**MATCH**
*Communications in Mathematical and in Computer Chemistry*

MATCH Commun. Math. Comput. Chem. **75** (2016) 583-592

ISSN 0340 - 6253

# Quantitative Structure Property Relationship Models for the Prediction of Gas Heat Capacity of Benzene Derivatives Using Topological Indices

**H. Hosseini, F. Shafiei**[1]

*Department of Chemistry, Science Faculty, Arak Branch, Islamic Azad University, Arak, Iran*

(Received October 10, 2015)

## Abstract

Graph-theoretical approach represents simple and efficient means to quantitative structure-property relationship (QSPR) studies. Graph-theoretical topological indices are high potential descriptors for modeling and predicting physicochemical properties of chemical compounds. A QSPR study was performed for prediction of gas heat capacity ($C_v$) of 69 benzene derivatives using Wiener (W), Szeged (Sz), first order molecular connectivity ($^1\chi$), Balaban (J), hyper-Wiener(WW), Wiener polarity ($W_P$) and Harary (H) topological indices. The calculation was performed by the ab-initio method at HF/6-31G(d) level of theory. The relationship analysis between heat capacity ($C_v$) and topological indices was done by using multiple linear regression (MLR) method, with heat capacity ($C_v$) as dependent variable and seven independent variables to generate the equation that relates the structural features to the heat capacity ($C_v$) properties. The results show good models with three-seven parameters linear equations. The best model in this study is contains three descriptors ($^1\chi$, W, Sz) are included, with values of the correlation coefficient (r=0.943), the standard error (s=9.667Jmol$^{-1}$K$^{-1}$), the Fisher-ratio (F=172.475), the adjusted coefficient of determination ($r^2_{adj} = 0.883$) and Durbin-Watson value (D=1.728), which indicate that these descriptors, play an important role in effect on heat capacity ($C_v$) of benzene derivatives.

## 1 Introduction

A topological representation of a molecule can be carried out through molecular graph. The descriptors are numerical values associated with chemical constitution for correlation of chemical structure with various physical properties, chemical reactivity or biological activity. A topological index is the graph invariant number calculated from a graph representing a

[1] Corresponding author, e-mail: f-shafiei@iau-arak.ac.ir

molecule. To translate chemical structures into a single number, the graph theory visualizes chemical structures as mathematical object sets consisting of vertices or points, which symbolize atoms, and vertices or lines, linking a pair of edges, which represent covalent bonds or shared electron pairs. In this notation, adjacent vertices stand for pairs of covalently linked atoms situated at a topological distance equal to unit.

In this study, the relationship analysis between heat capacity ($C_v$) and topological indices was done by MLR analysis.

Heat capacity or thermal capacity is a measurable physical quantity and an extensive property of matter and it is defined as the derivative of the energy of the system with respect to the temperature under specified conditions. If the system is maintained at constant volume the heat capacity is shown with $C_v$ and if the system is maintained at constant pressure, the heatcapacity is shown with $C_p$. The values of heat capacity reported in this paper are those at constant volume and correspond to one mole of a specified substance; the units are thus, J $mol^{-1}K^{-1}$.This property is involved in processes such as distillation, evaporation, extraction and heating, found in the petrochemical, pharmaceutical and food industry, among others [1].

There is abundant information in the literature on the heat capacity of organic and inorganic compounds; theoretical approaches, experimental data, generalized correlations, empirical equations, and group contribution methods.

QSPR models based on molecular descriptors for the prediction of liquid heat capacity at $25^o$ C using a set of organic compounds have been developed [2,3].

Heat capacities and entropies of organic compounds in the condensed phase were examined and evaluated [4].

Estimation of liquid heat capacity at constant pressure and $25^o$ C by using additive rules has been proposed [5, 6].

Prediction of the heat capacity of ionic liquids by using the mass connectivity index and a group contribution method has been developed [7].

A new topological index for the studies on structure-properties of alkanes has been examined [8].

A model for evaluation the heat capacity of alkanes by using artificial neural network (ANN) has been developed [3,9].

A linear model for predicting the heat capacity of alcohols and aldehydes in liquid phase at 298 K by using topologic, electronic, and geometric descriptors has been examined [10].

Group contribution methods for estimating the heat capacity of fluids at room temperature (20 or $25^o$ C) have been proposed [11,12].

Relationship between topological indices and thermodynamic properties such as heat capacity, thermal energy and entropy of the monocarboxylic acids has been searched [13].

The main aim of this study is to illustrate the usefulness of topological indices in QSPR study of gas heat capacity ($C_V$) of benzene derivatives. As far as we are aware, this is the first QSPR study for prediction of benzene derivatives heat capacities using topological indices.

## 2 Materials, Mathematical Method and Graphs

The heat capacity ($C_V$) of 68 benzene derivatives (68 compound and benzene) is taken from the quantum mechanics methodology with Hartree-Fock (HF) level using the ab initio 6-31G basis sets. The benzene derivatives in this set have seven different substituents; each substituent being present in at least six compounds. These substituents are amino, bromo, chloro, hydroxyl, methyl, methoxyl and nitro groups. Studied benzene derivatives and their heat capacity are listed in Table 1.

**Table 1.** Benzene derivatives and their heat capacity, used in present study.

| Compounds | Comp. No. | $C_V$(J/molK) | Compounds | Comp. No | $C_V$(J/molK) |
|---|---|---|---|---|---|
| Bromobenzene | 1 | 79.299 | 4-Methylphenol | 36 | 106.226 |
| phenol | 2 | 81.815 | 4-Methyl-3,5-dinitroaniline | 37 | 180.325 |
| 1,2-Dichlorobenzene | 3 | 93.887 | 1,3,5-Trichlorobenzene | 38 | 110.665 |
| 3-Chlorotoluene | 4 | 102.677 | Benzene | 39 | 62.115 |
| 1,3-Dihydroxybenzene | 5 | 101.337 | 2-Nitrotoluene | 40 | 119.539 |
| 3-Hydroxyanisol | 6 | 119.205 | 1,4-Dinitrobenzene | 41 | 131.636 |
| 4-Methyl-3-nitroaniline | 7 | 145.961 | 2-Methyl-3,6-dinitroaniline | 42 | 174.925 |
| 2,4-Dimethylphenol | 8 | 129.881 | 2-Methyl-4,6-dinitrophenol | 43 | 170.594 |
| 2,6-Dimethylphenol | 9 | 129.471 | 2,5-Dinitrotoluene | 44 | 154.731 |
| 3-Nitrotoluene | 10 | 121.107 | 1,2-Dinitrobenzene | 45 | 115.648 |
| 2,6-Dinitrotoluene | 11 | 154.254 | 1,4-Dimethoxybenzene | 46 | 138.876 |
| 4-Methyl-2,6-dinitroaniline | 12 | 165.152 | 2-Methyl-3-nitroaniline | 47 | 145.548 |
| 5-Methyl-2,6-dinitroaniline | 13 | 172.107 | 2-Methyl-4-nitroaniline | 48 | 144.03 |
| 5-Methyl-2,4-dinitroaniline | 14 | 174.047 | 4-Hydroxy-3-nitroaniline | 49 | 140.009 |
| 2,4-Dinitrotoluene | 15 | 154.568 | 4-Chloro-3-methylphenol | 50 | 121.554 |
| 4-Nitrophenol | 16 | 115.757 | 2,4,6-Tribromophenol | 51 | 131.172 |
| 4-Chlorotoluene | 17 | 102.598 | 2,4,6-Trinitrotoluene | 52 | 124.969 |
| 2,4,6-Trichlorophenol | 18 | 129.082 | 1,2,4,5-Tetrachlorobenzene | 53 | 125.68 |
| Toluene | 19 | 86.367 | 3-Methyl-2,4-dinitroaniline | 54 | 174.348 |
| 3-Methyl-6-nitroaniline | 20 | 141.589 | 2-Methyl-3,5-dinitroaniline | 55 | 106.226 |
| 4-Methyl-2-nitroaniline | 21 | 141.911 | 3,5-Dinitrotoluene | 56 | 180.325 |
| 1,2,4-Trichlorobenzene | 22 | 110.026 | 3,4-Dinitrotoluene | 57 | 110.665 |
| 3,4-Dichlorotoluene | 23 | 118.306 | 1,2,4-Trimethylbenzene | 58 | 62.115 |
| 2,4-Dichlorotoluene | 24 | 117.993 | 2,4-Dinitrophenol | 59 | 119.539 |
| Chlorobenzene | 25 | 78.275 | 3,4-Dimethylphenol | 60 | 131.636 |
| 1,3,5-Trinitrobenzene | 26 | 166.94 | 2,4-Dichlorophenol | 61 | 174.925 |
| 1,2,3,4-Tetrachlorobenzene | 27 | 125.362 | 1,2,3-Trichlorobenzene | 62 | 109.641 |
| 2,3,4,5,6- | 28 | 159.713 | 2-Methyl-6-nitroaniline | 63 | 140.352 |

| | | | | | |
|---|---|---|---|---|---|
| Pentachlorophenol | | | | | |
| 1,3-Dichlorobenzene | 29 | 94.459 | 2-Methyl-5-nitroaniline | 64 | 146.17 |
| 2-Chlorophenol | 30 | 98.109 | 1,3-Dinitrobenzene | 65 | 131.649 |
| 3-Methylphenol | 31 | 106.084 | 4-Nitrotoluene | 66 | 121.003 |
| 2,3-Dinitrotoluene | 32 | 155.897 | 1,2-Dimethylbenzene | 67 | 109.357 |
| 1,4-Dimethylbenzene | 33 | 102.339 | 2-Methylphenol | 68 | 105.407 |
| 2,3,4,5-Tetrachlorophenol | 34 | 144.853 | 1,4-Dichlorobenzene | 69 | 94.368 |
| 2,3,6-Trinitrotoluene | 35 | 190.487 | | | |

For obtaining appropriate QSPR model we have used multiple linear regression (MLR) techniques procedure of SPSS version 16, and backward stepwise regression was used to construct the QSPR models.

For drawing the graphs of our results, we used the Microsoft Office Excel – 2003program.

## 3 Topological Indices

As known, each molecule may be represented by a topological graph G={V,E} where V(G) and E(G) are the vertex and edge sets, respectively. Vertices correspond to individual atoms in the graph and the edges correspond to chemical bonds between them.

A large number of topological indices have been defined and used. The majority of the topological indices are derived from the various matrices corresponding to the molecular graphs. The adjacency matrix (A) and the distance matrix ($D_m$) of the molecular graph have been most widely used in the definition of topological indices. Various definitions of topological indices have been used in order to obtain molecular descriptors. The most used ones are the following:

**3.1 Wiener index,** W [14],

**3.2 Hyper-Wiener index,** WW[15-17],

**3.3 Wiener polarity index,** $W_P$ **[**18,19],

**3.4 Randić index,** $^1\chi$ [20,21],

**3.5 Balaban index**, J [22,23],

**3.6 Harary number**, H [24],

**3.7 Szeged index**, Sz [25,26].

All the used topological indices were calculated using hydrogen suppressed graph by deleting all the carbon hydrogen as well as heteroatomic hydrogen bonds from the structure of the benzene derivatives. The calculations of topological indices used in this paper are well documented.

The descriptors were calculated with chemicalize program [27]. Seven topological indices tested in the present study are recorded in Table 2.

**Table 2.** Benzene derivatives and their topological indices, used in present study.

| Comp. No. | $^1\chi$ | J | H | W | WW | $W_p$ | Sz | Comp. No. | $^1\chi$ | J | H | W | WW | $W_p$ | Sz |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 3.39 | 1.82 | 12.92 | 42 | 71 | 5 | 78 | 36 | 3.79 | 2.19 | 16.03 | 62 | 115 | 7 | 110 |
| 2 | 3.39 | 1.82 | 12.92 | 42 | 71 | 5 | 78 | 37 | 6.43 | 2.70 | 39.02 | 282 | 669 | 21 | 420 |
| 3 | 3.80 | 2.28 | 16.17 | 60 | 106 | 8 | 106 | 38 | 4.18 | 2.08 | 19.50 | 84 | 159 | 9 | 144 |
| 4 | 3.79 | 2.23 | 16.08 | 61 | 110 | 7 | 108 | 39 | 3.00 | 2.00 | 10.00 | 27 | 42 | 3 | 54 |
| 5 | 3.79 | 2.23 | 16.08 | 61 | 110 | 7 | 108 | 40 | 4.72 | 2.40 | 22.90 | 114 | 231 | 12 | 180 |
| 6 | 4.33 | 1.98 | 19.15 | 88 | 176 | 9 | 146 | 41 | 5.61 | 2.30 | 29.74 | 206 | 521 | 15 | 314 |
| 7 | 5.11 | 2.25 | 26.67 | 148 | 315 | 14 | 232 | 42 | 6.45 | 2.64 | 38.87 | 289 | 717 | 22 | 434 |
| 8 | 4.20 | 2.09 | 19.53 | 84 | 160 | 10 | 144 | 43 | 6.43 | 2.66 | 3.85 | 286 | 691 | 21 | 428 |
| 9 | 4.22 | 2.15 | 19.67 | 82 | 151 | 11 | 140 | 44 | 6.02 | 2.28 | 34.14 | 246 | 616 | 18 | 372 |
| 10 | 4.70 | 2.32 | 22.73 | 117 | 245 | 11 | 186 | 45 | 5.63 | 2.54 | 30.43 | 188 | 416 | 16 | 278 |
| 11 | 6.04 | 2.40 | 34.60 | 234 | 545 | 19 | 348 | 46 | 4.86 | 2.17 | 22.24 | 125 | 287 | 11 | 200 |
| 12 | 6.43 | 2.70 | 39.02 | 282 | 669 | 21 | 420 | 47 | 5.13 | 2.28 | 26.80 | 146 | 306 | 15 | 228 |
| 13 | 6.45 | 2.72 | 39.13 | 281 | 667 | 22 | 418 | 48 | 5.11 | 2.18 | 26.50 | 152 | 337 | 14 | 240 |
| 14 | 6.43 | 2.65 | 38.83 | 287 | 698 | 21 | 430 | 49 | 5.11 | 2.25 | 26.67 | 148 | 315 | 14 | 232 |
| 15 | 6.02 | 2.33 | 34.30 | 240 | 576 | 18 | 360 | 50 | 4.20 | 2.09 | 19.53 | 84 | 160 | 10 | 144 |
| 16 | 4.70 | 2.26 | 22.60 | 120 | 262 | 11 | 192 | 51 | 4.61 | 2.49 | 23.28 | 110 | 215 | 13 | 184 |
| 17 | 3.79 | 2.19 | 16.03 | 62 | 115 | 7 | 110 | 52 | 7.34 | 2.80 | 47.72 | 408 | 1044 | 25 | 594 |
| 18 | 4.61 | 2.49 | 23.28 | 110 | 215 | 13 | 184 | 53 | 4.61 | 2.46 | 23.23 | 111 | 220 | 13 | 186 |
| 19 | 3.39 | 1.82 | 12.92 | 42 | 71 | 5 | 78 | 54 | 6.45 | 2.72 | 39.13 | 281 | 667 | 22 | 418 |
| 20 | 5.11 | 2.22 | 26.60 | 150 | 327 | 14 | 236 | 55 | 6.43 | 2.66 | 38.85 | 286 | 691 | 21 | 428 |
| 21 | 5.11 | 2.25 | 26.67 | 148 | 315 | 14 | 232 | 56 | 6.00 | 2.33 | 34.23 | 240 | 573 | 17 | 360 |
| 22 | 4.20 | 2.09 | 19.53 | 84 | 160 | 10 | 144 | 57 | 6.02 | 2.40 | 34.53 | 234 | 542 | 18 | 348 |
| 23 | 4.20 | 2.09 | 19.53 | 84 | 160 | 10 | 144 | 58 | 4.20 | 2.09 | 19.53 | 84 | 160 | 10 | 144 |
| 24 | 4.20 | 2.09 | 19.53 | 84 | 160 | 10 | 144 | 59 | 6.02 | 2.33 | 34.3 | 240 | 576 | 18 | 360 |
| 25 | 3.39 | 1.82 | 12.92 | 42 | 71 | 5 | 78 | 60 | 4.20 | 2.09 | 19.53 | 84 | 160 | 10 | 144 |
| 26 | 6.91 | 2.46 | 42.60 | 354 | 906 | 21 | 516 | 61 | 4.20 | 2.09 | 19.53 | 84 | 160 | 10 | 144 |
| 27 | 4.63 | 2.52 | 23.37 | 109 | 211 | 14 | 182 | 62 | 4.22 | 2.15 | 19.67 | 82 | 151 | 11 | 140 |
| 28 | 5.46 | 2.76 | 31.60 | 174 | 357 | 21 | 282 | 63 | 5.13 | 2.28 | 26.8 | 146 | 306 | 15 | 228 |
| 29 | 3.79 | 2.23 | 16.08 | 61 | 110 | 7 | 108 | 64 | 5.11 | 2.18 | 26.5 | 152 | 337 | 14 | 240 |
| 30 | 3.80 | 2.28 | 16.17 | 60 | 106 | 8 | 106 | 65 | 5.61 | 2.40 | 30.02 | 197 | 464 | 15 | 296 |
| 31 | 3.79 | 2.23 | 16.08 | 61 | 110 | 7 | 108 | 66 | 4.70 | 2.26 | 22.6 | 120 | 262 | 11 | 192 |
| 32 | 6.04 | 2.47 | 34.83 | 228 | 511 | 19 | 336 | 67 | 3.80 | 2.28 | 16.17 | 60 | 106 | 8 | 106 |
| 33 | 3.79 | 2.19 | 16.03 | 62 | 115 | 7 | 110 | 68 | 3.80 | 2.28 | 16.17 | 60 | 106 | 8 | 106 |
| 34 | 5.04 | 2.39 | 27.32 | 140 | 281 | 17 | 230 | 69 | 3.79 | 2.19 | 16.03 | 62 | 115 | 7 | 110 |
| 35 | 7.36 | 2.83 | 47.97 | 405 | 1036 | 26 | 588 | | | | | | | | |

# 4 Regression Analysis

In the present work, linear regression analyses were performed using SPSS/PC software package (version 16.0).

The gas heat capacity ($C_v$ J/molK) is used as the dependent variable and $^1\chi$, J, H, W, $W_P$, WW and Sz indices as the independent variables.

Criteria for selection of the best multiple linear regression model were the statistics: squared multiple correlation coefficient ($r^2$), adjusted correlation coefficient ($r^2_{adj}$), Fisher-ratio (F), standard error of estimate (s), Durbin-Watson value (D) and significance (Sig).

## 5 Result and Discussion

Several linear QSPR models involving three-seven descriptors are established and strongest multivariable correlations are identified by the Back ward step wise regression routine implemented in SPSS is used to develop the linear model for the prediction of gas heat capacity using calculated topological descriptors. In the first of this study we drown scattering plots of heat capacity versus the seven topological indices. Some of these plots are given in Figures (1-3), respectively.

Distribution of the dependent variable against the independent variable for 69 chemicals employed in developing quantitative structure-property relationship.

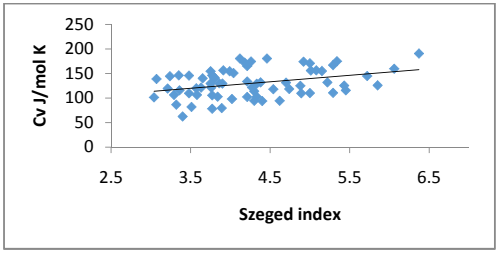


**Figure 1.** Plot of the Szeged index (Sz) versus heat capacityof 69 benzene derivatives.

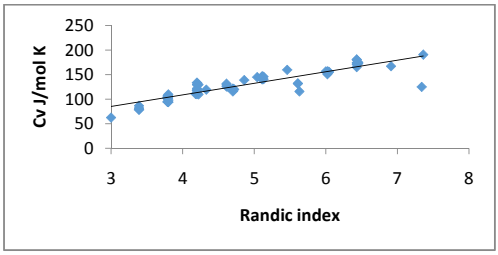


**Figure 2.** Plot of the Randić index ($^1\chi$) versus heat capacity of 69 benzene derivatives.

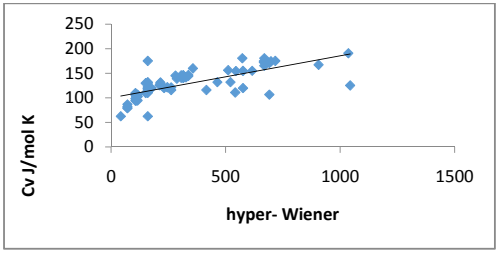


**Figure 3.** Plot of the hyper- Wiener (WW) versus heat capacity of 69 benzene derivatives.

### 5.1. QSPR models for heat capacity

In Table 3 are given the regression parameters and quality of correlation of the proposed models for the gas heat capacity of 69 benzene derivative compounds.

**Table 3.** Statistical parameters of models calculated with SPSS software

| Model | independent variables | r | $r^2$ | $r^2_{adj}$ | s | F | Sig |
|---|---|---|---|---|---|---|---|
| 1 | Sz, J, H, $W_p$, $^1\chi$, HW, W | 0.946 | 0.895 | 0.883 | 9.688 | 74.121 | 0.000 |
| 2 | Sz, J, H, $^1\chi$, WW, W | 0.946 | 0.895 | 0.885 | 9.610 | 87.892 | 0.000 |
| 3 | Sz, J, $^1\chi$, WW,W | 0.946 | 0.895 | 0.886 | 9.541 | 106.966 | 0.000 |
| 4 | Sz, $^1\chi$, WW , W | 0.944 | 0.891 | 0.884 | 9.638 | 130.484 | 0.000 |
| 5 | Sz, $^1\chi$, W | 0.943 | 0.888 | 0.883 | 9.667 | 172.475 | 0.000 |

The best linear model contains three topological descriptors, namely, Randić ($^1\chi$), Wiener (W) and Szeged (Sz) indices.

The regression parameters of the best three descriptors correlation models are gathered in equation (1).

Model $5C_v$=-84.569+43.970 $^1\chi$-2.298W+1.463Sz(9)

n=69 r=0.943$r^2$=0.888 $r^2_{adj}$ = 0.883 s=9.667Jmol$^{-1}$ K$^{-1}$F=172.475 Sig=0.000 D=1.728 (1)

This model produced a standard error of 9.667 Jmol$^{-1}$ K$^{-1}$, a correlation coefficient of 0.943, and the adjusted correlation coefficient (adjusted r-squared) was calculated as 0.883.

The result is therefore very satisfactory. Figure 4 shows the linear correlation between the observed and the predicted heat capacity values obtained using equation (1).

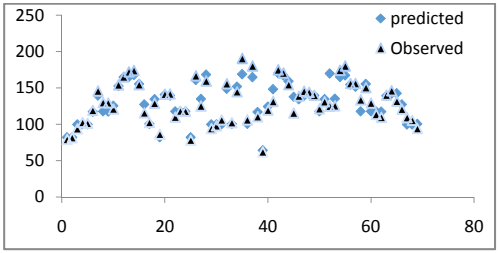


**Figure 4.** Comparison between the predicted and observed of heat capacity by MLR method.

### 5.2 The Durbin-Watson statistics

In this section for verification and validity of the regression models, we will focus on the Durbin-Watson (D) statistics, non-standardized predicted and residual values.

The Durbin-Watson statistics ranges in value from 0 to 4. A value near 2 indicates non-autocorrelation; a value toward 0 indicates positive autocorrelation; a value toward 4 indicates negative autocorrelation. Therefore, the value of Durbin-Watson statistics is close to 2 if the errors are uncorrelated. In our model, the value of Durbin-Watson statistics for model 5 is close to 2 (see Eq. (1)) hence the errors are uncorrelated.

### 5.3 The residuals values

The residuals values of heat capacity expressed by equation (1) show in Table 4. The residual values show a fairly random pattern (see Figure 5). This random pattern indicates that a linear model provides a decent fit to the data.

**Table 4.** Comparison between predicted and observed values of heat capacity ($C_v$) of respect benzene derivatives.

| Comp. No. | Observed $C_v$(J/molK) | Predicted $C_v$(J/molK) | Residual | Comp. No. | Observed $C_v$(J/molK) | Predicted $C_v$(J/molK) | Residual |
|---|---|---|---|---|---|---|---|
| 1 | 79.299 | 82.107 | -2.808 | 36 | 106.226 | 100.560 | 5.666 |
| 2 | 81.815 | 82.107 | -0.292 | 37 | 180.325 | 164.704 | 15.621 |
| 3 | 93.887 | 99.742 | -5.855 | 38 | 110.665 | 116.903 | -6.238 |
| 4 | 102.677 | 99.931 | 2.746 | 39 | 62.115 | 64.310 | -2.195 |
| 5 | 101.337 | 99.931 | 1.406 | 40 | 119.539 | 124.388 | -4.849 |
| 6 | 119.205 | 117.234 | 1.971 | 41 | 131.636 | 148.186 | -16.550 |
| 7 | 145.961 | 139.495 | 6.466 | 42 | 174.925 | 169.982 | 4.943 |
| 8 | 129.881 | 117.783 | 12.098 | 43 | 170.594 | 167.218 | 3.376 |
| 9 | 129.471 | 117.405 | 12.066 | 44 | 154.731 | 159.165 | -4.434 |
| 10 | 121.107 | 125.394 | -4.287 | 45 | 115.648 | 137.754 | -22.106 |
| 11 | 154.254 | 152.503 | 1.751 | 46 | 138.876 | 134.530 | 4.346 |
| 12 | 165.152 | 164.704 | 0.448 | 47 | 145.548 | 139.117 | 6.431 |
| 13 | 172.107 | 164.955 | 7.152 | 48 | 144.03 | 142.008 | 2.022 |
| 14 | 174.047 | 167.846 | 6.201 | 49 | 140.009 | 139.495 | 0.514 |
| 15 | 154.568 | 155.394 | -0.826 | 50 | 121.554 | 117.783 | 3.771 |
| 16 | 115.757 | 127.279 | -11.522 | 51 | 131.172 | 134.594 | -3.422 |
| 17 | 102.598 | 100.560 | 2.038 | 52 | 124.969 | 169.784 | -44.815 |
| 18 | 129.082 | 134.594 | -5.512 | 53 | 125.68 | 135.222 | -9.542 |
| 19 | 86.367 | 82.107 | 4.260 | 54 | 174.348 | 164.955 | 9.393 |
| 20 | 141.589 | 140.752 | 0.837 | 55 | 106.226 | 100.560 | 5.666 |
| 21 | 141.911 | 139.495 | 2.416 | 56 | 180.325 | 164.704 | 15.621 |
| 22 | 110.026 | 117.783 | -7.757 | 57 | 110.665 | 116.903 | -6.238 |
| 23 | 118.306 | 117.783 | 0.523 | 58 | 62.115 | 64.310 | -2.195 |
| 24 | 117.993 | 117.783 | 0.210 | 59 | 119.539 | 124.388 | -4.849 |
| 25 | 78.275 | 82.107 | -3.832 | 60 | 131.636 | 148.186 | -16.550 |
| 26 | 166.94 | 160.832 | 6.108 | 61 | 174.925 | 169.982 | 4.943 |
| 27 | 125.362 | 134.845 | -9.483 | 62 | 109.641 | 117.405 | -7.764 |
| 28 | 159.713 | 168.298 | -8.585 | 63 | 140.352 | 139.117 | 1.235 |
| 29 | 94.459 | 99.931 | -5.472 | 64 | 146.17 | 142.008 | 4.162 |
| 30 | 98.109 | 99.742 | -1.633 | 65 | 131.649 | 142.530 | -10.881 |
| 31 | 106.084 | 99.931 | 6.153 | 66 | 121.003 | 127.279 | -6.276 |
| 32 | 155.897 | 148.732 | 7.165 | 67 | 109.357 | 99.742 | 9.615 |
| 33 | 102.339 | 100.560 | 1.779 | 68 | 105.407 | 99.742 | 5.665 |
| 34 | 144.853 | 151.872 | -7.019 | 69 | 94.368 | 100.560 | -6.192 |
| 35 | 190.487 | 168.778 | 21.709 | | | | |

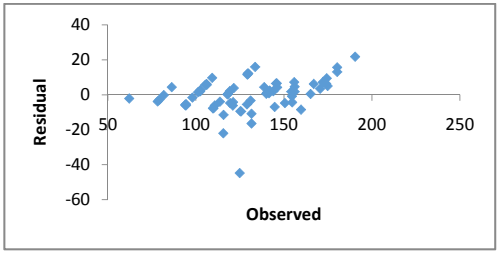


**Figure 5.** Plot of residuals against observed value

Analysis of the obtained results indicates that the MLR model can well represent the structure-property relationships of these compounds, and that we can use only three topological indices for predicting the gas heat capacities of studied compounds.

## 6 Conclusion

QSPR models for the prediction of gas heat capacity for a training set of benzene derivatives using MLR based on topological descriptors calculated from molecular structure alone have been developed. MLR model is proved to be a useful tool in the prediction of gas heat capacity. This model contains fewer parameters having to be optimized: the Wiener (W), Szeged (Sz), and Randić ($^1\chi$) indices.

## References

[1] R. Ceriani, R. Gani, A. J.A. Meirelles, Prediction of heat capacities and heats of vaporization of the organic liquids, *Fluid Phase Equilibria* **283** (2009) 49-55.

[2] X. Yao, B. Fan, J. P. Doucet, A. Panaye, Quantitative structure-property relationship models for the prediction of liquid heat capacity, *QSAR Comb. Sci*. **22 (**2003) 29-48.

[3] A. Ivanova, V. A. Palyulin, A. N. Zefirov, N. S. Zefirov. Fragment descriptors in QSPR: Application to heat capacity calculation, *Russ. J. Org. Chem.* **40** (2004) 644–649.

[4] E. S. Dolmaski, W. H. Evans, E. D. Hearing, Heat capacities and entropies of organic compounds in the condensed phase, *J. Phys. Chem. Ref. Data* **19** (1990) 881–1047.

[5] V. Ružička, E. S. Domalski, Estimation of the heat capacities of organic liquids as a function of temperature using group additivity. I. Hydrocarbon compounds, *J. Phys. Chem. Ref. Data* **22** (1993) 597–619.

[6] M. Luria, S. W. Benson, Heat capacities of liquid hydrocarbons. Estimation of heat capacities at constant pressure as a temperature function, using additivity rules, *J. Chem. Eng. Data* **22** (1977) 90–100.

[7] J. O. Valderrama, A. Toro, R. E. Rojas, Prediction of the heat capacity of ionic liquids using the mass connectivity index and a group contribution method, *J. Chem. Thermodyn.* **43** (2011) 1068–1073.

[8] C. H. Ni, L. C. Li, Z. Y. Feng, A novel topological index for the studies on structure-properties alkanes, *Wuhan Univ. J. Nat. Sci.* **5** (2000) 464–468.
[9] A. A. Gakh, E. G. Gakh, B. G. Sumpter, D.W. Noid, Neural network - graph theory approach to the prediction of the physical properties of organic compounds, *J. Chem. Inf. Comput. Sci.* **34** (1994) 832–839.
[10] F. Ashrafi, R. Saadati, A. Behboodi-Amlashi, Modeling and theoretical calculation of liquid heat capacity of alcohols and aldehydes using QSPR *African J. Pure. Appl. Chem*. **2** (2008) 116–120.
[11] T. A. Albahri, R. S. George, Artificial neural network investigation of the structural group contribution method for predicting pure components auto ignition temperature, *Ind. Eng. Chem. Res.* **42** (2003) 5708–5714.
[12] T. P. Thinh, J. L. Duran, R. S. Ramalho, Estimation of ideal gas heat capacities of hydrocarbons from group contribution techniques. New and accurate approach, *Indian Eng. Chem. Process Des. Dev.***10** (1971) 576–582.
[13] F. Shafiei, Relationship between topological indices and thermodynamic properties of the monocarboxylic acids – Applications in QSPR, *Iran. J. Math. Chem*. **6 (**2015) 15–28.
[14] M. Randić, On generalization of Wiener index for acyclic structures, *Acta. Chim. Slov*. **49** (2002) 483–496.
[15] B. Zhou, I. Gutman, Relations between Wiener, hyper-Wiener and Zagreb indices, *Chem. Phys. Lett*. **394** (2004) 93–95.
[16] D. J. Klein, W. Yan, Y.N. Yeh, Hyper-Wiener vector, Wiener matrix sequence, and Wiener polynomial sequence of a graph, *Int. J. Quantum Chem*. **106** (2006) 1756–1761.
[17] D. J. Klein, I. Lukovits, I. Gutman, On the definition of the hyper-Wiener index for cycle-containing structures, *J. Chem. Inf. Comput. Sci*. **35** (1995) 50–52.
[18] M. Liu, B. Liu, On the Wiener polarity index, *MATCH Commun. Math. Comput. Chem.* **66** (2011) 293–304.
[19] H. Deng, H. Xiao, F. Tang, On the extremal Wiener polarity index of trees with a given diameter, *MATCH Commun. Math. Comput. Chem*. **63** (2010) 257–264.
[20] M. Randić, Generalized molecular descriptors, *J. Math. Chem*. **7** (1991) 155–168.
[21] M. Randić, Charactrization of molecular branching, *J. Am. Chem. Soc.* **97** (1975) 6609–6615.
[22] A. T. Balaban, Highly discriminating distance based topological indices, *Chem. Phys. Lett.* **89** (1982) 399–404.
[23] L. Blaha, J. Damborsky, M. Nemec, QSAR for acute toxicity of saturated and unsaturated halogenated compounds, *Chemosphere* **36** (1998) 1345–1365.
[24] K. C. Das, B. Zhou, N. Trinajstić, Bounds on Harary index, *J. Math. Chem*. **46** (2009) 1369–1376.
[25] I. Gutman, S. Klavžar, An algorithm for the calculation of the Szeged index of benzenoid hydrocarbons, *J. Chem. Inf. Comput. Sci*. **35** (1995) 1011–1014.
[26] P. V. Khadikar, N. V. Deshpande, P. P. Kale, A. Dobrynin, I. Gutman, G. Dömötör, The Szeged index and an analogy with the Wiener index, *J. Chem. Inf. Compt. Sci.* **35** (1995) 547–550.
[27] Web search engine developed by ChemAxon; software available at http:// WWW. Chemicalize. Org.