Turkish Journal of Computer and Mathematics Education Vol.12 No.10 (2021), 3992-3999

Research Article

QSPR Analysis of Novel Indices with Priority Polycyclic Aromatic Hydrocarbons(PAHs)

C.Tamilarasi[a], and F. Simon Raj[b]

a,b
Hindustan Institute of Technology and Science, Chennai, Tamilnadu, India.

Article History: Received: 10 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online:28April 2021

Abstract: In this paper, proposed degree-based topological index,*FORAN* index = $\sum_{u,v\epsilon E}\left(\deg(u)\sqrt{\frac{\deg(u)}{\deg(v)}}+\deg(v)\sqrt{\frac{deg(v)}{\deg(u)}}\right)$and its corresponding neighborhood degree-sum index *NFORAN* index = $\sum_{u,v\epsilon E}\left(S(u)\sqrt{\frac{S(u)}{S(v)}}+S(v)\sqrt{\frac{S(v)}{S(u)}}\right)$are established by known property dependence relationships(Degeneracy test with eighteen octane isomers). And developed linear models based on n-octane isomers &sixteen priority Polycyclic Aromatic Hydrocarbons(*PAH*s) by*QSPR* study.

Keywords: Topological indices, physico-chemical properties, QSPR study.

1. Introduction

Chemical graph theory is a branch of mathematics that combines mathematical graph theory and chemistry. The molecular graph is an appropriate model for any molecule in a chemical transformation. Chemical compounds are expressed as molecular graph$G = (V,E)$. It is a simple hydrogen-suppressed graph, representing the carbon atom skeleton of an organic molecule, where the vertex$v\epsilon V$represents non-hydrogen atom and the edge $e(u,v)\epsilon E$implies covalent bond between two atoms.

Topological indices are representing the molecular descriptors that are numbers linked with intermolecular relationships and biological activities of chemical compounds. One of the main branches of the topological indices isthe degree based topological indiceswhich are widely used in structure-property relationships[13] and attendant the theoretical basis of manufacturing of drugs. Product connectivity index or Randic index$R(G)$ was the first degree based topological indices defined by Milan Randic in 1975. It is very much popular in drug designing and is mathematically connected with normalized Laplacian matrix[4,12]. Similar toanother notable degree-based topological index -Forgotten index[5]$F(G)$ was defined to test the Physico-chemical, pharmacological properties of the drug molecular structures. Based on their common phenomenon,Physico-chemical & pharmacological modeling of the molecular structure of drugs, the new indices have been defined such as *FORAN* index $FR(G)$and its neighborhood degree sum topological index*NFORAN* index$NFR(G)$which are holding properties of parent indices.In *QSPR*[6,10,11,14,17] study, the linear regression models based on n-octane isomers[8,9] and sixteen priority *PAH*s[7,15,16]are used as tools to predict Physico-chemical properties of the related chemical compounds.

Definitions

Let $G = (V,E)$ is a simple, connected graph that contains vertex set V, as well as edges, set E. The degree of a vertex is the number of edges connect to u and is denoted as $\deg(u)$ where neighborhood degree sum is the summation of the degree of neighborhood vertices of vertex $u$and is denoted as $S(u)$. $FR(G)$ index is the combination of Randic index $R(G) = \sum_{u,v\epsilon E}\frac{1}{\sqrt{\deg(u)\deg(v)}}$ as well as Forgotten index $F(G) = \sum_{u,v\epsilon E}(\deg(u)^2 + \deg(v)^2)$and is defined as

FORAN index $FR(G)= \sum_{u,v\epsilon E}\left(\deg(u)\sqrt{\frac{\deg(u)}{\deg(v)}}+\deg(v)\sqrt{\frac{deg(v)}{\deg(u)}}\right)$

A similar process is followed in defining *NFORAN* index and is denoted as

NFORAN index $NFR(G)= \sum_{u,v\epsilon E}\left(S(u)\sqrt{\frac{S(u)}{S(v)}}+S(v)\sqrt{\frac{S(v)}{S(u)}}\right)$

Results and Discussions

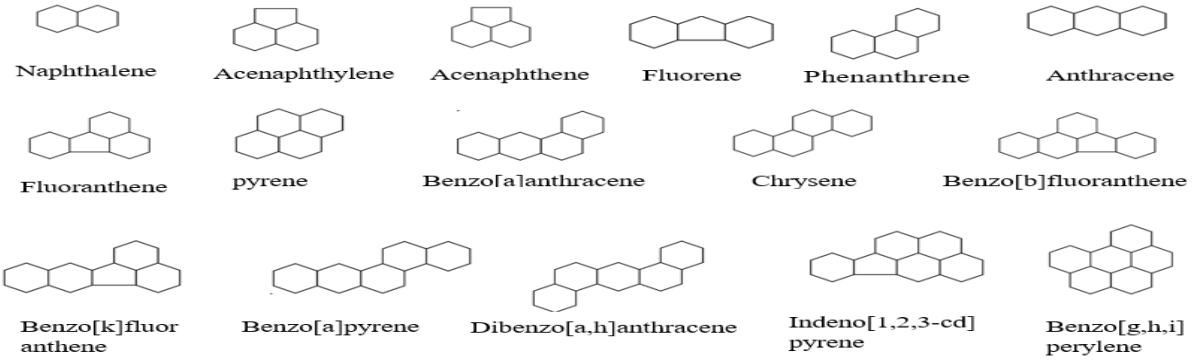
This section contains three parts. The first part shows the discrimination level of proposed novel indices through the degeneracy test. The second part shows, comparison analysis of the novel indices with an existing notable degree-based topological index. The third partshows *QSPR* analysis of *n*-octane isomers and sixteen priority *PAH*s.

Part 1

Degeneracy Test(Sensitivity Test): Eighteen octane isomers are having the same molecular formula,molecular weight, and molecular structure difference.Additionally, these isomers are large enough to create statistical inference reliable.

D(Topological index) = $\frac{A-B}{A}$ where *A*epresentsthe number of isomers in the particular class of isomers and *B*represents the number of isomers who possess the same topological values.The sensitivity test shows that the proposed indices with octane isomers produce the high discriminate level of outcomes. *FR*(*G*) shows 0.88889 and *NFR*(*G*) shows 1.00000(Table 1). These results help us to move forward towards a *QSPR* study based on these indices with *n*-octane isomers and *PAH*s.

Figure 1. shows the molecular structures of sixteen priority *PAH*s.



Octane isomers	*FORAN*(*G*)	*NFORAN*(*G*)	*S*	*A.F*	*HVAP*	*DHVAP*
n-octane	27.07106	49.04822	111.67	0.397898	73.19	9.915
2-methyl heptane	32.38976	53.29363	109.84	0.377916	70.30	9.484
3-methyl heptane	31.45902	56.7097	111.26	0.371002	71.3	9.521
4-methyl heptane	31.45902	56.95074	109.32	0.371504	70.91	9.483
3-ethyl hexane	30.52828	60.59408	109.43	0.362472	71.70	9.476
2,2-dimethyl hexane	44.1066	61.77924	103.42	0.339426	67.70	8.915
2,3-dimethyl hexane	36.16326	64.52417	108.02	0.348247	70.20	9.272
2,4-dimethyl hexane	36.77772	61.20216	106.98	0.344223	68.50	9.029
2,5-dimethyl hexane	37.70846	57.20328	105.72	0.356830	68.60	9.051
3,3-dimethyl hexane	42.2132	69.45039	104.74	0.322596	68.50	8.973
3,4-dimethyl hexane	35.23252	68.58426	106.59	0.340345	70.20	9.316
2-methyl-3-ethyl pentane	35.23252	71.09961	106.06	0.332433	69.70	9.209
3-methyl-3-ethyl pentane	40.3198	77.32052	101.48	0.306899	69.30	9.081
2,2,3-trimethyl pentane	47.33314	76.97072	101.31	0.300816	67.30	8.826
2,2,4-trimethyl pentane	49.4253	66.72953	104.09	0.305370	64.87	8.402
2,3,3-trimethyl pentane	46.37048	80.57557	102.06	0.293177	68.10	8.897
2,3,4-trimethyl pentane	40.8675	72.78824	102.39	0.317422	68.37	9.014
2,2,3,3-tetramethyl butane	59	87.7031	93.06	0.255294	66.20	8.410

Table 1. shows the results of *FORAN*(*G*) and *NFORAN*(*G*) indices and the experimental values(www.moleculardescriptors.eu) of n-octane isomers.

Part 2

Comparative analysis

The investigation of this part reports that, the correlation of the parent indices with the *FORAN* index and the *NFORAN* index(Tables 2 & 3). The *FORAN* index is highly correlated with the Randic index as well as the Forgotten index($|r|$ = 0.99 app.). Likewise, the *NFORAN* index displays a very good correlation with neighborhood version of the Randic index and the Forgotten index (($|r|$ = 0.94 app.). From the above result, we will comprehend that the novel indices also play important role instructure-property relationships in related series of chemical compounds, biological activities of chemical and drug designing. The NForgotten index is the neighborhood version of Forgotten index(Using edge partition method) whereas the NForgotten* index is the neighborhood version of Forgotten index(Using vertex partition method).

Table 2 shows the cross- correlation matrix of *FORAN* index, Randic index and Forgotten index.

Topological Indices	*FORAN* Index	Randic Index	Forgotten Index
FORAN Index	1.0000		
Randic Index	-0.9843	1.0000	
Forgotten Index	0.9907	-0.9596	1.0000

Table 3 shows the cross-correlation matrix of neighborhood version of the mentioned above topological indices.

Neighborhood Topological Indices	*NFORAN* Index	NRandic Index	NForgotten* Index	NForgotten Index
NFORAN Index	1.0000			

NRandic Index	-0.942	1.0000		
NForgotten* Index	0.9491	-0.9236	1.0000	
NForgotten Index	0.9743	-0.9686	0.9311	1.0000

Part 3

(a) Novel indices in *QSPR* analysis(Octane isomers)

The correlation coefficients of the novel indices *FORAN*(*G*) and *NFORAN*(*G*) with Entropy(*S*), Acentric Factor(*A.F*), Enthalpy of vaporization(*HVAP*) and Standard enthalpy of vaporization(*DHVAP*) are remarkably high(Table 4). These results lead to do *QSPR* analysis of novel indices with the physicochemical properties of n-octane isomers.

The linear regression formulafor n-octane isomers are as follows:

Ꝑ = m(Topological indices) + c, where Ꝑ represents the properties of isomers, m & c represent slope & intercepts respectively.

With the help of the above formula, the linear regression models for *FORAN*(*G*) and *NFORAN*(*G*) are represented here.

Table 4. shows the correlation coefficients of *S*, *A.F*, *HVAP* and *DHVAP* of n-octane isomers with novel indices.

Index	*S*	*A.F*	*HVAP*	*DHVAP*
FORAN(*G*)	-0.9317	-0.9321	-0.9117	-0.9482
NFORAN(*G*)	-0.8928	-0.9574	-0.6322	-0.5959

Theoretically,if the absolute correlation coefficient value(|r|) of topological index with chemical compounds is less than 0.8, then topological index is not useful to predict the physicochemical properties of the chemical compounds.Based on this, linear regression models are not done as *NFORAN* index shows a moderate correlation with *HVAP* and *DHVAP*.

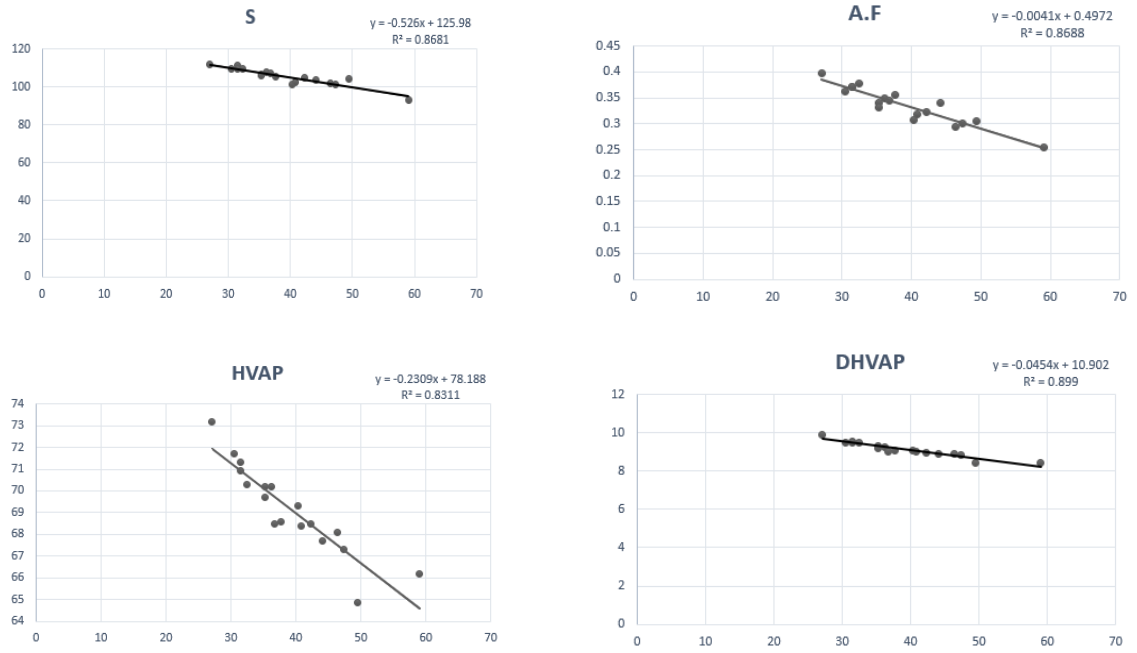
FORAN(*G*) index

1. *S* = -0.526(*FORAN*(*G*)) + 125.98
2. *A.F* = -0.0041(*FORAN*(*G*)) + 0.4972
3. *HVAP* = -0.2309(*FORAN*(*G*)) + 78.188
4. *DHVAP* = -0.0454(*FORAN*(*G*)) + 10.902

NFORAN(*G*) index

1. *S* = -0.3932(*NFORAN*(*G*)) + 131.46
2. *A.F* = -0.0033(*NFORAN*(*G*)) + 0.555

Figure 2. shows the fitted graphs of linear regression models of *FORAN*(*G*) index with Entropy(*S*), Acentric Factor(*A.F*), *HVAP*&*DHVAP* of n-octane isomers.



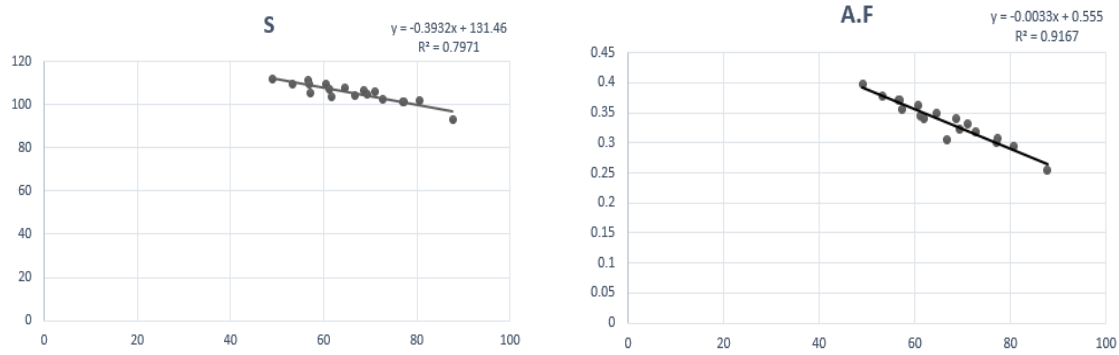


Figure 3. shows the fitted graphs of linear regression models of *NFORAN*(*G*) index with Entropy(*S*), Acentric Factor(*A.F*) of n-octane isomers.

(b) Novel indices in *QSPR* analysis(*PAH*s)

Theresults of the *FORAN* index and the *NFORAN* index with sixteen priority *PAH*s &the Physico-chemical properties(boiling point,melting point,molecular weight, Solubility in water-C_w^{sat}, n-octanol water partition coefficient-*Log* K_{ow}, n-octanol air partition coefficient-*Log* K_{oa}) are shown in Table 5. The correlation coefficients of *FR*(*G*) &*NFR*(*G*) with bp, mw, *Log*K_{ow}, *Log* $K_{oa}$of *PAH*s are significantly high(Table 6). Based on this, the linear regression models are formed.

Table 5.

*PAH*s	*FR*(*G*)	*NFR*(*G*)	*bp* (℃)	*mp* (℃)	mw g/mol	C_w^{sat} m g/l	Log K_{ow}	Log K_{oa} 25℃
Naphthalene	51.22891	116.70464	218	80.2	128	31.0	3.5	5.13
Acenaphthylene	69.84337	170.85658	280	92.5	152	16.1	3.85	6.34
Acenaphthene	69.84337	170.85658	279	93.4	152	3.8	3.92	6.52
Fluorene	73.84337	178.26446	295	115	166	1.9	4.11	6.90
Phenanthrene	77.84337	187.04944	340	99.2	178	1.1	4.47	7.68
Anthracene	78.45782	183.1678	340	215	178	0.045	4.45	7.71
Fluoranthene	95.84337	242.86768	384	108	202	0.26	4.90	8.76
Pyrene	96.45782	240.2516	404	151	202	0.132	5.18	8.61
Benz[a]anthracene	105.07228	253.26671	435	167	228	0.011	5.91	10.28
Chrysene	104.45782	257.19372	448	258	228	0.0015	5.79	10.30
Benzo[b] fluoranthene	122.45782	295.7913	481	168	252	0.0015	5.78	11.34
Benzo[k] fluoranthene	123.07228	308.83906	480	217	252	0.0008	6.11	11.37
Benzo[a]pyrene	123.07228	309.42111	495	177	252	0.0038	6.35	11.56
Dibenzo[a,h] anthracene	131.68673	294.4981	524	270	278	0.0005	6.75	12.59
Indeno[1,2,3-cd]pyrene	141.07228	328.29109	536	164	276	0.062	6.70	12.55
Benzo[g,h,i] perylene	141.07228	364.5727	550	278	276	0.00026	6.90	12.43

Table 6.

TIs	bp (℃)	Mp (℃)	mw (g/mol)	C_w^{sat} m g/l	Log K_{ow}	Log K_{oa}25℃
FORAN	0.9914	0.7294	0.9929	-0.6233	0.9807	0.9902
NFORAN	0.9797	0.7161	0.9724	-0.6325	0.9648	0.9694

FORAN(*G*) index

1. bp = 3.66522(*FORAN*(*G*)) + 37.82051
2. mw = 1.7478(*FORAN*(*G*)) + 37.13825
3. Log K_{ow} = 0.0397(*FORAN*(*G*)) + 1.3045
4. Log K_{oa} = 0.0873(*FORAN*(*G*)) + 0.6206

NFORAN(*G*) index

1. bp = 1.4518(*NFORAN*(*G*)) + 51.522
2. mw = 0.6861(*NFORAN*(*G*)) + 45.19
3. Log K_{ow} = 0.0157(*NFORAN*(*G*)) + 1.4703
4. Log K_{oa}= 0.0343(*NFORAN*(*G*)) + 1.0253

The results of the above linear models are shown in Tables 6-10. Here, we have used the terms n, m, c, |r|, *SE*, *F*, *SF* stands for the number of chemical compounds, slope, intercept, absolute value of correlation coefficient, standard error, *F*-test and significance of *F* respectively and are shown in figures 2-5.

Figure 4. shows the fitted graphs of linear regression models of *FR*(*G*) with *bp*, *mw*, *Log* K_{ow}&*Log* $K_{oa}$of *PAH*s.

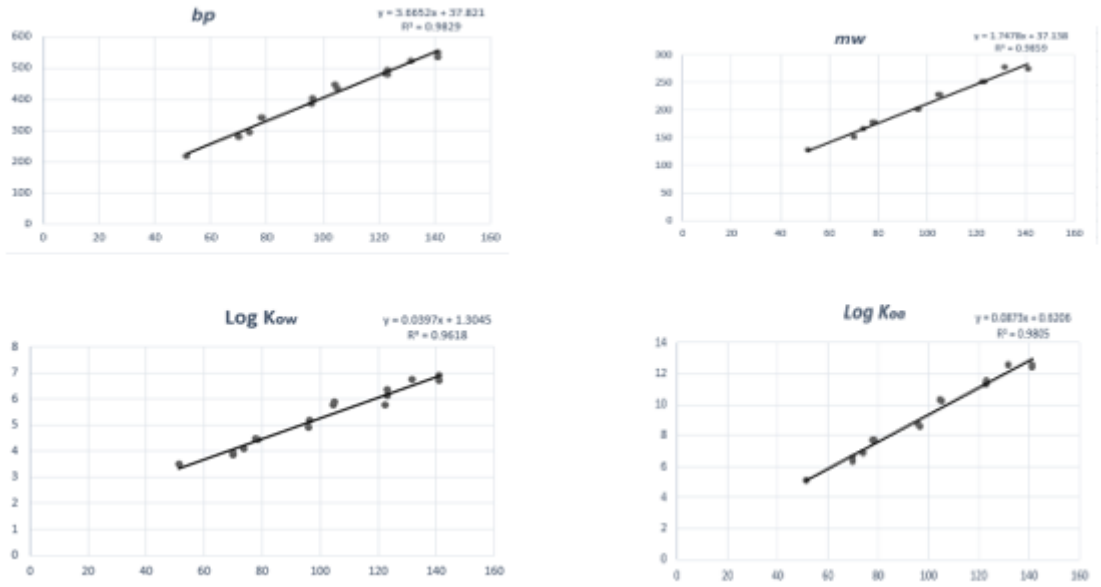
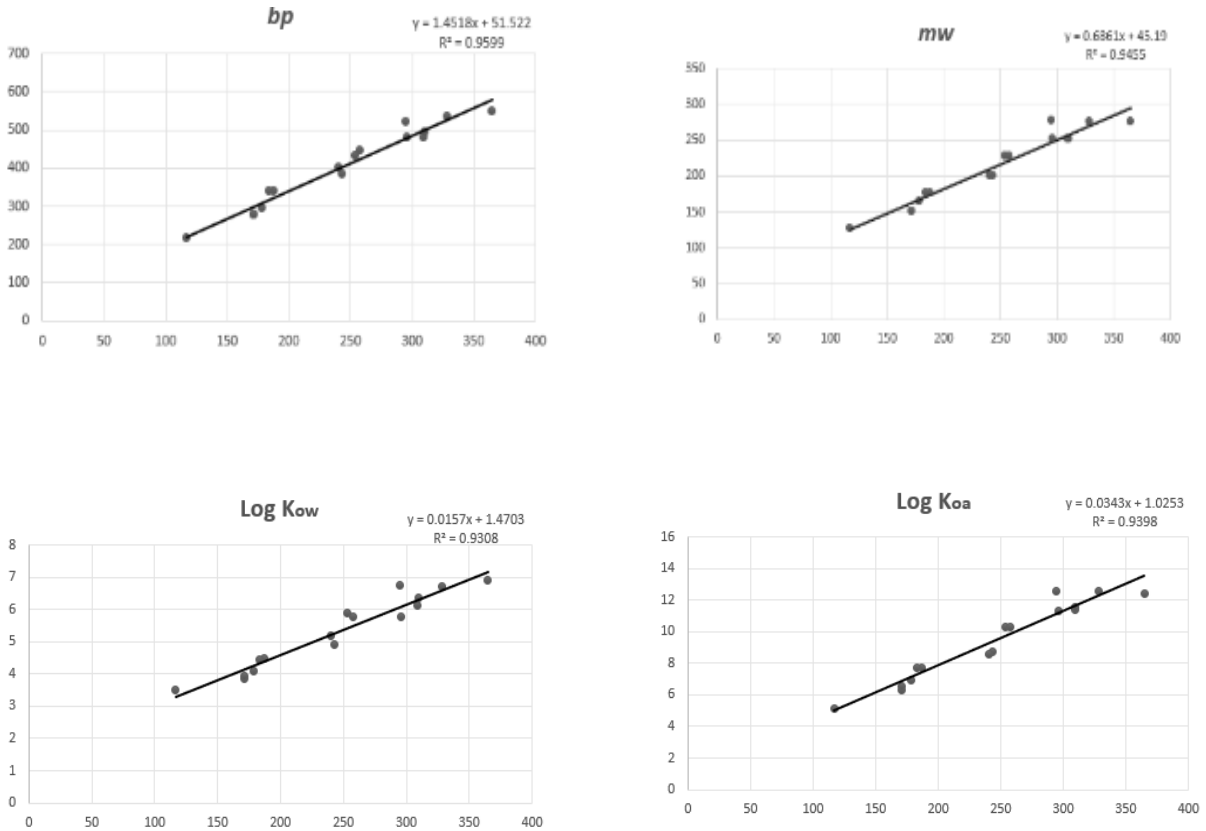


Figure 5. shows the fitted graphs of linear regression models of *NFR*(*G*) with *bp*, *mw*, *Log* K_{ow}&*Log* $K_{oa}$of *PAH*s.



Concluding Remarks

Based on the linear regression analysis of *FORAN*&*NFORAN* indices with octane isomers and *PAH*s, the following results are drawn:

Table 7. shows the statical parameters for the linear *QSPR* model for *FORAN*(*G*) with octane isomers.

Physical Properties	n	m	c	|r|	*SE*	*F*	*SF*
Entropy	18	-0.526	125.95	0.93173	1.69106	105.3222	1.91E-08
Acentric Factor	18	-0.0041	0.4972	0.93211	0.01324	105.9703	1.83E-08
Enthalpy of vaporization	18	-0.2309	78.188	0.91167	0.85828	78.75032	1.41E-07
Standard enthalpy of vaporization	18	-0.0454	10.902	0.94816	0.12557	142.4229	2.23E-09

Table 8. shows the statical parameters for the linear *QSPR* model for *NFORAN*(*G*) with octane isomers.

Physical Properties	n	m	c	|r|	*SE*	*F*	*SF*
Entropy	18	-0.3932	131.46	0.89283	2.09729	62.8757	6.22E-07
Acentric Factor	18	-0.0033	0.555	0.95743	0.01055	175.9987	4.75E-10

Table 9. shows the statical parameters for the linear *QSPR* model for *FORAN*(*G*) with sixteen priority *PAH*s.

Physical Properties	n	m	c	|r|	*SE*	*F*	*SF*
Boiling point	16	3.66522	37.82051	0.99143	14.05653	806.1693	8.19E-14
Molecular weight	16	1.7478	37.13825	0.99292	6.08608	977.8895	2.35E-14
Log K_{ow}	16	0.0397	1.3045	0.98070	0.230557	352.3044	2.54E-11
Log K_{oa}	16	0.0873	0.6206	0.99021	0.35810	704.6674	2.25E-13

Table 10. shows the statical parameters for the linear *QSPR* model for *NFORAN*(*G*) with sixteen priority *PAH*s.

Physical Properties	n	m	c	|r|	*SE*	*F*	*SF*
Boiling point	16	1.4518	51.522	0.97973	21.55268	334.8647	3.57E-11
Molecular weight	16	0.6861	45.19	0.97238	11.95769	242.9476	3.06E-10
Log K_{ow}	16	0.0157	1.4703	0.96478	0.31024	188.3033	1.64E-09
Log K_{oa}	16	0.0343	1.0253	0.96944	0.62944	218.6005	6.16E-10

The absolute correlation coefficient(|r|)shows the eligibility level of the topological index as a tool, to predict the Physico-chemical properties of chemical compounds in *QSPR* analysis. In this point of view, the *FORAN*index shows a high correlation with entropy, acentric factor, *HVAP*&*DHVAP* of n-octane isomers(|r| ranging from 0.91 to 0.95). Correspondingly, inpredicting the physicochemical properties of *PAH*s such as boiling point, molecular weight, *Log* K_{ow}&*Log* K_{oa}, *FORAN* index is a more suitable tool as such(|r| ranging from 0.98 to 0.99).By the observation of Tables 8 & 10, the *NFORAN* index plays important role in predicting the physicochemical properties of n-octane isomers and *PAH*s(|r| ranging from 0.89 to 0.98).

These two novel indices are holdingsignificantly high diverse structural formulae(Degeneracy test). This discriminating level of structural information stimulated us to study the structure-property relationships of n-octane isomers as well as sixteen priority *PAH*s. The proposed indices(*FORAN* index, *NFORAN* index) are very much useful in structure-property relationships, and the results are shown in Tables 7-10.

These novel indices will show promising application prospects in chemical and pharmacological fields.

References

1. J. A. Bondy and U.S.R. Murty, *Graph Theory*, Spring, Berlin, Germany, **2008**.
2. H.J. Wiener, "Structural determination of paraffin boiling points" *Journal of the American Chemical Society*, vol. 69, no. 1, pp. 17-20,**1947**.
3. Todeschini, *Handbook of Molecular Descriptors*, Wiley-VCH, Weinheim, pp. 120-121, **2000**.
4. Gutman, Degree-Based Topological Indices, *Croat, Chem, Acta*, 86,(4), pp. 351-361, **2013**.
5. Furtula, B.; Gutman, I.A forgotton topological index. *Journal of Math. Chem*., 53, pp. 1184-1190, **2015**.
6. Devillers. A. T. Balaban, Topological indices and related descriptors in QSAR and QSPR, *God*. &*Bre*. *Sci*. *Pub*., Netherland, **1999**.
7. Dias, Handbook of Polycyclic Hydrocarbons, *Elsevier*, Amsterdam,**1987.**
8. *CRC Handbook of Chemistry and Physics*(CRC Press, Boca Raton, **1995**.
9. Sunilkumar M. Hosamani, Correlation of domination parameters with physicochemical properties of octane isomers, 10.13140/*RG*.2.1.3808.4004, **2015**.
10. Sunilkumar Hosamani, Deepa, QSPR Analysis of Certain Degree Based Topological Indices, *J*. *Stat*. *App*. *Pro*.6, No.2, pp. 1-11, **2017**.
11. Kier, Hall, Molecular Connectivity in structure-activity analysis(Research Studies Press- Wiley) Chichester,**1986**.

12. Li, I. Gutman, Mathematical aspects of Randic-type molecular structure descriptors, *Univ*. *Kragujevac*, **2006**.
13. Mihalic, Trinajstic, A graph theoretical approach to structure property relationships. *J*.*Chem*.*Educ*.69, 701-712, **1992**.
14. Sourav Mondal, Nilanjan De, QSPR analysis of some novel neighbourhood degree based topological descriptors, *arXiv*: 1906.06660v1 [cs.DM],**2019**.
15. Abiodun Olagoke Adeniji; Omobola Oluranti Okoh; Anthony Ifeanyi Okoh, Analytical Methods for polycyclic Aromatic Hydrocarbons and their Global Trend of Distribution in Water and Sediment, Recent Insights in Petroleum Science and Engineering, 3138, **2018**.
16. Karishma Hussain; Raza Rafiqul Hoque, Monitoring and Risk Analysis of PAHs in the Environment, Handbook of Environmental Materials Management, pp. 1-35, **2018**.
17. Gozalbes, R.; Doucet,J.P.; Derouin,F., Application of Topological Descriptors in QSAR and Drug Design: *History and New Trends, Current Drug Targets- Infectious Disorders*, *Vol*. 2, no. 1, pp. 93-102, **2002**.
18. Peter, J. Hansen; Peter, Chemical applications of graph theory. Part 1. Fundamentals and topological indices, J.*Chem*. *Educ*. 65,7,574, **1988**.
19. Balban, A.T., Chemical applications of graph theory, Academic Press, **1976**.
20. Randic, On characterization of molecular branching, *J*. *Am*. *Chem*. *Soc*. 97, pp. 6609-6615, **1975**.