CREAT. MATH. INFORM. Volume **25** (2016), No. 2, Pages 151 - 158 Online version at https://creative-mathematics.cunbm.utcluj.ro/ Print Edition: ISSN 1584 - 286X; Online Edition: ISSN 1843 - 441X DOI: https://doi.org/10.37193/CMI.2016.02.05

Development of a QSPR model for predicting octane number of alkanes using SD topological index

ZOIŢA-MĂRIOARA BERINDE, CLAUDIA BUTEAN and THOMAS DIPPONG

ABSTRACT. The octane number (ON, MON and PON) for the molecular structures of 18 octane isomers have been correlated using the quantitative structure-property relationship (QSPR) method, with topological index SD. For single parameter correlation the index SD shows poor results (RON, r = 0.406; MON, r = 0.490; PON, r = 0.448), whereas for two-parameter correlation almost any combination among the above D_C was found to give relatively high r value. The best correlation coefficients are as follows: for RON, r = 0.993; MON, r = 0.968; PON, r = 0.985. For RON, the best model obtained by our regression analysis is

 $RON = -227.218 + 7.63 * SD - 37.111 * D_{C_{t}}$ with r = 0.993, s = 4.8, F = 534.

1. INTRODUCTION

The octane number (ON) represents a conventional measure that is used to describe the quality of benzine from the point of view of its resistance to detonation, for the sparkignition engines. Benzine is obtained by the oil fractionate distillation and contains mostly acyclic saturated hydrocarbons, linear or branched (called alkanes or paraffins), but also cyclical saturated hydrocarbons (called cycloalkanes, cycloparaffins or naphthenes) and aromatic hydrocarbons in smaller quantities. Based on the oil origin and on the obtention method, the chemical composition of benzines is very different.

The detonation occurs by the autoignition of the benzine-air mixture in front of the spark produced by the plug and followed by the abnormal combustion of the fuel at an explosive speed. Under these circumstances, the pressure increase due to the sudden heat release takes place so fast that it acts out of time on the piston, before it gets to the end of the compression stroke or is lost against the piston inertia. These leads to a decrease of the engine power.

It has been found out that hydrocarbons (containing between 4 and 12 carbons atoms in the molecule) that are part of benzines have a different behaviour in the engine as far as the detonation phenomenon is concerned. Normal alkanes (with linear chain) cause a powerful detonation in the engine, whilst the isoalkanes (with branched chain) and the aromatic hydrocarbons prevent this phenomenon. The resistance to detonation decreases as the length of the carbon chain increases and increases as the carbon chain develops more branches. Therefore, a scale has been set, with limits between 0 and 100, randomly selected, regarding the resistance to detonation of a fuel or of the octane number.

The higher the resistance of a fuel to detonation is, the higher its octane number. The heptane, with linear chain, has a weak resistance and determines a strong detonation in the engine, therefore is has been assigned an octane number of zero. The isooctane or 2,2,4-trimethylpentane, a highly branched compound, is little sensitive to auto-ignition, burns smoothly, with very small detonation. The isooctane has been assigned the octane

2010 Mathematics Subject Classification. 92E10.

Received: 23.02.2016. In revised form: 19.05.2016. Accepted: 22.05.2016

Key words and phrases. Motor octane number (MON), research octane number (RON), posted octane number (PON), topological index SD, isooctanes, regression analysis.

Corresponding author: Zoița-Mărioara Berinde; zoitaberinde@ubm.ro

number of 100. As there are fuels that presents an even higher resistance to detonation than the isooctane, the definition has been extended in order to be possible to assign them an octane number > 100. Examples of such fuels are to be found in the field of car races or aviation benzine (kerosene), ethanol (which has the octane number of 129).

As it contains a mixture of branched-chain hydrocarbons, benzine may have different octane number. In order to assess the octane number of a fuel, an experimental testing engine is used in the same standard conditions. The resistance to detonation of the tested benzine, namely the octane number, is compared to the resistance to detonation of a mixture of isooctane and heptane. The volume percentage of isooctane of this mixture represents the octane number of the mixture and at the same time the octane number of the tested benzine. For example, the benzine with the same anti-compression features as those of a mixture of 95% isooctane and 5% heptanes has an octane number of 95. This leads us to say that benzine with octane number 95 does not contain isooctane and heptanes at the given rates, but that it has the same resistance to detonation as the mixture made of 95% isooctane and 5% heptane.

According to the speed of the engine used experimentally, there are two types of octane numbers:

- Octane number determined by the Research method (Research Octane Number, RON), with motor speed 600 rpm
- Octane number determined by the Motor method (Motor Octane Number, MON), with motor speed 900 rpm

In Australia, Europe and therefore in Romania, too, the gas stations display the octane number determined by the Research method (RON) with 600 rpm motor speed. In the USA and Canada, the octane number displayed at the pump is the average RON-MON, abbreviated PON and called pump octane number. Hence PON = (RON + MON)/2 is the current ON that is reported at the pump.

In Romania, gas stations sell benzine with octane number determined by the Research Method-RON 95, 98, 99, 100.

The octane number may also be increased by adding different substances such as tetraethyl lead (TEP), methyl-tert-butyl-ether (MTBE) or toluene (TL). Because of its toxicity, the use of TEP as a benzine additive has been limited after 1970.

In the present paper, we have approached the categories of octane numbers, RON, MON and PON for 18 molecular structures representing octane isomers. Using the topological descriptors SD and D_C we have elaborated a mathematical model for the calculation of the octane number of alkanes.

2. The SD topological index

In a series of papers published in the last years [7]-[18], the first author introduced and studied the concept of *weighted electronic distance*, denoted *w.e.d.*, as an edge invariant in a chemical graph. Then, using the concept of weighted electronic distance, we defined the so-called weighted electronic connectivity matrix, CEP(G), of a chemical graph *G*. For a molecule having *N* atoms, whose graphs is G = (V(G)), E(G)), the weighted electronic connectivity matrix CEP(G) is a symmetric and quadratic $N \times N$ matrix given by:

$$CEP(G) = \{ [CEP]_{ij} ; i, j \in V(G) \},$$
(2.1)

where V(G) is the set of vertices of the molecular graph G, E(G) is the set of edges, and

$$CEP_{ij} = w.e.d._{(i,j)}$$
, if $i \neq j$ and $(i,j) \in E(G)$ and $CEP_{ij} = 0$, otherwise, (2.2)

where $w.e.d._{(i,j)}$ denotes the weighted electronic distance between the atoms (vertices) i and j. The concept of weighted electronic distance, w.e.d., was defined in [8] by the

Development of a QSPR model...

following formula:

$$w.e.d.(i,j) = \frac{1}{b_r} \cdot \frac{Z'_i + Z'_j}{v_i \cdot v_j},$$
(2.3)

where:

 $-b_r$ is the bond weight with values: 1, for single bond, 2 for double bond, 3 for triple bond and 1.5 for aromatic bond, like in Barysz et al. [6];

 $-v_i, v_j$ denote the degree of vertices *i*, and *j*, respectively;

 $-Z'_i$ denotes the formal degree of vertex *i*, and it is defined by

$$Z_i' = Z_i \cdot v_i, \tag{2.4}$$

where Z_i is the order number of atom *i* in Mendeleev's periodic system, that is, the number of all electrons in the atom *i*.

In Figure 1 it is given the corresponding CEP matrix for the hydrogen-deleted graph (G) of 2,2,4-trimethylpentane

$$CEP(G) = \begin{bmatrix} 0 & 7.5 & 0 & 0 & 0 & 0 & 0 & 0 \\ 7.5 & 0 & 4.5 & 0 & 0 & 7.5 & 7.5 & 0 \\ 0 & 4.5 & 0 & 5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 5 & 0 & 8 & 0 & 0 & 8 \\ 0 & 0 & 0 & 8 & 0 & 0 & 0 & 0 \\ 0 & 7.5 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 7.5 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 8 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 8 & 0 & 0 & 0 & 0 \\ \end{bmatrix} \begin{bmatrix} SEP_i \\ 27 \\ 9.5 \\ 21 \\ 8 \\ 7.5 \\ 7.5 \\ 8 \end{bmatrix}$$

Figure 1. The electronic connectivity matrix CEP(G) of 2,2,4-trimethylpentane

The sum of all entries on the *i*-th row in CEP (G) is denoted by SEP_i (and is written at the matrix right hand side):

$$SEP_i = \sum_{j=1}^{n} [CEP]_{ij}, \ i = 1, 2, \dots, n$$
 (2.5)

From the electronic connectivity matrix we can obtain several new topological indices, see [7]-[18]. The SD index is defined, see [15], by:

$$SD = \sum_{i=1}^{n} \frac{SEP_i}{z_i}.$$
(2.6)

In Figure 2, the values of SEP_i for the hydrogen-suppressed graph G of 2,2,4-trimethylpentane are indicated.



Figure 2. The values of the SEP_i for 2,2,4-trimethylpentane.

153

Hence, the calculation of the SD index for 2,2,4-trimethylpentane is done as follows

$$SD(G) = 3 \cdot \frac{7.5}{1} + 2 \cdot \frac{8}{1} + \frac{27}{4} + \frac{9.5}{2} + \frac{21}{3} = 57.$$

Because ON decreases with the number of CH_2 groups and increases with the number of CH_3 groups [29], we also considered the parameter that we have introduced in [15], denoted by D_C , which is calculated by using the formula:

$$D_C = NC - NP, \tag{2.7}$$

where

- *NC* represents the number of carbon atoms in the molecular graph;
- *NP* represents the number of primary carbon atoms in the molecular graph.

The calculation technique of parameter D_C is illustrated in Figure 3 for the hydrogensuppressed graph G of 2,2,4-trimethylpentane:



 $D_C = 8$ (carbon atoms) -5 (primary carbon atoms) = 3

Figure 3. Calculation of D_C for the hydrogen-suppressed graph of 2,2,4-trimethylpentane.

Note that SD and D_C have been used in [15] in order to develop quantitative structureproperty models for estimation of the normal boiling points of alkanes using two-variable linear regression analysis.

3. RESULTS AND DISCUSSION

The values of *SD* index were calculated for a set of 18 isooctanes and are presented in Table 1. The values RON and MON for octane number were taken from [3], [23], [22], while and PON was calculated with the formula

$$PON = \frac{RON + MON}{2}.$$

The structures of alkanes are presented in shorthand, e.g., 8 is n-octane, 224MMM5 is 2,2,4-trimethylpentane, 3E3M5 is 3-ethyl-3-methylpentane etc.

The two-variable linear equations obtained by regression analysis for predicting RON, MON and PON was well as their statistical parameters are given in the following:

$$RON = -227.218 + 7.63SD - 37.111D_C; r = 0.993, s = 4.8, F = 534,$$
 (3.8)

$$MON = -410.18 + 10.318SD - 29.952D_C; r = 0.968, s = 8.9, F = 113,$$
 (3.9)

$$PON = -318.699 + 8.974SD - 33.532D_C; r = 0.985, s = 6.5, F = 252.$$
 (3.10)

The values of observed and calculated RON, MON and PON by means of the previous QSPR models are presented in Table 1 for all the 18 isooctanes.

Table 1. Observed and calculated RON, MON and PON

Development of a QSPR model...

N	Alkanes	RON	MON	PON	SD	D_C	RON	MON	PON
		obs.	obs.	obs.			Calc.	Calc.	Calc.
1	8	-19.0	-19.0	-19.0	57.00	6	-14.97	-17.66	-8.37
2	2M7	21.7	23.8	22.75	57.00	5	22.14	28.18	25.15
3	3M7	26.8	35.0	30.90	58.00	5	29.77	38.50	34.13
4	4M7	26.7	39.0	32.85	58.00	5	29.77	38.50	34.13
5	3E6	33.5	52.4	42.95	59.00	5	37.39	48.82	43.11
6	22MM6	72.5	77.4	74.95	58.33	4	69.40	71.86	70.63
7	23MM6	71.3	78.9	75.10	59.00	4	74.50	78.77	76.64
8	24MM6	65.2	69.9	67.55	58.00	4	66.88	68.46	67.67
9	25MM6	55.5	55.7	55.60	57.00	4	59.24	58.14	58.69
10	33MM6	75.5	83.4	79.45	58.50	4	70.69	73.62	72.15
11	34MM6	76.3	81.7	79.00	59.33	4	77.03	82.18	79.60
12	3E2M5	87.3	88.1	87.70	59.33	4	77.03	82.18	79.60
13	3E3M5	80.8	88.7	84.75	60.00	4	82.14	89.10	85.61
14	223MMM5	109.6	99.9	104.75	58.50	3	107.81	103.57	105.68
15	224MMM5	100.0	100.0	100.00	57.00	3	96.36	88.10	92.22
16	233MMM5	106.1	99.4	102.75	59.00	3	111.62	108.73	110.17
17	234MMM5	102.7	95.9	99.30	58.66	3	109.02	105.22	107.12
18	2233MMMM4	137.0	117.0	127.00	57.75	2	139.19	125.78	132.48

In order to compute the values of RON calc., MON calc. and PON calc., we used the equations (3.8), (3.9), (3.10), respectively.

We then correlated the values of RON observed versus RON calculated with a correlation coefficient r = 0.994, MON observed versus MON calculated with a correlation coefficient r = 0.948, and PON observed versus PON calculated with a correlation coefficient r = 0.989. These results are validating our three QSPR models (3.8), (3.9), (3.10).

Of all these three QSPR models, the best bi-variable linear model involving the topological index SD and the parameter D_C is the one corresponding to Research Octane Number (RON).

Our results in the present paper are very good and are certainly better than those obtained previously by other authors, see the next section for details.

4. OTHER RELATED DEVELOPMENTS ON OCTANE NUMBER

There are many developments regarding the study of octane numbers, most of them focusing on two main directions: 1) the first one that uses topological indices; 2) by means of spectroscopy and multivariate calibration. We summarise some of the most relevant contributions amongst these developments.

In [3], Balaban and Moţoc obtained correlation between octane numbers of 34 alkanes with 4-8 carbon atoms and 14 topological indices. The best correlation coefficient of the obtained models for C_8H_{18} were ranging from 0.593 to 0.945, in the case of Research Octane Number (RON) and from 0.587 to 0.959, in the case of Motor Octane Number (MON).

In another publication [4], for the same set of 34 alkanes, Balaban obtained better mono parametric correlation between MON and the mean square distance ($r^2 = 0.90$ and $r^2 = 0.94$, respectively).

Later, Balaban, Kier and Joshi [5] have studied the correlation between chemical structure of hydrocarbons and their posted octane number (PON), which is defined as the

155

arithmetic mean of Research Octane Number (RON) and Motor Octane Number (MON):

$$PON = \frac{MON + RON}{2}.$$

The tests have been performed separately for 45 alkanes, 35 cycloalkanes and 75 alkanes, respectively. For the first group the statistical parameters reported are better ($r^2 = 0.92$) than the second ($r^2 = 0.78$) and third group ($r^2 = 0.80$), due to the variation range of PON values.

More recently, Gutman et al. [22] considered Research Octane Number only and studied the correlation between the octane numbers related to heptanes and $\ln \pi$ (r = -0.949) and the correlation between the octane numbers related to octanes and $\ln \pi$ (with r = -0.975). Both correlations are significant at the 99% confidence level and octane numbers decrease with the increasing values of $\ln \pi$. Note that $\ln \pi$ is the multiplicative version of the Wiener index (i.e., it equals the product of distances between all pairs of vertices).

Singh et al. [32] studied the Motor Octane Number (MON) for 30 alkanes $C_4 - C_7$ by means of the PJ index. Their predictive model developed to estimate MON of the 30 alkanes $C_4 - C_7$ has low correlation coefficient (r = 0.2794), while by splitting them into 5 families they obtain better results: $r \in [0.2794, 0.98]$.

Hosoya [23] performed correlation studies of the octane number (ON) of heptane and octane isomers with various topological indices: Z, w, p, and B. For single parameter correlation, the index B by Balaban and w by Wiener have good results, while for two-parameter correlation almost any combination among the above four indices give relatively high r value. These models can be improved by adding a third parameter to get r = 0.982. From these studies one can conclude that in order for a gasoline isomer to have high ON or to be burnt without knocking, it should be highly branched (small Z or low boiling point).

Perdih and Perdih [29] have studied empirically the dependence of ON on the structure of alkanes and have shown that ON decreases with the number of CH_2 groups and increases with the number of CH_3 groups. They thus give a more thorough understanding of the relationship between the structure of alkanes and their physicochemical properties and the empirical rules of octane number (ON) dependence on the structure of alkanes are discussed.

Ryzhov et al. [30] used optimal topological indices for studying the octane numbers corresponding to 68 alkanes that have not been characterised experimentally, thus improving the results obtained by Smolenskii et al. [33] that introduced a new QSPR method based on transition numerical values to topological equivalents of physicochemical properties of alkanes.

A more comprehensive QSPR study has been performed by Al-Fahemi et al. [1] to predict octane number of hydrocarbons by means of various parameters calculated from molecular structure: molecular mass (M), hydration energy (E_n), boiling point (BP), octanol/water distribution coefficient ($\log P$), molar refraction (MR), cirtical pressure (C_p), critical volume (C_v), and critical temperature (C_T). Their best linear model for validation of octane number has $r^2 = 0.942$.

Nikolaou et al. [26] presented a calculation method that effectively utilizes the compositional data from high-resolution capillary GC analysis and the measured pure and blending RON values of various hydrocarbons, which are widely published.

For other interesting research work on related topics see also [20]-[25] and [28].

5. CONCLUSIONS

A very good QSPR bi-variable linear model involving the topological index SD and the parameter D_C has been obtained for Octane Number of isooctanes.

REFERENCES

- Al-Fahemi, J. H., Albis, N. A. and Gad, E. A. M., QSPR models for octane number prediction, J. Theoretical Chem., 2014, Article ID 520652, 6 pp.
- [2] Al-Ghouti, M. A., Al-Degs, Y. S. and Amer, M., Determination of motor gasoline adulteration using FTIR spectroscopy and multivariate calibration, Talanta, 76 (2008), No. 5, 1105–1112
- [3] Balaban, A. T. and Motoc, I., Chemical Graphs XXXVI. Correlations between Octane Numbers and Topological Indices in Alkanes, Commun. Math. Chem. (MATCH), 5 (1979), 197–218
- [4] Balaban, A. T., Pure Appl. Chem., 55 (1983), 199-206
- [5] Balaban, A. T., Kier, L. B. and Joshi, N., Structure-property analysis of octane numbers for hydrocarbons (alkanes, cycloalkanes, alkenes), Commun. Math. Chem. (MATCH), 28 (1992), 13–27
- [6] Barysz, M., Jashari, G., Lall, R. S., Srivastava, V. K. and Trinajstić, N., Chemical Applications of Topology and Graph Theory (King R. B., ed.), Elsevier, Amsterdam, 1983, 222–230
- [7] Berinde, Z., Applications of Molecular Topology in the Study of Physico-chemical Properties of Organic Compounds (in Romanian), Cub Press 22, Baia Mare, 2001
- [8] Berinde, Z., Considerații privind modelarea matricială a compuşilor halogenați, Rev. Chim. (Bucureşti), 52 (2001), No. 12, 788–792
- [9] Berinde, Z., Un nou indice topologic de tip Randić, Rev. Chim. (Bucureşti), 53 (2002), No. 1, 812-816
- [10] Berinde, Z. and Drinkal, C., On a local invariant for modeling molecular double bonds graphs, Chem. Bull. "Politehnica" Univ. (Timişoara), 49 (63) (2004), No. 1-2, 8–10
- [11] Berinde Z. and Drinkal, C., Molecular modelling of molar refraction for alkanes, Chem. Bull. "Politehnica" Univ. (Timişoara), 49 (63) (2004), No. 1-2, 4–7
- [12] Berinde, Z., Vertex- and edge-weighted molecular graphs for amines, Rev. Roum. Chim., 51 (2006), No. 11, 1131– 1135
- [13] Berinde, Z., QSPR modelling of molar volume of alkanes using the ZEP topological index, Creat. Math. Inform., 17 (2008), No. 3, 308–312
- [14] Berinde, Z., Matrix mathematical models used in the representation of molecular structures, Sc. Stud. Res. Ser. Math. Inf., 19 (2009), No. 2, 59–70
- [15] Berinde, Z., Modelling normal boiling points of alkanes by linear regression using the SD index, Creat. Math. Inform., 19 (2010), No. 2, 135–139
- [16] Berinde, Z., A QSPR study of hydrophobicity of phenols and 2-(aryloxy-α-acetyl)-phenoxathiin derivatives using the topological index ZEP, Creat. Math. Inform., 22 (2013), No. 1, 33–40
- [17] Berinde, Z., QSTR mathematical models for the toxicity of aliphatic carboxylic acids on tetrahymena pyriformis, Creat. Math. Inform., 22 (2013), No. 2, 151–160
- [18] Berinde, Z., Comparing the molecular graph degeneracy of Wiener, Harary, Balaban, Randić and ZEP topological indices, Creat. Math. Inform., 23 (2014), No. 2, 165–174
- [19] Billingsley, D., Gordon, L. and Tzanzalian, Z., Octane prediction of gasoline blends using neural nets, National Petroleum Refiners Association-Publications-All series, 1995
- [20] Chung, H., Lee, H. and Jun, C.-H., Determination of research octane number using NIR spectral data and ridge regression, Bull. Korean Chem. Soc., 22 (2001), No. 1, 37–42
- [21] Ghosh, P., Hickey K. J. and Jaffe, S. B., Development of a detailed gasoline composition-based octane model, Ind. Eng. Chem. Res., 45 (2006), No. 1, 337–345
- [22] Gutman, I., Linert, W., Lukovits, I. and Tomović, Z., The multiplicative version of the Wiener index, J. Chem. Inf. Comput. Sci., 40 (2000), 113–116
- [23] Hosoya, H., Chemical meaning of octane number analyzed by topological indices, Croat. Chem. Acta, 75 (2002), No. 2, 433–445
- [24] Kelly, J. J., Barlow, C. H., Jinguji, T. M. and Callis, J. B., Prediction of gasoline octane numbers by near infrared spectroscopy in the spectral range 680–1235 nm, Analytical Chem., 61 (1989), No. 4, 313–320
- [25] Meusinger, R. and Moros, R., Determination of octane numbers of gasoline compounds from their chemical structure by 13C NMR spectroscopy and neural networks, Fuel, 80 (2001), No. 5, 613–621
- [26] Nikolaou, N., Papadopoulos, C. E., Gaglias, I. A. and Pitarakis, K. G., A new non-linear calculation method of isomerization gasoline research octane number based on gas chromatographic data, Fuel, 83 (2004), No. 4-5, 517–523
- [27] Nikolić, S., Plavšić, D. and Trinajstić, N., On the Balaban-like topological indices, MATCH Commun. Math. Comput. Chem., 44 (2001), 361–386

- [28] Pasadakis, N., Gaganis, V. and Foteinopoulos, C., Octane number prediction for gasoline blends, Fuel Processing Technology, 87 (2006), No. 6, 505–509
- [29] Perdih, A. and Perdih, F., Chemical interpretation of octane number, Acta Chem. Slovenica, 53 (2006), No. 3, 306–315
- [30] Ryzhov, A. N., Strizhakova, Yu. A., Smolenskii, E. A. and Lapidus, A. L., Modeling the octane numbers of alkenes by the inverse function method, Petroleum Chem., 51 (2011), No. 5, 354–362
- [31] Scott, E. J. Y., Knock characteristics of hydrocarbon mixtures, SAE Journal, 38 (1958), 90
- [32] Singh, S., Joshi, S., Shrivastova A. and Khadikar, P., A novel method of estimating Motor Octane Number (MON)-A Structure Property-Relationship Approach, J. Sci. & Ind. Res., 61 (2002), 961–965
- [33] Smolenskii, E. A., Ryzhov, A. N., Bavykin, V. M., Myshenkova, T. N. and Lapidus, A. L., Octane numbers (ONs) of hydrocarbons: a QSPR study using optimal topological indices for the topological equivalents of the ONs, Russian Chem. Bull., 56 (2007), No. 9, 1681–1693

DEPARTMENT OF CHEMISTRY AND BIOLOGY NORTH UNIVERSITY CENTER AT BAIA MARE TECHNICAL UNIVERSITY OF CLUJ-NAPOCA VICTORIEI 76, 430122 BAIA MARE, ROMANIA Email address: zoitaberinde@ubm.ro Email address: dee1168@yahoo.com Email address: dthomas2005@gmail.com