	382.65	53	3-methylbutyl butanoate	452.20	449.16
11	Ethyl butanoate	394.20	393.52	54	heptyl acetate	464.00	462.41
12	Butyl acetate	399.10	396.57	55	pentyl butanoate	451.75	456.92
13	ethyl 2-methylpropionate	385.70	386.39	56	butyl pentenoate	460.00	455.74
14	2-methylpropyl acetate	390.40	389.30	57	2-methylpropyl 3-methylbutanoate	444.35	441.66
15	methyl pentanoate	401.20	395.46	58	hexyl propionate	458.65	458.08
16	propyl propionate	395.70	393.89	59	octyl formate	471.00	471.06
17	methyl 3-methylbutanoate	389.70	388.15	60	1,4-dimethylpentyl acetate	440.00	444.76
18	Isopropyl propionate	383.00	386.54	61	ethyl n-octanoate	480.20	478.53
19	1,1-dimethylethyl acetate	370.70	380.05	62	1-octyl acetate	483.20	481.74
20	3-methylbutyl formate	397.40	398.68	63	3-methylbutyl 3-methylbutanoate	465.85	462.24
21	1-methylpropyl acetate	385.40	388.87	64	methyl nonanoate	486.70	481.75
22	methyl pivalate	375.05	378.84	65	2-ethylhexyl acetate	472.20	466.73
23	methyl n-hexanoate	422.20	417.81	66	butyl hexanoate	480.89	474.17
24	n-pentyl acetate	420.70	418.66	67	hexyl 2-methylpropionate	472.00	470.05
25	ethyl 3-methylbutanoate	407.90	408.24	68	heptyl propionate	483.19	478.57
26	ethyl n-pentanoate	415.90	415.32	69	pentyl pentanoate	479.00	475.70
27	2-propyl butanoate	403.70	407.59	70	methyl decanoate	497.20	502.11
28	n-propyl butanoate	416.00	414.69	71	ethyl nonanoate	500.15	498.77
29	2-methylbutyl acetate	414.00	409.17	72	octyl propionate	501.00	498.64
30	2-methylpropyl propionate	410.00	408.54	73	nonyl acetate	497.00	501.85
31	n-hexyl formate	427.20	427.91	74	propyl octanoate	499.58	496.61
32	1-methylbuthyl acetate	408.00	408.60	75	heptyl butanoate	499.00	496.54

Table 1 (continued)

33	1-methylpropyl propionate	406.20	406.29	76	butyl heptanoate	499.30	495.41
34	1,1-dimethylpropyl acetate	396.65	399.42	77	hexyl pentanoate	499.50	495.37
35	propyl 2-methylpropionate	412.00	407.87	78	1,3,5-trimethylhexyl acetate	474.65	466.01
36	2-pentyl acetate	408.00	408.60	79	ethyl decanoate	518.20	518.63
37	1,1-dimethylethyl propionate	391.50	399.79	80	hexyl hexanoate	518.58	513.46
38	3-pentyl acetate	407.00	405.75	81	decyl acetate	517.15	521.57
39	n-hexyl acetate	442.20	440.02	82	octyl butanoate	515.40	516.10
40	n-butyl butanoate	437.70	435.67	83	heptyl pentanoate	518.36	514.75
41	methyl heptanoate	445.20	439.58	84	1-methylheptyl butanoate	492.00	505.87
42	2-methylpropyl butanoate	430.00	428.57	85	heptyl hexanoate	534.12	532.96
43	3-methylbutyl propionate	433.90	429.99	86	nonyl butanoate	535.00	535.04

No.	Compound	NBP_{exp} (K)	NBP_{pred} (K)	No.	Compound	NBP_{exp} (K)	NBP_{pred} (K)
87	pentyl octanoate	533.36	533.05	104	ethyl hexanoate	441.20	436.80
88	hexyl heptanoate	534.04	532.72	105	2-methylpropyl 2-methylpropionate	420.70	421.69
89	decyl butanoate	543.00	554.22	106	1-methylpropyl butanoate	424.65	433.43
90	heptyl heptanoate	550.36	551.09	107	heptyl formate	451.27	449.72
91	hexyl octanoate	550.59	551.14	108	propyl hexanoate	458.70	456.23
92	heptyl octanoate	563.80	569.06	109	3-methylbutyl 2-methylpropionate	442.95	442.42
93	octyl heptanoate	563.90	569.24	110	*n*-hexyl *n*-butanoate	478.25	477.13
94	octyl octanoate	580.00	586.80	111	butyl hexanoate	477.50	474.17
95	ethyl acetate	350.20	350.59	112	3-methylbutyl pentanoate	466.00	468.50
96	methyl butanoate	375.40	372.60	113	propyl heptanoate	481.09	476.59
97	*tert*-butyl formate	355.70	366.25	114	pentyl hexanoate	498.00	495.01
98	methyl 2-methylbutanoate	388.20	387.49	115	methyl 2-butylhexanoate	479.15	482.95
99	n-pentyl formate	405.50	405.56	116	butyl octanoate	518.17	514.85
100	ethyl 2-methylbutanoate	406.20	406.20	117	octyl pentanoate	534.75	533.81
101	*n*-butyl propionate	418.70	415.75	118	*n*-octyl hexanoate	548.40	551.55
102	3-methylbutyl acetate	415.20	411.60	119	ethyl tetradecanoate	582.00	594.61
103	1-methylethyl 2-methyl propionate	393.90	399.77				

Linear regression analyses were performed using SPSS/PC software package.[19] Criteria for selection of the best multiple linear regression (MLR) model were the statistics of squared correlation coefficient (R^2), adjusted correlation coefficient (R^2_{adj}), Fisher-ratio (F) and standard error of estimate (SE). Statistical validity of the model was tested using external validation method.20,21 by dividing the entire data set into five subsets, leaving out one subset as prediction set, and regenerating the model coefficients for the training set composed of the other four subsets. Then, the standard errors of calibration (SEC) and prediction (SEP) were employed to evaluate quality of the models.

Descriptor generation

Topological indices for aliphatic esters were drived from the molecular graph G=[V,E], where V and E symbolize the atoms of a molecule and the chemical bonds, respectively. The indices was calculated using the distance matrix, $\mathbf{D}=[d_{ij}]_{nxn}$ and the vertex-adjacency matrix, $\mathbf{A}=[a_{ij}]_{nxn}$ in which the elements d_{ij} are the length of shortest path between vertex i and vertex j and the elements a_{ij} are 1, if atom si and j are adjacent and and 0 otherwise. Also,. The sums over row i or column j of **A** and **D** matrices yields local vertex-degree v_i and the distance sums s_i, respectively. The ^{m}Xu index for a molecule with n atoms is defined as:[12]

$$Xu = n^{1/2}\log\left(\sum_{i=1}^{n} v_i s_i^2 \Big/ \sum_{i=1}^{n} v_i s_i\right) \quad (1)$$

where the sum is over all i vertices in graph. For any atom i that belongs to jth atom type in the graph, corresponding AI index is defined as:[13]

$$AI_i(j) = 1 + \phi_i(j) = 1 + \left(v_i(j)\, s_i^2(j) \Big/ \sum_{i=1}^{n} v_i s_i \right) \qquad (2)$$

where $\phi_i(j)$ is a perturbing term reflecting the effect of the structural environment of the ith atom on its topological index value and the topological index for jth atom-type in a graph, AI(j), is the sum of all $AI_i(j)$ values of the same atom type in the graph. Moreover, for differentiation of the hetero atoms and multiple bonds, the degree of vertex, v^m [14] based on the valence connectivity of Kier–Hall[22] is used.

RESULTS AND DISCUSSION

Linear regression models for normal boiling points of aliphatic esters

In the first step of the study, simple linear regression model was developed for NBPs using ^mXu index alone. Specifications of the model found for the data set along with the statistical parameters are as follows:

$$NBP = 251.721 (\pm 1.787) + 39.416 (\pm 0.347)\, {}^mXu \qquad (3)$$

N=119 $R^2 = 0.991$ $R^2_{adj} = 0.991$ SE= 5.67 F= 12883.0

$$NBP = 254.219 (\pm 1.826) + 40.649 (\pm 0.543)\, {}^mXu - 1.381 (\pm 0.209) AI(-CH_3) + 1.380 (\pm 0.597) AI(-O-) \qquad (4)$$

It can be seen that ^mXu index shows good correlation with NBPs values judging from high values of R^2_{adj} and F and relatively low SE. However, to obtain a more accurate model and to account the contribution of atom types or groups to the NBPs of the esters, a MLR model, added AI topological indices, was generated. Equation 4 indicates the three-parameter MLR model developed for predicting NBPs of esters. Values of the topological descriptors entered the model are given in Table 2.

As can be found, combination of ^mXu and AI indices significantly improves quality of the model. The statistics indicate that the regression explained by the model is significant at 99% confidence level and the standard error is about 20% less than SE for the simple linear model. According to the model predictions for NBPs (Table 1), maximum relative error (2.99%) belonged to isopropyl formate. Fig. 1 demonstrates the correlation between the experimental and predicted NBPs indicating all the predicted values are within ±3% error band. The results evidently proves high efficiency of the topological indices in modeling NBPs of aliphatic esters.

Table 2

Topological indices employed in the MLR model developed for NBP of aliphatic esters

No.	^mXu	AI ($-CH_3$)	AI (-O-)	No.	^mXu	AI ($-CH_3$)	AI (-O-)	No.	^mXu	AI ($-CH_3$)	AI (-O-)
1	1.371	2.183	2.127	41	4.679	8.322	4.841	81	6.753	10.909	5.740
2	1.865	4.637	2.403	42	4.566	11.049	2.916	82	6.685	11.296	4.172
3	1.995	2.562	2.205	43	4.593	10.969	3.048	83	6.667	11.423	3.828
4	2.595	2.937	2.494	44	4.676	8.299	3.119	84	6.534	13.786	3.682
5	2.480	4.778	2.323	45	4.642	8.494	3.190	85	7.139	12.232	3.938
6	2.434	5.440	2.767	46	4.592	11.943	3.091	86	7.162	12.058	4.604
7	3.012	6.184	2.584	47	4.544	10.593	4.699	87	7.134	12.275	4.187
8	2.946	7.926	2.374	48	4.427	12.76	2.741	88	7.134	12.276	3.952
9	2.894	8.009	3.092	49	4.512	10.299	3.183	89	7.642	12.723	5.034
10	3.174	3.311	2.898	50	4.474	12.634	3.042	90	7.606	13.017	4.110
11	3.565	6.970	2.917	51	5.206	8.986	5.454	91	7.604	13.036	4.224
12	3.634	6.680	2.794	52	5.176	9.154	4.270	92	8.067	13.787	4.332
13	3.460	8.958	2.818	53	5.100	12.079	3.130	93	8.070	13.752	4.321
14	3.532	8.847	2.691	54	5.280	8.689	4.046	94	8.523	14.529	4.496
15	3.583	6.886	3.710	55	5.188	9.076	3.165	95	2.472	5.318	2.339

Table 2 (continued)

16	3.581	6.897	2.643	56	5.163	9.245	3.219	96	3.013	6.177	3.208
17	3.477	8.985	3.627	57	5.008	14.78	3.096	97	2.909	6.939	2.436
18	3.475	8.999	2.540	58	5.203	9.000	3.467	98	3.456	8.579	3.369
19	3.361	10.237	2.439	59	5.325	4.805	5.100	99	3.734	3.684	3.377
20	3.643	5.896	3.281	60	5.037	13.673	3.399	100	3.964	9.732	3.112
21	3.517	8.535	2.506	61	5.690	9.865	4.803	101	4.136	7.599	2.836
22	3.306	10.450	3.378	62	5.763	9.478	4.604	102	4.099	9.769	3.084
23	4.138	7.587	4.257	63	5.550	15.924	3.210	103	3.903	12.208	2.731
24	4.188	7.369	3.166	64	5.720	9.691	6.090	104	4.650	8.440	3.773
25	4.019	10.008	3.241	65	5.504	11.918	3.793	105	4.478	13.380	2.844
26	4.113	7.716	3.319	66	5.710	13.311	4.531	106	4.543	6.635	2.695
27	4.018	10.014	2.796	67	5.626	12.722	3.412	107	4.808	4.432	4.490
28	4.112	7.720	2.880	68	5.718	9.704	3.868	108	5.163	9.242	3.559
29	4.036	9.568	2.975	69	5.674	9.979	3.352	109	5.023	14.648	3.068
30	4.042	9.979	2.751	70	6.223	10.399	6.746	110	5.700	9.808	3.446
31	4.278	4.058	3.912	71	6.193	10.575	5.365	111	5.710	13.311	4.531
32	4.023	9.408	2.788	72	6.220	10.412	4.312	112	5.606	13.15	3.297
33	3.982	9.720	2.633	73	6.263	10.191	5.159	113	5.674	9.974	3.975
34	3.863	11.149	2.579	74	6.176	10.697	4.431	114	6.168	10.763	3.579
35	4.021	9.889	2.796	75	6.192	10.585	3.797	115	5.870	13.154	6.009
36	4.023	9.408	2.788	76	6.168	10.761	3.857	116	6.658	11.497	4.245
37	3.889	11.584	2.542	77	6.175	10.704	3.561	117	7.150	12.141	4.144
38	3.959	9.470	2.663	78	5.767	18.771	2.388	118	7.612	12.958	4.200
39	4.723	8.079	3.605	79	6.686	11.285	5.951	119	8.567	14.148	8.482
40	4.650	8.447	2.988	80	6.655	11.541	3.378				

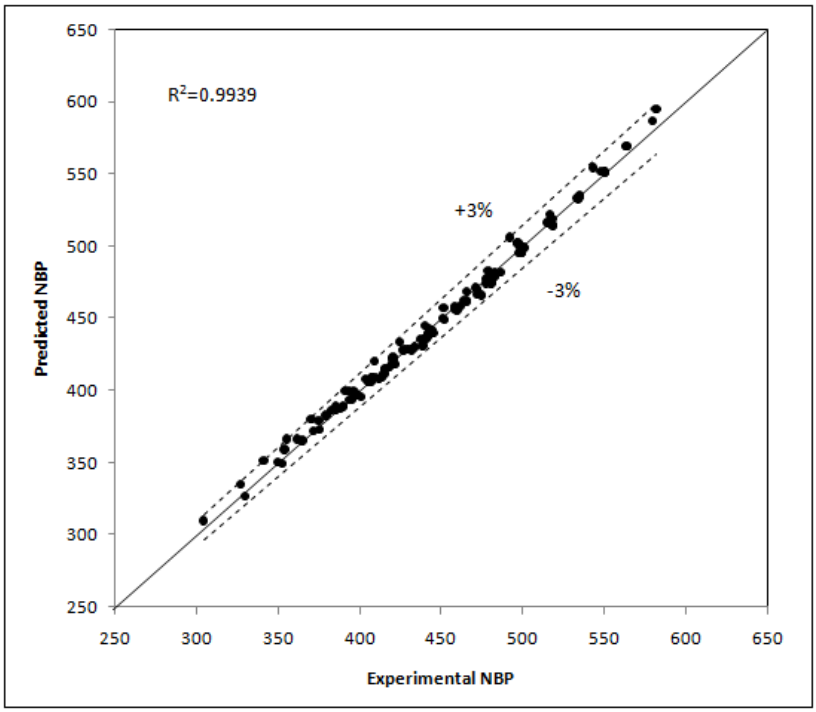


Fig. 1 – Predicted *versus* experimental NBPs for aliphatic esters based on the MLR model.

Table 3

Specifications of the MLR models based on the external validation test

Parameter	Training subsets				
	1	2	3	4	5
Constant	255.669	254.697	255.265	250.726	253.940
^{m}Xu	41.025	40.687	40.599	40.605	40.199
AI ($-CH_3$)	-1.520	-1.333	-1.483	-1.231	-1.313
AI (-O-)	0.768	1.054	-1.437	2.085	1.899
R^2_{adj}	0.994	0.994	0.994	0.995	0.994
F	4972.4	5062.3	5381.6	6479.2	5320.9
SEC	4.58	4.53	4.41	3.98	4.57
SEP	3.84	4.17	4.67	6.19	4.02
n_t / n_p [a]	95/24	95/24	95/24	95/24	96/23

[a] n_t and n_p are the number of molecules in the training and prediction sets, respectively.

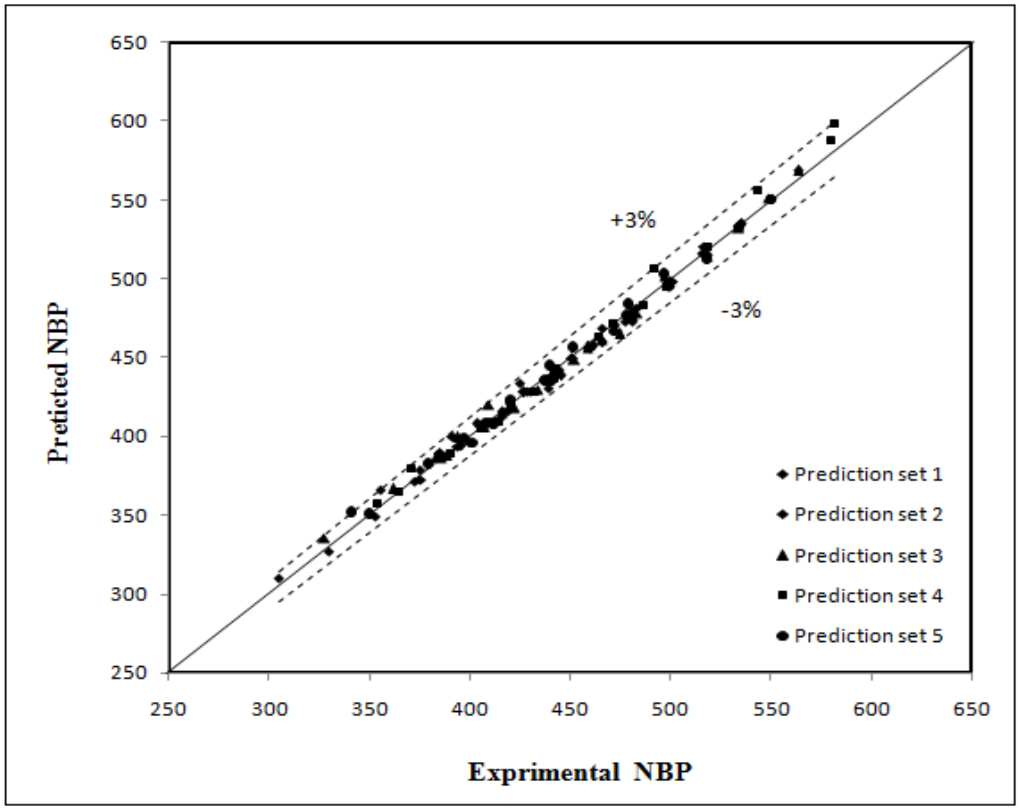


Fig. 2 – Predicted *versus* experimental NBPs for different prediction sets in external validation test.

Table 3 indicates the results of validation test of the MLR model. Average values of 4.58 for SEC and 4.41 for SEP show that the models developed for the training sets can satisfactorily predict NBPs for aliphatic esters not used for model generation. Good agreement between the predicted and experimental NBPs for different prediction sets can be easily observed in Fig. 2.

Structural interpretation of normal boiling points of aliphatic esters

To obtain insights into the role of the structural features of esters in determining their NBPs, the relative contribution (Ψ_r) and fraction contribution (Ψ_f) of the topological indices were calculated as follows:[23]

$$\Psi_r(i) = c_i TI_i \quad (5)$$

$$\Psi_f(\mathrm{i}) = \left[\mathrm{R}^2 \times \frac{|\Psi_r(i)|}{\sum_i |\Psi_r(i)|} \right] \times 100\% \quad (6)$$

where c_i and TI_i are the coefficient and the average value of the ith topological index entered in the MLR model and the sum is over all the topological indices. The results showed that ^{m}Xu index with Ψ_f value of 91.00% is the most important descriptor appeared in the model showing that bulkiness or size of the ester molecule plays a dominant role in determining its NBP because ^{m}Xu index characterizes the molecular size.[12] Relative contribution of +200.25 for this descriptor indicates that the larger the size of the molecule is, the greater its NBP. Fraction contribution values of AI(–CH_3) and AI(–O–) indices to the NBPs were 6.15% and 2.25%, respectively. As expected, the relative contribution value for AI(–CH_3) index was negative (Ψ_r=−13.52) because the AI(–CH_3) index is clearly related to the number of methyl groups which is a crude measure of branching.[23] Obviously, branching prevents close contact with neighboring molecules in space and reduces the interactions between them. Consequently, NBP decreases with increasing the AI(–CH_3) index. On the other hand, positive value of +4.96 for relative contribution of AI(–O–) index shows that higher AI(–O–) tends to increase the extent of NBP for the aliphatic ester. This is due to the fact that the atomic –O– group belongs to the carboxyl group of ester molecules and reflects their polar character and it is no doubt that polar interactions between molecules become more important with increasing AI(–O–) index.

CONCLUSION

In this study, an accurate three-parameter QSPR model for NBPs of a group of aliphatic esters was successfully developed using a combination of ^{m}Xu index and atom-type-based AI topological indices. The MLR model is very efficient and provides satisfactory results in both accuracy and stability for predicting normal boiling point values of aliphatic esters. The results indicate that NBPs of the aliphatic esters depend not only on the molecular size but also on various atomic group of the molecules. Contribution of the topological indices included in the model decreased in the order of ^{m}Xu> AI(–CH_3)> AI(–O–).The developed model allows prediction of NBPs for similar compounds using only the knowledge of two dimensional structure of the molecules.

REFERENCES

1. D. Ghaslani, Z. Eshagi Gorgi, A. Ebrahimipoor Gorgi and S. Riahi, *Chem. Eng. Res. Des.,* **2017**, *120*, 15- 25.
2. N. Basant and Sh. Gupta, *Atmos. Environ.*, **2018**, *177*, 166-174
3. I. Mehraein and S. Riahi, *J. Mol. Liq.,* **2017**, *225*, 521-530.
4. A. Weinebeck, S. Kaminski, H. Murrenhoff and K. Leonhard, *Tribol. Int.*, **2017**, *115*, 274-284.
5. T. Nejad Ghaffar Borhani, A. Afzali and M. Bagheri, *Process Saf. Environ. Prot.*, **2016**, *103*, 115–125.
6. P. Kumar and O. Kunel Roy, *Food Chem. Toxicol.*, **2018**, *112*, 551-562
7. W. Zhang, J. Wu, B. Li, H. Wu, L. Wang, J. Hao, S. Wu and Q. Zhou, *Org. Biomol. Chem.,* **2016**, *14*, 2840-2845.
8. N. Basant, Sh. Gupta and K.P. Singh, *Neurotoxicol.*, **2016**, *53*, 45-52 .
9. A. Can, *Toxicol. Lett.*, **2014**, *230*, 434-443.
10. L. B. Kier and L. H. Hall, "Molecular connectivity in chemistry and drug research", New York, Academic Press, 1976.
11. L. B. Kier, L. H. Hall and J. W. Frazer, *J. Math. Chem.*, **1991**, *7*, 229-241.
12. B. Ren, *J. Chem. Inf. Comput. Sci*., **1999**, *39*, 139-143.
13. B. Ren, *Comput. Chem*., **2002**, *26*, 223-235.
14. B. Ren, *Comput. Chem*., **2002**, *26*, 357-369.
15. B. Ren, *J. Chem. Inf. Comput. Sci*., **2003**, *43*, 1121-1131.
16. K. Panneerselvam, M. P. Antony, T. G. Srinivasan and P. R. V. Rao, *Thermochim. Acta*, **2010**, *511*, 107-111.
17. F. Safa and M. Yekta, *Arabian J. Chem*., **2017**, *10*, 439-447.
18. R. L. Brown and S. E. Stein, "Boiling Point Data in NIST Chemistry WebBook", Gaithersburg, National Institute of Standards and Technology, http://webbook. nist.gov, 2010.
19. SPSS version 20.0, IBM Inc, United State.
20. R. Kramer, "Chemometric techniques for quantitative analysis", New York, Marcel Dekker, 1998.
21. R.D.M.C. Amboni, B.S. Junkes, R.A. Yunes and V.E.F. Heinzen, *J. Mol. Struct.* (*Theochem.*), **2002**, 586, 71-80.
22. L.B. Kier and L. H. Hall, "Molecular connectivity in structure–activity studies", Letchworth, Research Studies Press, 1986.
23. D.E. Needham, I.C. Wei and P.G. Seybold, *J. Am. Chem. Soc*., **1988**, *110*, 4186-4194.