Asian Journal of Chemistry; Vol. 23, No. 11 (2011), 5079-5084

ASIAN JOURNAL OF CHEMISTRY

www.asianjournalofchemistry.co.in

Associative Neural Network and Multi Linear Regression Based Quantitative Structure Property Relationship for Modeling Viscosity of Alcohols

P. NEELAMEGAM[1,*] and S. KRISHNARAJ[2]

[1]SEEE, SASTRA University, Thirumalaisamudram, Thanjavur-613 401, India
[2]Department of Physics, PRIST University, Thanjavur-614 904, India

*Corresponding author: E-mail: neelkeer@yahoo.com

(*Received*: 4 February 2011; *Accepted*: 27 July 2011) AJC-10215

A quantitative structure property relationship (QSPR) study was conducted based on molecular descriptors derived from molecular structures have been used for the prediction of viscosity of alcohols at 20 ºC. To perform this research, a set of 35 alcohols as data series was selected then variable Zagreb index ($^{\lambda}M_2$), number of carbon atoms (N_c), molecular weight (MW) and number of hydroxyl groups (N_{OH}) for data series was calculated. Variable Zagreb index involving a parameter λ to be determined during regression fitting. The optimal value of the variable λ was determined by minimizing the standard error of regression. Multi linear regression (MLR) and associative neural network (ASNN) methods were used to construct the linear and non-linear prediction models respectively. The results were cross-validated by leave-one-out (LOO) procedure. The predicted results are in good agreement with the experimental viscosity of alcohols. Comparison of these two methods reveals that those model obtained by ASNN are better.

Key Words: QSPR models, Viscosity, Topological indices, Multilinear regression, Associative neural network.

INTRODUCTION

Quantitative structure property/activity (QSPR/QSAR) studies are unquestionably of great importance in computational chemistry and has been widely used in the prediction of physico-chemical properties and biological activities of organic compounds. The general idea of QSAR and QSPR is that property/activity of a new untested molecule can be readily estimated from the molecular structure of similar compounds whose properties/activities have already been determined. In other words, we may study the correlation between the properties/activities of the molecules and their structures. If a good correlation is found, then it would be easy to determine the properties/activities of various compounds, including those not yet synthesized[1]. Quantitative structure property relationship studies are performed on the basis of the correlation between the experimental values of the property and molecular descriptors reflecting the molecular structure of the respective compounds. Rigorous testing of the predictive power of the equations obtained is possible[2,3]. Hence the QSPR approach is a general and reliable method to predict various physico-chemical properties. To develop a QSPR model the following steps are usually involved *i.e.*, data collection, molecular geometry optimization, molecular descriptors generation, descriptors selection, model development and finally model performance evaluation. One of the important problems in QSPR is the description of molecular structures using molecular descriptors, which can include structural information as much as possible. At present there exist a great number of molecular descriptors that encode constitutional, topological, geometry and electronic features of organic compounds[3-5]. Among various structure descriptors, those derived from molecular structure alone have a particular advantage of the possibility to calculate them based only on the molecular structural feature and to be applicable to different families of compounds. After the calculation of molecular descriptors, linear methods, such as multiple linear regression (MLR) and non linear neural network can be used in the mathematical relationship between the molecular descriptors and the property to be predicted.

Liquid viscosity of organic compounds is one of the most significant transport property for many studies related to the transfer or movement of bulk quantities of liquids and simulation of the processes in chemical and petroleum industries[6,7]. Accurate experimental viscosity measurement of ever-growing number of actual and potential chemical products is time consuming and laborious[8]. On the other hand, a statistically significant QSPR approach, that requires only chemical structure data could serve as a tool to predict reliable viscosity data in a fraction of the time and expense. Numerous QSPR models for predicting the liquid viscosities have been proposed using fixed molecular descriptors. The concept of the variable molecular descriptors was proposed as an alternative way of

characterizing heteroatoms in molecules. The idea behind the variable molecular descriptors is that the variables are determined during regression fitting so that the standard of estimate for a studied property (viscosity) in as small as possible[9]. Several molecular descriptors have already been tested in their variable forms in QSPR. Here we report the use of the variable Zagreb index $^{V}M_2$ in the structure-viscosity modeling of alcohols.

EXPERIMENTAL

In parametric methods, one series of digital variables named molecular descriptors were used for evaluation of molecules properties. By using a method either multi linear regression (MLR) or a non-linear method such as associative neural network (ASNN), which states the relation between molecule structure and relative variable[10].

Molecular descriptor generation: For modeling, the descriptors which have relation with considered property have to be selected. The descriptors which are used in the present study are listed below: (1) Variable Zagreb index ($^{V}M_2$), (2) Number of hydroxyl groups (N_{OH}), (3) Number of carbon atoms (N_c), (4) Molecular weight (MW).

Variable Zagreb index: Topological indices are characterized by fixed numerical values, which are independent of property considered[11]. Hence they can be computed once the bonding pattern of molecule is known. In contrast, the variable Zagreb index $^{v}M_2$ a flexible descriptor depends on the property considered. The variable Zagreb index $^{v}M_2$ offers a powerful tool for the study of physico-chemical properties.

Here we report the uses of the variable Zagreb $^{V}M_2$ index in the structure-viscosity modeling of alcohols. The variable Zagreb index is calculated as below[12].

$$^{V}M_2 = \sum_{edges} [d(i)\, d(j)]^{\lambda} \quad (1)$$

where $^{V}M_2$ denote variable Zagreb M_2 index, d(i) is the degree of valence vertex, d(i) d(j) is the weight of edges i-j, λ is the variable parameter.

For example from the hydrogen suppressed graph of 2-methyl 1-propanol (Fig. 1) we obtained:

$$^{V}M_2 = (1 \times 3)^{\lambda} + (2 \times 3)^{\lambda} + (5 \times 2)^{\lambda} + (1 \times 3)^{\lambda} \quad (2)$$

The numbers in the graph (Fig.1) represents degree of valance vertex computed by valency of 'C' or 'O' minus number of hydrogen atoms attached. A C++ program was developed to calculate the variable Zagreb index for various λ values. The structure of 35 considered alcohols were taken from the work of Lide[13].

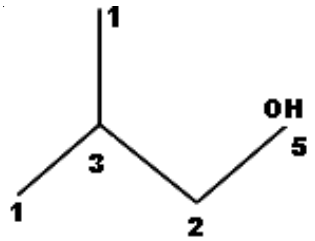

Fig . 1. Graph representing molecular skeleton of carbon atoms and oxygen atom for 2-methyl 1-propanol

The descriptor N_c and MW are constitutional descriptors and N_{OH} is a functional group calculated with E-dragon software[14].

Multi linear regression (MLR): MLR is one of the earliest methods used for constructing QSPR/QSAR models, but it is still one of the most commonly used one to date. The advantage of MLR is simple form and easily interpretable mathematical expression[15]. The selected descriptors were employed with MLR to develop the linear model of the property of interest, which takes the form:

$$Y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n \quad (3)$$

in this equation Y is the predicted property value of the dependent variable, x_1 to x_n represent the specific descriptors, while b_1 to b_n represent the coefficient of those descriptor, b_0 is the intercept of this equation. The software package SPSS 11.0 for windows was used to implement multi linear regression.

Artificial neural network (ANN): The growing interest in the application of artificial neural network in computational chemistry is a result of their demonstrated superiority over the traditional models. Neural network were used in quantitative structure property relationships studies to predict various physical and chemical properties.

An artificial neural network is a information processing paradigm composed of simple elements operating in parallel. The connections of complex system among these elements mostly determine the network functions. The ANNs are trained to perform a particular function by adjusting the values of the connections, or weights, between elements until a particular input leads to a specific output.

The ANN consists of three layers, input layer, hidden layer and output layer. These three layers are connected each other. The input layer receives the input data (descriptors) outside the network and sends them to the hidden layer. The hidden layer contains interconnected neurons for the pattern recognition and the relevant information interpretation for adjusting the weights on the connections. Afterwards, the result from the hidden layer are sent to the output layer for the output (property). The neurons contain several functions and variable including weights, non-linear transfer functions, methods to add up all inputs and bias values. The sum of products of all the inputs multiplying the weights and the bias values passes through a non-linear transfer function as the output of each neuron[16].

Associative neural network(ASNN): The traditional artificial feed forward neural network is a memory-less approach this means that after training is complete all informations about the input patterns is stored in the neural network weights and input data are no longer needed *i.e.*, there is no explicit storage of any presented example in the system. Contrary to that associative neural network is a combination of memory based and memory less methods. The recently proposed associative neural network offers an elegant approach to incorporate on the fly the user's data[17]. The ASNN is an extension of the committee of machines that goes beyond a simple/weighted average of different models. An ASNN represents a combination of an ensemble of feed forward neural networks (memory less) and the K-nearest neighbour technique (memory based). This

TABLE-1
NUMERICAL VALUES OF THE VARIABLE ZAGREB INDEX $^{V}M_2$ FOR SELECTED VALUES OF λ AND OTHER DESCRIPTORS USED IN THE PRESENT WORK.

S. No.	Compound Name	λ						N_{OH}	MW	N_C	$1/N_C^2$
		0.1	0.8	-0.1	-0.9	-1.4	-1.6				
1	Methanol	1.174619	3.623899	0.85134	0.234924	0.105061	0.076146	1	32.04	1	1
2	2-Methyl propan-1-ol	4.687403	15.31899	3.422204	1.069346	0.5508	0.426853	1	74.12	4	0.0625
3	2-Methyl butan-1-ol	5.843846	20.10089	4.298266	1.481875	0.841274	0.681132	1	88.2	5	0.04
4	3-Methyl butan-1-ol	5.836102	18.35042	4.292755	1.356521	0.694387	0.535672	1	88.15	5	0.04
5	2-Methyl butan-2-ol	5.949597	24.06761	4.227521	1.331593	0.735598	0.591697	1	88.15	5	0.04
6	2-Methoxy ethanol	4.885944	20.83435	3.280815	0.71928	0.295633	0.209581	1	76.09	3	0.111111
7	Ethanol	2.330699	8.050675	1.727361	0.661779	0.41874	0.354996	1	46.07	2	0.25
8	2-Ethoxy ethanol	6.043575	25.68286	4.157866	1.162635	0.624012	0.501341	1	90.12	4	0.0625
9	2-Ethyl hexan-1-ol	8.136682	24.90769	6.036338	2.006913	1.083499	0.85882	1	130.23	8	0.015625
10	1-Propanol	3.479397	11.08211	2.597912	0.948954	0.562327	0.463815	1	60.1	3	0.111111
11	2-Propanol	3.543266	13.54361	2.554682	0.831483	0.452163	0.357984	1	60.1	3	0.111111
12	Propyn-1,2-diol	4.882299	21.63793	3.289011	0.784707	0.358569	0.267556	2	76.09	3	0.111111
13	Propyn-1,3-diol	4.815248	18.68202	3.329758	0.826134	0.366796	0.267875	2	76.09	3	0.111111
14	1-Butanol	4.628096	14.11354	3.468462	1.236128	0.705914	0.572633	1	74.122	4	0.0625
15	2-Butanol	4.695147	17.06945	3.427716	1.194701	0.697687	0.572314	1	74.122	4	0.0625
16	Butan-1,3-diol	6.030998	24.66936	4.159561	1.071881	0.502156	0.376374	2	90.122	4	0.0625
17	1-Pentanol	5.776794	17.14498	4.339013	1.523303	0.849502	0.681452	1	88.15	5	0.04
18	1-Hexanol	6.925492	20.17641	5.209563	1.810477	0.993089	0.790271	1	102.17	6	0.027778
19	1-Heptanol	8.074191	23.20784	6.080114	2.097652	1.136676	0.89909	1	116.2	7	0.020408
20	2-Heptanol	8.141242	26.16375	6.039367	2.056225	1.128449	0.89877	1	116.201	7	0.020408
21	1-Octanol	9.322515	29.8184	6.88451	2.258533	1.200825	0.94398	1	130.23	8	0.015625
22	1-Nonanol	10.47121	32.84983	7.75506	2.545707	1.344413	1.052799	1	144.25	9	0.012346
23	1-Decanol	11.61991	35.88126	8.625611	2.832882	1.488	1.161618	1	158.28	10	0.01
24	Allyl alcohol	10.03717	49.38654	6.377501	1.047125	0.340457	0.217467	1	58.03	3	0.111111
25	Benzyl alcohol	3.479397	11.08211	2.597912	0.948954	0.562327	0.463815	1	108.14	7	0.020408
26	Thiophenol	8.896384	48.78454	5.512055	0.834804	0.261323	0.164737	1	140.2028	7	0.020408
27	Ethylene glycol	3.666549	15.65058	2.459207	0.53896	0.223209	0.159056	2	46.07	2	0.25
28	Glycerol	6.169239	27.31465	4.054902	0.776421	0.282218	0.189116	3	92.1	3	0.111111
29	Diethylene glycol	7.379425	33.28276	4.889712	1.039816	0.42848	0.305402	2	106.12	4	0.0625
30	Pentane-1,5-diol	7.112644	24.74488	5.070858	1.400483	0.65397	0.485513	2	90.121	4	0.0625
31	2-Ethyl hexane-1,3-diol	10.69661	39.92881	7.604328	2.308772	1.254141	1.007192	2	146.23	8	0.015625
32	2-Methyl pentane-2,4-diol	8.501198	37.65486	5.78917	1.45452	0.675427	0.504257	2	216.32	12	0.006944
33	2,2-Diethoxy ethanol	9.756451	43.31504	6.588371	1.663491	0.829283	0.647687	1	134.17	6	0.027778
34	Triethylene glycol	11.0923	50.91494	7.320217	1.540672	0.633752	0.451747	2	150.17	6	0.027778
35	2-Propyn-1-ol	3.772159	18.88798	2.386558	0.386626	0.125062	0.079779	1	56.06	12	0.006944

method uses the correlation between ensemble responses as a measure of distance among the analyzed cases for the nearest neighbour techniques. This method uses the correlation between ensemble responses as a measure of distance among the analyzed cases for the nearest neighbour techniques. This provides an improved prediction by the bias correction of the neural network ensemble. An associative neural network has a memory that can coincide with training set. If new data become available the network further improves its predictive ability and provides a reasonable approximation of unknown function without a need to retrain the neural network ensemble. This feature of the method dramatically improves its predictive ability over traditional neural network and K-nearest neighbour techniques.

In K-nearest neighbour approach, it keeps the entire database of examples in memory and their predictions are based on some local approximation of the stored examples. The neural network can be considered global models, while the other approach is usually considered local model. The developed algorithm of ASNN is available online at[18]. In the present study, the best model obtained from Multi linear regression analysis was applied to ASNN for comparative study. The model developed in this ASNN study were evaluated by the co-efficient of correlation R^2 which is the proportion of the variance that could be explained by the model and root mean square error (RMSE) of the predicted values to the experimental values.

Over-all viscosity data of 35 alcohols in the present investigation were obtained from the[6,7]. The compounds include a diverse set of substituted alcohols. Name of the compounds were presented in Table-1 and their viscosities are shown in Table-2. In this work, we searched for the optimal value of λ. In order to determine λ, we used the set of C_1-C_{12} alcohols and their viscosities. We considered the following two descriptor relationship.

$$\log(\eta) = a + b\,TI + cN_{OH} \quad (4)$$

where log(η) stands for logarithmic viscosity and TI for Topological index (variable Zagreb index).

Then we use the three descriptor relationship.

$$\log(\eta) = a + b\,TI + c\,N_{OH} + d\,[1/N_c^2] \quad (5)$$

Finally we tested the four-descriptor relationship.

$$\log(\eta) = a + b\,TI + c\,N_{OH} + d\,1/N_C^2 + e\,MW \quad (6)$$

The structure-viscosity modeling was based on the CROMRsel procedure[9]. This is a multivariate procedure, which

TABLE-2
THE RESULT OF MLR AND ASNN ANALYSIS AND CALCULATED VALUES OF VISCOSITIES. THE NUMBERING OF MOLECULES CORRESPONDS OF THE MOLECULAR LISTING OF TABLE-1

Sl. No.	Experimental Log(η) Value	Calculated Log(η) values 2 descriptors model	Residual	Calculated Log (η) values 3 descriptors model	Residual	Calculated Log(η) values 4 descriptors model	Residual	Predicted Log(η) Values by ASNN	Residual
1	-7.4236	-6.8407	-0.58291	-6.7908	-0.63277	-7.6786	0.25502	-7.43	0.01
2	-5.5130	-5.9713	-0.45834	-5.9343	0.42129	-5.8397	0.32668	-5.59	0.08
3	-5.2030	-5.3410	0.13801	-5.3256	0.12257	-5.2692	0.06617	-5.44	0.24
4	-5.2766	-5.7016	0.42503	-5.6602	0.38361	-5.5839	0.30737	-5.38	0.1
5	-5.3743	-5.5627	0.18837	-5.5662	0.19191	-5.4955	0.12120	-5.33	-0.04
6	-6.3654	-6.5099	0.14448	-6.5233	0.15790	-6.4561	0.09070	-6.57	0.2
7	-6.7254	-6.1495	-0.57598	-6.1290	-0.59640	-6.1888	-0.53661	-6.55	-0.18
8	-6.1948	-5.7867	-0.40812	-5.8280	-0.36678	-5.7729	-0.42196	-5.74	-0.45
9	-4.6254	-4.9006	0.27518	-4.9327	0.30732	-4.9575	0.33209	-4.75	0.16
10	-6.1040	-5.8797	-0.22425	-5.8549	-0.24905	-5.7940	-0.30995	-5.98	-0.12
11	-5.5450	-6.1420	0.59706	-6.1060	0.56101	-6.0303	0.48530	-6.04	0.5
12	-3.0865	-3.0682	-0.01835	-3.0306	-0.05593	-2.9814	-0.10509	-2.86	-0.23
13	-2.8824	-3.0674	0.18499	-3.0118	0.12944	-2.9638	0.08139	-2.79	-0.09
14	-5.8266	-5.6100	-0.21666	-5.5808	-0.24587	-5.5070	-0.31962	-5.53	-0.3
15	-5.5450	-5.6108	0.06577	-5.5995	0.05453	-5.5247	-0.02033	-5.55	0.01
16	-2.0379	-2.7984	0.76052	-2.7565	0.71857	-2.6945	0.65659	-2.18	0.14
17	-5.5215	-5.3402	-0.18124	-5.3066	-0.21483	-5.2512	-0.27021	-5.43	-0.09
18	-5.2903	-5.0705	-0.21987	-5.0325	-0.25788	-5.0077	-0.28263	-5.21	-0.08
19	-4.9598	-4.8007	-0.15912	-4.7584	-0.20149	-4.7700	-0.18980	-4.98	0.02
20	-5.0313	-4.8015	-0.22983	-4.7771	-0.25424	-4.7877	-0.24366	-5.00	-0.03
21	-4.7189	-4.6895	-0.02944	-4.6653	-0.05361	-4.7058	-0.01307	-4.66	-0.06
22	-4.2745	-4.4197	0.17221	-4.3911	0.14363	-4.4730	0.22549	-4.44	0.19
23	-4.2416	-4.1500	-0.09168	-4.1170	-0.12463	-4.2413	-0.00033	-4.18	-0.06
24	-6.5981	-6.4904	-0.10770	-6.3528	-0.24531	-6.2582	-0.33988	-6.32	-0.28
25	-5.1886	-5.8797	0.69114	-6.0369	0.84831	-5.9564	0.76787	-5.52	0.33
26	-6.6935	-6.6211	-0.07237	-6.8444	0.15092	-6.7827	0.08929	-6.61	-0.08
27	-3.9170	-3.3371	-0.57990	-3.2254	-0.69163	-3.2685	-0.64855	-3.89	-0.03
28	0.3988	0.0354	0.36338	0.0840	0.31475	0.1043	0.29451	0.181	0.218
29	-3.3326	-2.9744	-0.35824	-2.9850	-0.34760	-2.9427	-0.38991	-3.06	-0.27
30	-2.0557	-2.5279	0.47218	-2.4105	0.35474	-2.3689	0.31317	-1.92	-0.14
31	-1.1301	-1.2347	0.10464	-1.2551	0.12498	-1.3420	0.21186	-1.08	-0.05
32	-3.3697	-2.4814	-0.88826	-2.8396	-0.53013	-2.9679	-0.40182	-3.47	0.1
33	-5.5597	-5.4239	0.13576	-5.5270	-0.03266	-5.5394	-0.02026	-5.35	-0.21
34	-3.0159	-2.6116	-0.40433	-2.6840	-0.33194	-2.7093	-0.30667	-3.09	0.07
35	-6.3890	-6.8317	0.44271	-6.8362	0.44726	-6.5846	0.19563	-6.48	0.09

selects the best possible model among the set of models, the criterion being the standard error of estimate.

The exponent for $^{V}M_2$ varied from $\lambda = 1$ to $\lambda = -2$, in steps of 0.1 in order to detect the optimal values of λ, which the variable Zagreb indices give the structure-viscosity models with the lowest value of the standard error of estimate (SE) and the highest value of the correlation co-efficient (R^2).

RESULTS AND DISCUSSION

The overall performance of MLR is evaluated in terms of standard error of estimate and the other parameters: correlation co-efficient R^2 and F is the result of Fisher's test. The models were also cross-validated by using leave-one-out (LOO) procedure. LOO procedure is a procedure usually used for evaluating models stability. During LOO procedure each of N molecule is taken away only once. Statistical parameters for the cross-validated models are denoted by R^2(cv) and SE(cv), where cv stands for cross-validation procedure.

Result of MLR and cross validation model.

i. The best two Descriptor model.

$$\log(\eta) = 2.4788\ {}^{V}M_2^{[\lambda=-1.6]} + 3.2980\ N_{OH} - 10.327 \quad (7)$$

N = 35

$R^2 = 0.949$, SE = 0.40083,

F = 300.633

R^2 (cv) = 0.95, SE(cv) =0.393.

ii. The best three descriptor model.

$$\log(\eta) = 2.2791\ {}^{V}M_2^{[\lambda=-1.4]} + 3.3492\ N^{OH} - 0.003787\ MW - 10.258 \quad (8)$$

N = 35

$R^2 = 0.952$, SE = 0.3972,

F = 206.238

R^2(cv) = 0.960, SE(cv) = 0.3995.

iii.The best four descriptor model.

$$\log(\eta) = 2.1447\ {}^{V}M_2^{[\lambda=-1.4]} + 3.34\ N_{OH} - 0.005MW - 1.1948\ (1/Nc^2) - 9.868217 \quad (9)$$

N = 35

$R^2 = 0.962$, SE = 0.35878, F = 190.09, R^2 (cv) = 0.968, SE(cv) = 0.3333.

The above results shows that variable Zagreb index produce better models. The best linear model contains four

descriptors with $R^2 = 0.962$ for the optimal value of $\lambda = -1.4$. The calculated results of log (η) values from eqns.7-9 were given in the Table-2. The cross validation results show a small increase when compared with those obtained in the calibration process (R^2 (cv)= 0.968, SE (cv)= 0.333)

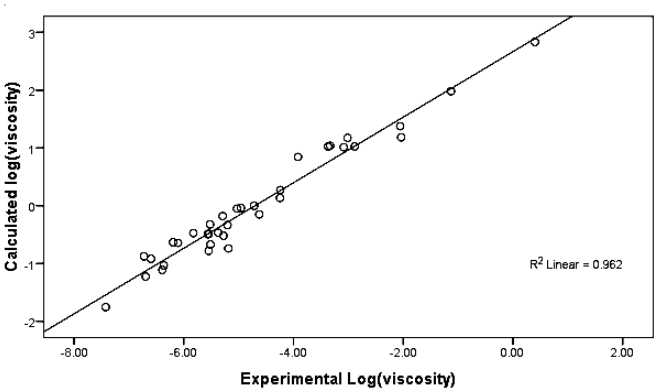

Fig. 2. Plot of experimental log (η) values against calculated log (η) values using MLR (four descriptor model)

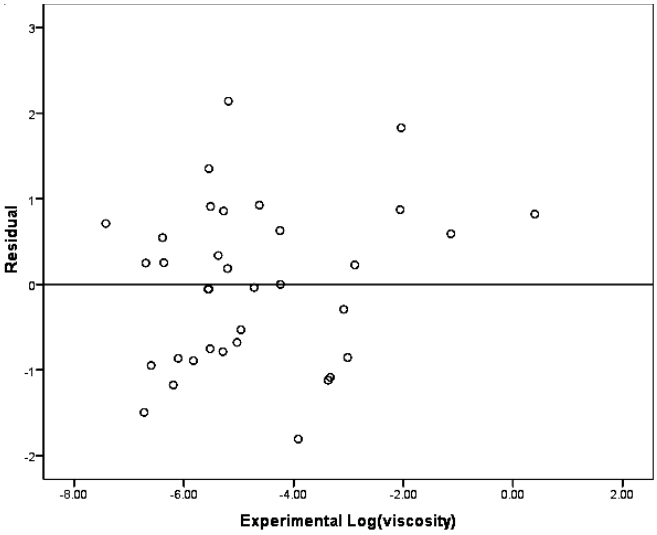

Fig. 3. Scatter plot of residuals against experimental log (η) values. (four descriptor model)

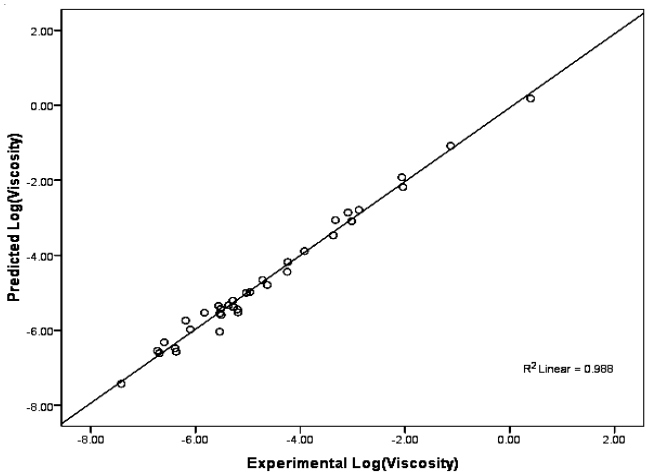

Fig. 4. Plot of experimental viscosity against calculated using ASNN.

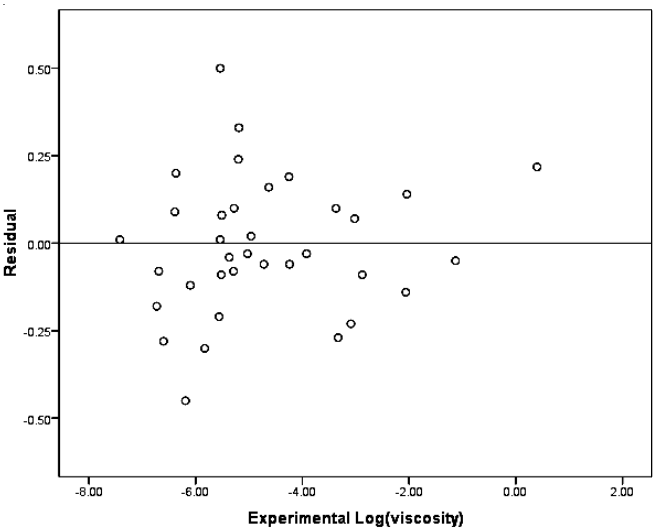

Fig. 5. Scatter plot of residual against experimental log (h) values (ASNN)

The plot of the calculated *versus* experimental log(η) of the best four descriptor model is shown in Figs. 2 and 3. shows the scatter plot between residual of MLR calculated values and experimental log (η) values. Fig. 2. shows that there is a good agreement between experimental and calculated values using MLR (4 descriptor model) and Fig. 3. shows that there is no systematical. The MLR analysis shows that all the three models to contains $^{V}M_2$ and N_{OH}, but the values of R^2 and SE were significantly reduced after the removal of N_{OH}. So that N_{OH} is a important descriptor which reflects hydrogen bonding and plays key factor in control liquid viscosity. The four descriptors used in MLR analysis were applied to train the ASNN for comparative study.

Associative neural network (ASNN) represents an innovative method to calculate nonlinear models between structure and property. In the present study, the network involves four neurons ($^{V}M_2^{[\lambda=-1.4]}$, N_{OH}, MW & $1/Nc^2$) in the input layer, seven neurons in the hidden layer and one neuron in the output layer (log η). The network is trained using the Leven Berg Marquardt algorithm. Number of hidden neuron was decided by training and predicting the training data by varying the number of hidden neurons in the hidden layer. A suitable configuration has to be chosen for the best performance of the network. Out of the different configuration tested, a hidden layer with 7 hidden neurons produced the best result for prediction of viscosity of alcohols. The performance of the associative neural network QSPR model for viscosity estimation is summarized in Table-2 and Figs. 4 and 5. The correlation coefficient (R^2) of 0.991 and RMSE of 0.2158 log units shows a good agreement of ASNN predicted values with experimental one (Fig. 4). The propagation of residuals on the both sides of the zero (Fig. 5) indicates that there is no systematic error in using ASNN model. The distribution of residual values are given in Table-2. Table-2 clearly shows the high statistical quality performance of the selected ASNN model. The number of alcohols not correctly predicted by the model is very limited. Thus, only two alcohols have their residual value within 0.50 log units (absolute value) and glycerol have residual of 0.218 log units, which is very low compared to MLR models.

Conclusion

The variable Zagreb index $^{V}M_2$ was used in the structure - viscosity modeling of alcohols. Two, three and four descriptor models were performed with MLR. The best one is the four descriptor models with $R^2 = 0.962$ and SE = 0.3858 log units for the optimal regression value of -1.4. The predictive ability of the MLR models were tested by leave-one-out cross validation method, showing QSPR model is stable and can be used to obtain good predictions for viscosity of alcohols. The four descriptors used in best MLR model was applied to the ASNN for training purpose. The results shows that, ASNN predicts viscosity very well with $R^2 = 0.991$ and RMSE = 0.2158 during training phase. The ASNN produces high statistical quality and low prediction error model compared with MLR analysis.

REFERENCES

1. R. Katritzky, U. Maran, U.S. Lobanov and M. Karelson, *J. Chem. Inf. Comput. Sci.,* **40**, 1 (2000).
2. A.R. Katritzky, V.S. Lobanov and M. Karelson, *Chem. Soc. Rev.,* 279 (1995).
3. A.R. Katritzky, M. Karelson and V.S. Lobanov, *Pure Appl. Chem.*, **69**, 245 (1997).
4. E. Estrada and E. Uriarte, *Curr. Med.*, **8**, 1699 (2001).
5. M. Karelson, Molecular Descriptors in QSAR/QSPR, John Wiley & Sons, NY (2000).
6. A.R. Katritzky, K. Chen, Y. Wang and M. Karelson, *J. Phys. Org. Chem.*, **13**, 80 (2000).
7. O. Ivanciuc, T. Ivanciuc, P.A. Fillip and D. Cabrol, *J. Chem. Int. Sci.*, **39**, 515 (1999).
8. J. Homer, S.C. Generalis and J.H. Robson, *Phys. Chem. Chem. Phys.*, **1**, 4075(1999).
9. S. Nikolic, A. Milicevic, N. Trinajstic and A. Juric, *Molecules*., **9**, 1208 (2004).
10. F. Ashrafi, R. Saadati and A.B. Amlashi, *Afr. J. Pure Appl. Chem*., **2**, 116 (2008).
11. M. Randic, *New J. Chem.*, **24**,165 (2000).
12. A. Milicevic and S. Nikolic, *Croat. Chem. Acta*, **77**, 97 (2004).
13. D.R. Lide, CRC Handbook of Chemistry and Physics, CRC Press, Boca Raton, edn. 75 (1994).
14. http://www/vcclab.org/lab/edragon/
15. P. Liu and W. King, *Int. J. Mol. Sci. Rev.*, **10**, 1978 (2009).
16. J. Devillers, Neural Networks in QSAR Drug Design, Academic Press, London (1996).
17. I.V. Tetko, *J. Chem. Inf. Comput. Sci.*, **42**, 717 (2002).
18. http://www.vcclab.org/asnn/