International Journal of Chemical Science
Online ISSN: 2523-2843, Print ISSN: 2523-6075; Impact Factor: RJIF 5.22
Received: 18-03-2020; Accepted: 19-03-2020; Published: 19-04-2020
www.chemicaljournals.com
Volume 4; Issue 2; 2020; Page No. 16-22

Quantitative structure-property relationship study of water solubility for aliphatic alcohols based on a new model combined modified autocorrelation method and electro-topological indices

Chaymae Jermouni[1*], Mohamed Nohair[2]
[1, 2] Laboratory of chemistry, physics and Bio-organic chemistry, Faculty of Sciences and Technologies of Mohammedia, Hassan II University of Casablanca, Morocco

Abstract

In this study, structure water solubility modeling was performed to describe a set of 50 of aliphatic alcohols in a Quantitative Structure-Property Relationship model by developing two descriptors types based on multifunctional autocorrelation method and electro-topological descriptors. The Modified Autocorrelation Method was used to describe the local environment of the hydroxyl group. Multiple Linear Regression, Artificial Neural Networks and Principal Components Analysis were analyzed the alcohols series. The efficiency approach was evaluated through the predictive ability of models by leave-p-Out cross-validation method. The molecules were coded by means of SMILES system and stored as input files. The results showed that aliphatic alcohols solubility is dominated by the shape and molecule branching, also the electro-topological descriptors had a considered model effect.

Keywords: water solubility; aliphatic alcohols; modified autocorrelation method; electro-topological indices; multiple linear regression; artificial neural networks

1. Introduction

The present study is aimed at estimating water solubility of Aliphatic Alcohols, because it is considered as an important property of organic compounds [1]. Since alcohols are toxic materials, therefore it represents dangerous environmental pollutants; the first step in polluting alcohols action is their solubility in water. Various quantitative structure-property relationship (QSPR) models were proposed for estimating water solubility of Aliphatic Alcohols. In General, QSPR methods are based on the assumption that chemical compound property related to its structure through a certain mathematical algorithm. This relationship can be used in the prediction, interpretation, and new assessment compounds with desired properties reducing time, efforts and synthesis cost, and rationalizing of new product development [2-7].

Several factors influencing the physical property of molecule. Such us, the shape, the electronic structure and the molecular size. These factors are associated with different intermolecular interaction aspects, as though Van Der Waals forces, the polarity and the molecule ability to participate in hydrogen bonding caused by heteroatoms and which considered a very interesting factor influencing the physical properties of alcohols. A significant approach is based on various group contribution methods using atomic types, bonds, or molecular fragments as descriptors with respect to different properties. [8]. Among the many methods that describe the chemical structure for a set of molecules is that based on the molecular graph. There are now many different descriptions, among the methods used to describe molecular structures, many are based on the topological indices (TI) [1, 9-11]. derived by the molecular graph. In fact, these conventional indices don't take into account the each contributions of the individual types or groups of atoms to properties. For these reasons, the multifunctional autocorrelation method was used to describe the influence of the local environment of the molecule.

for the global molecule and the sum calculated for all pairs of atoms in the graph separated by a distance ***d*** of ***k*** bonds is considered as the component of order ***k*** (P_k) of an autocorrelation vector P corresponding to the specific property, and to describe the influence of the local environment of the molecule we used the modified autocorrelation method.

On the other hand, to increase the efficiency of topological indices, Kier and Hall [12]. set up the electro-topological indices that boil down in the valence connectivity indices, including information not present in the graph such as the valence electron and the number of electrons in an atom with respect largely to their use as descriptors in QSAR/QSPR modeling.

Three-layer feedforward neural network was used for statistical method trained by the back-propagation algorithm and Multiple Linear Regression (MLR).

Also predictive linear and non-linear models were obtained by forward stepwise multi-linear regression analysis (MRA) and Artificial Neural Network (ANN) approaches respectively.

2. Method

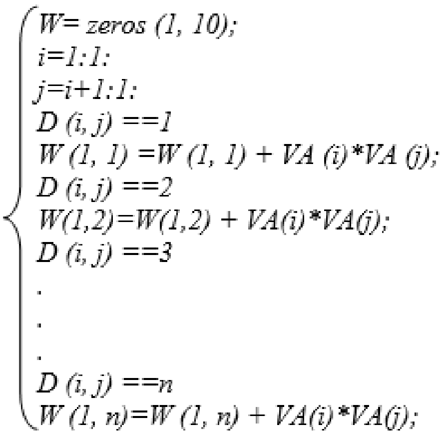
MAM method originate from Moreau Autocorrelation method [13], elected to implement structure-property relationships based on the autocorrelation function. This method can be useful for database characterize action and encoding various physicochemical properties [7, 14-16]. Moreau and Broto [8]. consider the molecule as a graph, in which the nodes are atoms (hydrogen atoms are not taken into account) linked by bonds. Properties *f* of atoms (surface, volume, electronegativity…) are defined on every node of the graph. The Components P, of the autocorrelation vector corresponding to that property are defined by the relation bellow:

$$Pk = \sum_{k=0}^{n} ((fi)(fj)) \tag{1}$$

Where P_k is autocorrelation component corresponding to topological distance of k bonds (smallest number of bonds between i and j) to the specific property f_i

To facilitate the computation for Big Data, the authors develop a computation program on Matlab which generates the connection matrices, distances matrices and the variables of the MAM was developed in the following way:

The first program was used for the calculation of global environment descriptors of the molecule.

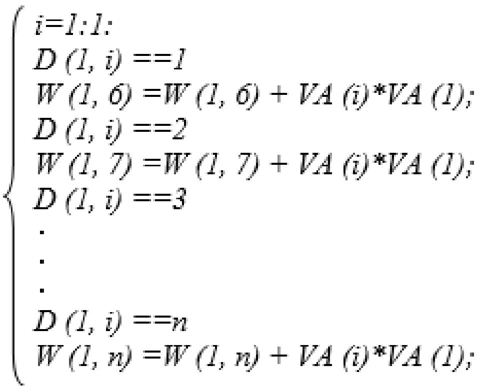


```
W= zeros (1, 10);
i=1:1:
j=i+1:1:
D (i, j) ==1
W (1, 1) =W (1, 1) + VA (i)*VA (j);
D (i, j) ==2
W(1,2)=W(1,2) + VA(i)*VA(j);
D (i, j) ==3
.
.
.
D (i, j) ==n
W (1, n)=W (1, n) + VA(i)*VA(j);
```

A new component Pik was specified to characterize the local environment of the atoms in molecules by the formula bellow:

$$Pk = \sum_{k=0}^{n} ((fi)(fj))^{x} \tag{2}$$

The principle of Modified Autocorrelation Method proposed by Nohair *et al* [16], comes from setting carbon atom i and P_{ik} the sum of f(i)*f(j) of all chemical bonds existing between all pairs of carbon atoms i(fixed atom) and j separated by a topological distance k, and x=1. The second program was applied to study the local environment of molecule, while fixing the atom concerned:



```
i=1:1:
D (1, i) ==1
W (1, 6) =W (1, 6) + VA (i)*VA (1);
D (1, i) ==2
W (1, 7) =W (1, 7) + VA (i)*VA (1);
D (1, i) ==3
.
.
.
D (1, i) ==n
W (1, n) =W (1, n) + VA (i)*VA (1);
```

The second model is based on the use of electro-topological descriptors of Kier and hall [17]. The molecular shape indices of Kier and Hall kappa aims to compare the molecular graph for minimum and maximum and intended to exploit the different aspects of molecular form. The molecular operating environment software (MOE) was used to generate the Kier and hall descriptors.

3. Results and discussion

3.1. Experimental Data set

A series of 50 Aliphatic Alcohols reported by Amic *et al* studied [18]. With short and long carbon chain with different category (linear and branched), primary, secondary and tertiary carbon. The QSPR analysis was performed using the experimental solubility of aliphatic alcohols in water (expressed in lnSol) which depends on two factors: influence of hydrophobic carbon chain and hydrophilic hydroxyl group.

3.2. Molecular descriptors Determination

In the data processing part, the descriptors were obtained by the MAM method, and the electro-topological indices were generated by the MOE (Molecular Operating Environment software in order to set up a model that combines these two types of descriptors for the QSPR modeling.

3.3. Statistical analysis

To determine the structure-property relationship by the four descriptors calculated by the Multifunctional (V_1, V_2) and the Modified Autocorrelation Method (V_{11}, V_{12}), the authors used Multiple linear regression and Artificial Neural Networks as methods of modeling. The firsts represent the global description to the environment of the molecule in space while the seconds describe the hydroxyl group in the molecule. The component V_1, takes into account the branching and the size of the molecule. it is calculated by the sum f(i)*f(j) of all the chemical interactions between all the pairs of adjacent carbon atoms in the molecule. A study has already been done to analyze the behavior of V_k components. It was based on Van Der Waals volumes (V_0 - V_5). The model provided the following statistical indicators for the solubility of aliphatic alcohols ($n = 50$, $r = 0.97$, $s = 0.38$) [16]. In the model, we add components that have a relationship with the description of the hydroxyl group Vik ($k = 1, 2$) and we add descriptors of Kiers, we observe the significant progression and we got for $n=50$, $s=0.18$; $r^2=0.99$.

3.3. Neural network

Artificial neural networks (ANNs) are very common for QSPR-type mathematical models that convert structural features into different properties of chemical compounds. In this study, we demonstrate that the multifunctional autocorrelation and kier methods are effective for modeling the water solubility of aliphatic alcohols. The main benefit of using neural networks in QSPR models is their ability to offer non-linear mapping of descriptors that describes a structure-property relationship. A comparison between the three different training algorithms was made to find out which one gives the best results with high performance. Table 4 generalizes the results obtained:

Table 1: Comparison of the three different training algorithms

Levenberg-Marquardt	Bayesian Regularization	Scaled Conjugate Gradient
S 0,66	0,53	0,80
R^2 0,95	0,97	0,93

3.4.1 Levenberg-Marquardt

This algorithm typically requires more memory but less time. Training automatically stops when generalization stops improving, as indicated by an increase in the mean square error of the validation samples.

3.4.2 Bayesian Regularization

This algorithm typically requires more time, but can result in good generalization for difficult, small or noisy datasets. Training stops according to adaptive weight minimization (regularization).

3.4.3 Scaled Conjugate Gradient

This algorithm requires less memory. Training automatically stops when generalization stops improving, as indicated by an increase in the mean square error of the validation samples.

Based on these results, (table 1) it can be concluded that Bayesian Regularization has better capability of a short-term forecast, however in the long run, it loses its accuracy and follows similar performance to that of Levenberg-Marquardt.

There are various techniques of cross-validation, and in our case the leave-20% out technique has been applied using p observations as the validation set and the remaining observations as the training set. This is repeated on all ways to cut the original sample on a validation set 20% and training set 80%.the results are shown in the table 5.The true error S_{cv} is estimated as the average error rate on test examples:

$$\mathrm{Scv} = \frac{1}{N\sum_{i=1}^{N} \mathrm{Si}} \qquad (3)$$

Scv= 0.20 by the Multiple linear regression modeling.

Scv= 0.70 by the Artificial Neural network modeling.

After dividing our data series into two parts 80% for calibration and 20% for validation, the results were obtained with five distributions, the significant progression of the model was observed by contributing the edge OC with respect to the model describing just the global environment of each molecule, the results of n=50, s=0.22; r=0.98. By the multiple linear regression and s=0.70; r=0.96 for Artificial Neural Network.

3.4. Variables Selection

In this study, the approach followed for determining the subset of variable selection problem is the forward stepwise regression [24, 25, 26], this method is simple to define. the operation is carried out at the start without variables; the variable that had the highest R-square will be selected. At each step, the variable that has increased R-Squared the most will be selected, then adding variables stops when none of the other variables is significant. The choice of relevant variables is confirmed by AIC: Akaike Information Criterion [27- 30]. The introduction of this criterion is widely used for good modeling selection, which is the most important aspect of statistical inference; also, AIC is the basis of a paradigm for the foundations of statistics.

A formula for least squares regression type analyses:

$$\mathrm{AIC} = n\log(\mathrm{RSS}) + 2k \qquad (4)\ ^{[13]}$$

Where:

RSS: Residual Sum of Squares/n, n = sample size, K is the number of model parameters.

The model containing MAM descriptors performed better than previous model, and that good statistical indicators were obtained when the two components V_{11} and V_{I2} were added. Excellent MLR regression: n = 50; s = 0.22; r^2 = 0.98 was obtained with the three components V_1, V_{11}, V_{12}. The forward stepwise regression was performed by the selection of the most significant term to each step, the operation when added variable stops improving the model.

The results of Akaike information criterion show that the (The higher the number, the better the fit. This was obtained from statistical output) that the two criterion of the model V_1, V_2, V_{11} and V_1, V_2, V_{11}; V_{12} are close: -147.83 and -144.39 respectively. The model wasn't improved when the component V_{12} (r = 0.98, s = 0.23) was added.

3.5. Results of Electro-topological indices by Kier and Hall

The same approach was followed for Kier and Hall descriptors and the same techniques used in the first type of descriptors, the statistical indices obtained are: s=0.42; r^2= 0.96 by the MLR, and s=0.44; r^2= 0.98 by ANN.

For the cross validation, validation errors was obtained by summing S_{cv}= 0.41 by MLR, and S_{cv}= 0.80 by ANN. Forward stepwise is a tractable approach and it gives a good sequence of models. the progression of the model was observed by contributing the most significant variables Kier 1; Kier 2; Kier3 (K1; K2 and K3) with the lowest value of AIC -82.42 (table 10).

3.6. Combination of two types of descriptors: MAM and KIER

3.6.1. Principal Component Analysis (PCA)

All the 11 descriptors encoding the 50 molecules were submitted to a Principal Component Analysis (PCA) and 11 components were obtained (Figure.1). The objective consists in transforming the correlated variables into new variables uncorrelated from each other. It reduces the number of variables and makes the information less redundant. This transformation is defined in such a way that the first component has the largest possible variance that is, accounts for as much of the variability in the data as possible. The first three axes, F1; F2 and F3 contributed respectively 55.74%; 23.87% and 8.42 to the total variability, and the total information 88, 04%.

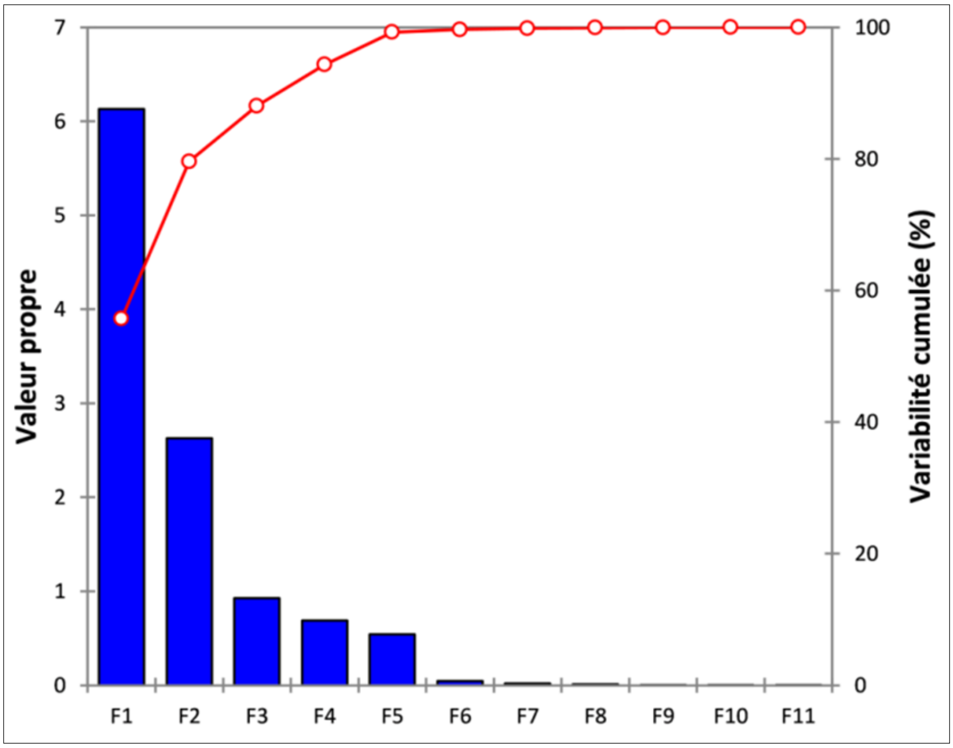


Fig 1: Principal Components and their variances

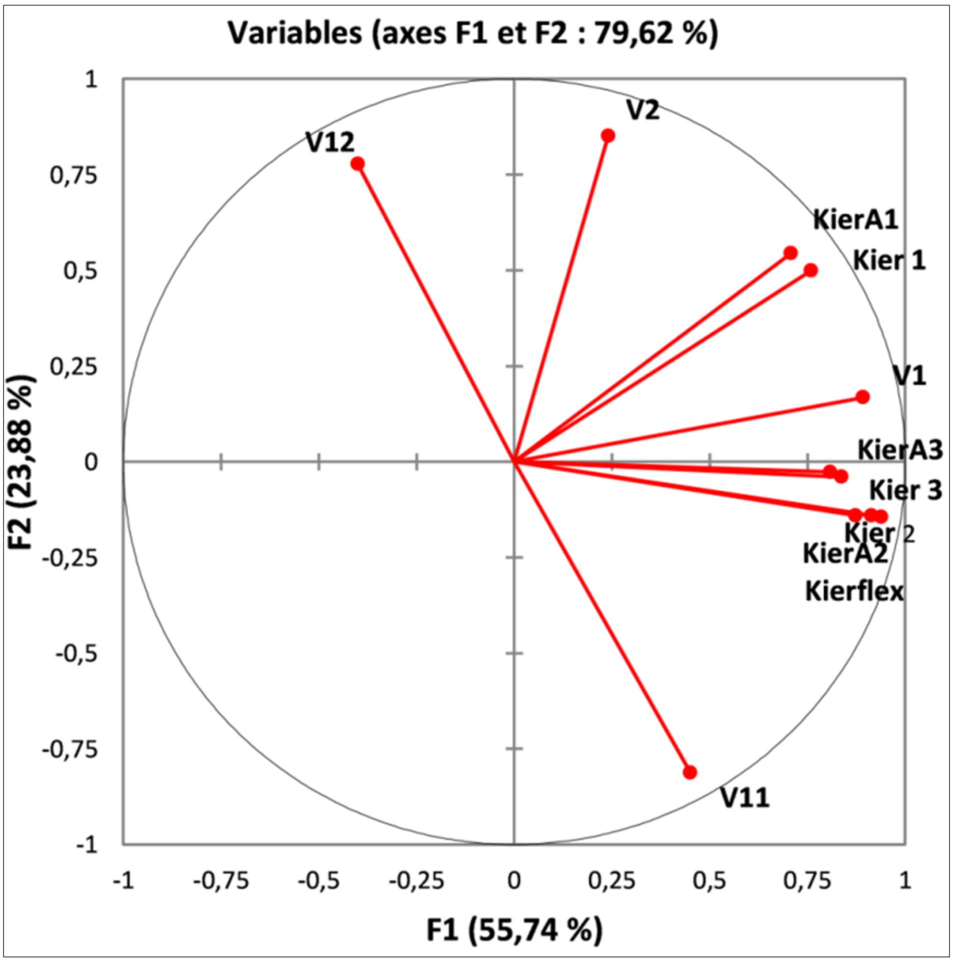


Fig 2: Representation of the descriptors in correlation circle

In the correlation matrix (Figure 2), V11 and V12 are perfectly correlated (r= 0, 98), the second descriptors of the Modified Autocorrelation Method is removed.

(Kier1; KierA1), (Kier2; KierA2), (Kier3; Kier A3) are correlated and consequently Kier A1, KierA2, KierA3 are removed. These results confirm the choice of descriptors by forward stepwise regression and by Akaike Information Criterion.

The link of these two types of descriptors table (2) was in order to know the complementary role of the electro-topological indices kier and hall on our first model.

so we notice the obvious improvement of our model which decreased of an error from s_{MLR}= 0.22 and s_{ANN}= 0.42 to 0.17 and 0,32 by MAM descriptor's and Kier descriptor's respectively with a very low value of AIC compared to the other two models (-173.61).

The regression model developed by combining the descriptors of the MAM and the descriptors of Kier for the water solubility of alcohols is excellent. The standard deviation, 0.18, is better than those obtained for the other two data sets. It is this statistical indicator that is most important because it relates directly to the interests of the experimental scientist who wishes to do so. It is also the first report for this particular data series that includes different forms of aliphatic alcohols, as described.

Table 2. Combination of MAM and KIER descriptors

The values of experimental water solubility of Aliphatic Alcohols and predicted values are illustrated in Figure. 3 for Multiple Linear Regression and Artificial Neural Network respectively, which shows a strong correlation between the predicted and experimental data.

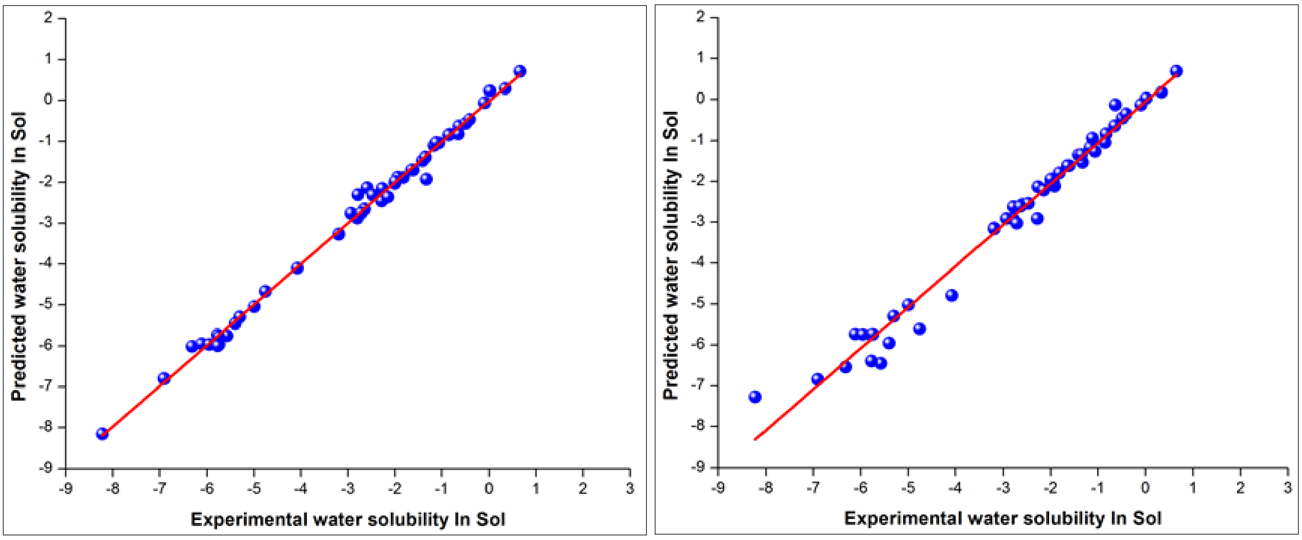


Fig 3: Predicted vs experimental water solubility of Aliphatic Alcohols by MLR and ANN

4. Conclusion

Multiple Linear Regression and Artificial Neural Network were used to construct a quantitative structure proprety relation model of 50 Aliphatic Alcohols for their water solubility; the study of this property is interesting because the toxic action of these compounds belongs primarily on it. the two modeling methods were compared and had significantly better predictive capability with a greater power, it is concluded that the solubility seems to be largely determined y the components (V_1, V_2, and V_{11}), which represent the size and the branching of the molecule without neglecting the influence of hydroxyl group caused by C–OH, also the addition of kier and hall descriptors (Kier1, Kier2, Kier3) resulted in a good meaningful model. The efficiency of the model is demonstrated through the combination of the two types of descriptors. The high quality of QSPR Model has been obtained for the property our data sets of alcohols.

Table 2. Combination of MAM and KIER descriptors

	V_1	V_2	V_3	K1	K2	K3	EXP	MLR	ANN
1	39.41	46.95	11.83	5	2.25	4.00	0.0227	0.23	0.03
2	39.41	46.95	9.63	5	2.25	4.00	0.658	0.71	0.69
3	49.64	57.47	11.83	6	3.20	5.33	-1.168	-1.10	-1.17
4	49.93	53.14	11.83	6	3.20	3.00	-1.0584	-1.04	-1.27
5	49.64	57.47	9.63	6	3.20	5.33	-0.6349	-0.63	-0.13
6	49.93	53.14	9.63	6	3.20	3.00	-0.4861	-0.56	-0.45
7	45.29	65.85	9.63	6	2.22	3.00	-0.405	-0.46	-0.35
8	37.9	83.23	6.75	6	1.63	5.33	0.3386	0.29	0.18
9	60.16	63.66	9.63	7	4.17	3.84	-1.326	-1.92	-1.54
10	48.82	84.7	6.75	7	2.34	2.67	-0.8301	-0.83	-0.84
11	48.13	94.15	6.75	7	2.34	6.00	-1.1178	-1.03	-0.94
12	55.82	72.33	9.63	7	3.06	2.67	-1.6094	-1.70	-1.62
13	55.82	72.33	9.63	7	3.06	2.67	-1.6399	-1.70	-1.62
14	44.25	97.06	6.75	7	1.85	2.67	-0.851	-0.84	-1.05
15	48.13	94.15	11.83	7	2.34	6.00	-2.5903	-2.14	-2.56
16	44.25	97.06	9.63	7	1.85	2.67	-1.4106	-1.47	-1.36
17	59.87	67.7	11.83	7	4.17	6.00	-2.2828	-2.46	-2.92
18	55.17	81.42	9.63	7	3.06	6.00	-1.814	-1.88	-1.79

19	60.46	59.57	11.83	7	4.17	2.67	-2.7871	-2.31	-2.63
20	58.36	104.38	6.75	8	3.11	7.20	-2.4734	-2.31	-2.55
21	59.05	95.61	6.75	8	3.11	3.67	-2.2634	-2.15	-2.13
22	59.73	86.41	6.75	8	3.11	2.22	-1.9173	-1.89	-2.12
23	54.77	103.93	6.75	8	2.52	2.81	-2.0025	-2.03	-2.06
24	55.16	98.81	6.75	8	2.52	2.22	-1.9379	-1.88	-2.03
25	53.66	118.56	6.75	8	2.52	7.20	-2.1456	-2.36	-2.22
26	61.7	91.88	9.63	8	3.11	2.81	-2.8018	-2.87	-2.92
27	54.77	103.93	9.63	8	2.52	2.81	-2.6437	-2.66	-2.62
28	70.39	73.89	9.63	8	5.14	5.00	-3.1942	-3.26	-3.15
29	70.39	74.19	9.63	8	5.14	5.00	-3.1966	-3.27	-3.16
30	54.73	124.65	6.75	9	2.32	2.00	-2.9318	-2.76	-2.92
31	80.33	88.16	9.63	9	6.13	8.00	-4.756	-4.67	-5.61
32	80.92	80.33	11.83	10	7.11	8.00	-4.9967	-5.04	-5.01
33	90.56	98.39	9.63	10	7.11	9.14	-6.32	-6.01	-6.54
34	90.85	94.35	9.63	10	7.11	7.00	-6.1193	-5.95	-5.74
35	90.85	94.65	9.63	10	7.11	7.00	-5.9522	-5.96	-5.75
36	90.85	94.65	9.63	10	7.11	7.00	-5.7446	-5.96	-5.75
37	81.45	122.08	9.63	10	4.76	7.00	-5.7764	-5.73	-5.76
38	82.75	104.85	9.63	10	4.76	3.11	-5.2983	-5.29	-5.29
39	80.88	99.28	11.83	10	7.11	9.14	-5.5728	-5.76	-6.45
40	90.56	98.39	11.83	10	4.00	7.00	-5.7446	-5.79	-5.74
41	74.42	135.27	11.83	10	4.00	7.00	-5.7699	-6.00	-6.39
42	44.11	33.88	11.83	5	4.00	4.00	-0.0953	-0.06	-0.13
43	54.34	44.11	11.83	6	5.00	5.33	-1.3471	-1.39	-1.34
44	37.9	83.23	11.83	6	1.63	5.33	-0.6463	-0.82	-0.64
45	64.57	54.34	11.83	7	6.00	6.00	-2.7181	-2.76	-3.03
46	59.87	67.7	9.63	7	4.17	6.00	-1.9951	-1.98	-1.95
47	74.8	64.57	11.83	8	7.00	7.20	-4.0745	-4.10	-4.79
48	85.03	74.8	11.83	9	8.00	8.00	-5.4015	-5.45	-5.96
49	95.26	85.03	11.83	10	9.00	9.14	-6.9078	-6.80	-6.84
50	105.49	95.26	11.83	11	10.00	10.00	-8.2208	-8.15	-7.28

5. Acknowledgements

The authors thank CNRST (Centre National pour la Recherche Scientifique et Technique - Morocco) and MESRSFC (Ministère de l'Enseignement Supérieur et de la Recherche Scientifique et de la Formation des Cadres – Morocco) for their financial support.

6. References

1. Catana C, Gao H, Orrenius C, Stouten PFW. Linear and Nonlinear Methods in Modeling the Aqueous Solubility of Organic Compounds," *J. Chem. Inf. Model.* 2016; 56(6):1127-1131,.
2. Winkler DA, "The role of quantitative structure--activity relationships (QSAR) in biomolecular discovery," *Brief. Bioinform.* 2002; 3(1):73-86
3. Kauppi AM, Andersson CD, Norberg HA, Sundin C, Linusson A, Elofsson M, *et al*. "Inhibitors of type III secretion in Yersinia: Design, synthesis and multivariate QSAR of 2-arylsulfonylamino-benzanilides," *Bioorganic Med. Chem.* 2007; 15(22):6994-7011.
4. Roy K. *Advances in QSAR Modeling: Applications in Pharmaceutical, Chemical, Food, Agricultural and Environmental Sciences*, 2017, 24.
5. Javaid M, Cao J. "Computing topological indices of probabilistic neural network," *Neural Comput. Appl. 2018;* 30(12):3869-3876.
6. [7] B. Ren, "Novel atomic-level-based AI topological descriptors: Application to QSPR/QSAR modeling," *J. Chem. Inf. Comput. Sci.*, vol. 42, no. 4, pp. 858–868, 2002.
7. Moreau G, Broto P. *Nouv J Chim*. 1980; 4:757-764
8. Rouvray DH, "The modeling of chemical phenomena using topological indices," *J. Comput. Chem.* 1987; 8(4):470-480.
9. Matsutani S. "The physical meaning. 1993; 26(1):5133–5143.
10. Randić M. "On molecular branching," *Acta Chim. Slov.* 1997; 44(1):57-77.
11. Nowell I. "Molecular connectivity in structure-activity analysis," *Endeavour*. 1986; 10(4):216-217.
12. Chastrettep M, Cretin D. "Structure-property relationships-determination of the vapor pressure of hydrocarbons and oxygenated compounds using multifunctional autocorrelation method (MAM)," *SAR QSAR Environ. Res.* 1995; 3(2):131-149.
13. Chastrette M, Tiyal F, Peyraud JF. C R *Acad Sci Paris* Ser. 1990; 310:514–515.
14. Zakarya D, Tiyal F, Chastrette M. *J Phys Org*. 1993; 6:574-582.
15. Nohair M, Zakarya D, Berrada A. "Autocorrelation method adapted to generate new atomic environments: Application for the prediction of 13-C chemical shifts of alkanes," *J. Chem. Inf. Comput. Sci.* 2002; 42(3):586-591.
16. Hall LH, Kier LB. "The Molecular Connectivity Chi Indexes and Kappa Shape Indexes in Structure-Property Modeling. 2007; 2:367-422.
17. Amić D, Basak SC, Lučić B, Nikolić S, Trinajstić N. "Structure-water solubility modeling of aliphatic

alcohols using the weighted path numbers," *SAR QSAR Environ. Res.* 2014; 13(2):281-295.

18. Weininger D, "J. Chem. Inj Comput.," *J. Chem. Inf. Comput. Sci.* 1988; 28(31):36.
19. Rumelhart DE, Hinton GE, Williams RJ. 1986; 323:533-536,.
20. Hinton GE, Srivastava N, Krizhevsky A, Sutskever I, Salakhutdinov RR. "Improving neural networks by preventing co-adaptation of feature detectors, 2012, 1-18.
21. Lecun Y, Bengio Y, Hinton G. "Deep learning," *Nature*. 2015; 521(7553):436-444.
22. Siregar SP, Wanto A. "Analysis of Artificial Neural Network Accuracy Using Backpropagation Algorithm In Predicting Process (Forecasting)," *IJISTECH (International J. Inf. Syst. Technol.* 2017; 1(1):34.
23. Zhang Z. "Variable selection with stepwise and best subset approaches," *Ann. Transl. Med.* 2016; 4(7):1-6,
24. Lv J, Liu Y, Zhang Z, Dai J. "Factorial kriging and stepwise regression approach to identify environmental factors influencing spatial multi-scale variability of heavy metals in soils," *J. Hazard. Mater.* 2013; 261:387-397.
25. Templ M, Kowarik A, Filzmoser P. "Iterative stepwise regression imputation using standard and robust methods," *Comput. Stat. Data Anal.* 2011; 55(10):2793-2806.
26. "As is well known, a fundamental difficulty in statistical analysis is the choice of an appropriate model, estimating and determining the order or dimension of a model. This is a common problem when a statistical model contains many parameters. The mai. 1987; 52(3):345-370.
27. Akalke H. "ANew Look at the Statistical Model Identification. 1974; 19:716-723.
28. Ingram T, Mahler DL, "SURFACE : detecting convergent evolution from comparative data by fitting Ornstein-Uhlenbeck models with stepwise Akaike Information Criterion, 2013, 416-425.
29. Dirick L, Claeskens G, Baesens B. "An Akaike information criterion for multiple event mixture cure models, 2014, 1-28.
30. Islam MA, Rahman MT, Islam AN, Ara MH, Dey SK. Volumetric and viscometric studies of l-proline in aqueous solution of alcohols at different temperatures.