

## MODELING OF BOILING POINTS OF HYDROCARBON COMPOUNDS

*Nikolay T. Kochev, Ognyan Pukalov, George N. Andreev  
Bulgaria, Plovdiv 4000, Tsar Assen St. 24, University of Plovdiv,  
Department of Analytical Chemistry*

### ABSTRACT

A method for calculation of boiling points of hydrocarbons is described. The method is based on the topological index of Wiener improved with additional topological descriptors. The descriptor values were calculated with in-house developed software JBSMM. Basic boiling point model is a quadratic function of the logarithm of Wiener index. The coefficient values are calculated by means of linear regression applied for a set of hydrocarbon compounds. Test results are presented and discussed.

*Keywords:* QSPR, QSAR, additive scheme, linear regression.

### INTRODUCTION

Applied chemistry is interested in creating chemical compounds with particular properties. Chemoinformatics identifies relationships known in the literature as QSPR/QSAR. These models are generally based on the abstract equation:

$$\text{property} = F(\text{structure, other properties}) \quad (1)$$

A cascading procedure is needed for calculation/prediction of chemical properties:

$$\text{structure} \rightarrow \text{property1} \rightarrow \text{property2} \rightarrow \dots \quad (2)$$

Boiling point (BP) is an important physicochemical property for modeling of other properties with practical value in environmental protection and pharmaceutical industry. However, BP data often is not available, and therefore it must be estimated.

Experimental boiling points are expected to be highly accurate, but it must be noted that the presence of impurities can cause inaccurate experimental measurement.

So far there were reported many QSPR approaches for modeling of BP. The latter shows the importance of this issue. Dearden [1] did a very good review of the most popular methods for BP prediction. Generally a sophisticated set of molecular descriptors is applied and very often high quality commercial software package is needed to make a QSPR model.

In this work we present very simple approach for estimation of normal boiling points of hydrocarbon compounds.

### THEORETICAL MODEL

Our model uses the early results of Wiener where he reported very good correlation between Wiener index and some properties of the organic compounds [2].

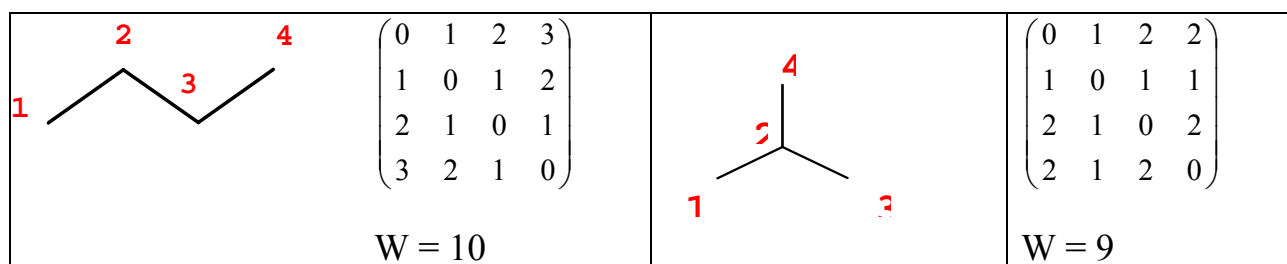
Wiener index [3] **W** is defined as the sum of shortest paths between all possible pairs of atoms in the structural graph. Mathematically it can be obtained from the distance matrix **D** as follows:

$$W = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n d_{ij} \quad (3)$$

where  $d_{ij}$  is the shortest graph path connecting atoms  $i$  and  $j$ .

Generally **W** describes the degree of branching. If a structure is more branched then the paths are shorter and accordingly the value of **W** decreases as well as the boiling point decreases. Also when **W** increases, BP increases accordingly as well as the path lengths in the structure since it is less branched.

To illustrate **W** calculation eq. 3 was applied for normal butane and for isobutane:

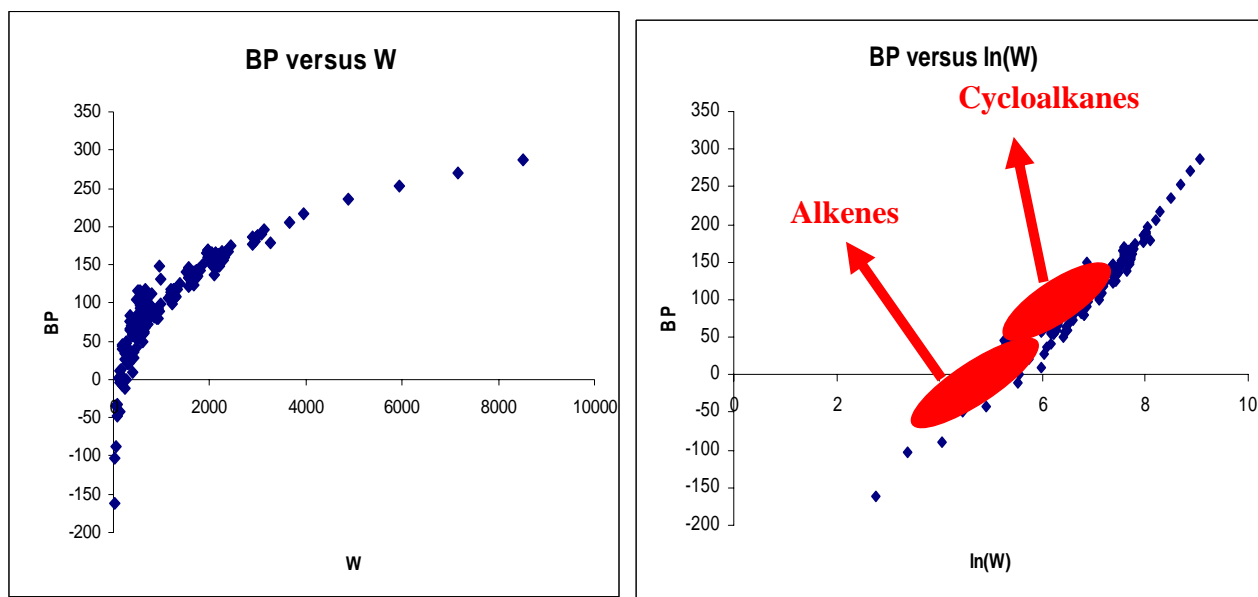


This is a version of the Wiener index where the hydrogen atoms are not taken into account. For the models described below, H atoms were counted in the **W** index calculation.

The original formula of Wiener described **BP** as a linear function of **W** plus a correction, **P** called polarity number [2]:

$$BP = aW + bP + c \quad (4)$$

Trying to establish a relationship of the type  $BP=f(W)$  without taking the correction of Wiener, it was obvious from the plot that  $BP$  depends from  $W$  in a logarithmic manner.



**Figure 1.** Graphical comparison of the values of  $W$  and  $BP$ .

Comparison of the values of  $\ln(W)$  and  $BP$  (see fig.1) showed that the relation is not a straight line. There is a small curve which successfully can be fixed with a quadratic function. Thus we obtained the basic model type for alkanes:

$$BP = a.\ln(W)^2 + b.\ln(W) + c \quad (5)$$

In order to improve model (5), there were needed corrections with simple descriptors Alkenes were modeled by adding two parameters:  $N_{2BR}$  – the relative number of double bonds, and  $N_{AA}$  - the number of Alene atoms. The correction for cyclic structures is  $N_R$  – the number of rings.

$$BP = a.\ln(W)^2 + b.\ln(W) + c + d.N_{2BR} + e.N_{AA} + f.N_R \quad (6)$$

## RESULTS AND DISCUSSION

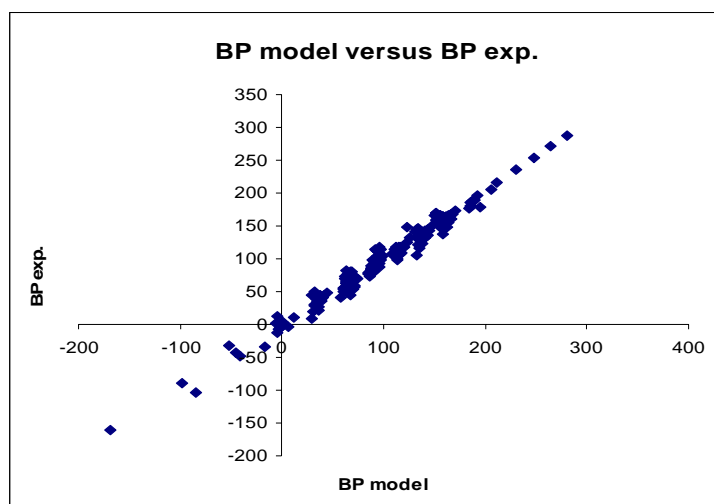
All calculations were performed with **JBSMM** (**J**ava **B**ased **S**ystem for **M**olecular **M**odeling). It is an in-house developed software system. It supports the main stages of the molecular modeling: structure representation  $\rightarrow$  descriptor calculation  $\rightarrow$  model creation and statistics  $\rightarrow$  prediction.

Model of type (5) was generated for a training set of 160 alkanes:  $BP(W) = 4.3 \ln(W)^2 + 19.6 \ln(W) - 247.0$  with a very good statistics: RMS Error = 6.4,  $R=0.993$ . When including alkenes and cyclic hydrocarbons the model of type (5) was not enough accurate prediction of BP. As it can be seen from fig.1, above the basic cluster of objects (they form a quadratic curve) two smaller clusters can be

recognized - alkenes and cyclic hydrocarbons. The cluster deviations were fixed by the correction terms introduced in equation (6).

The final model of the type (6) was generated with a training data set of 272 compounds:

$$BP = 3.35\ln(W)^2 + 31.86\ln(W) - 282.62 + 264.82N_{2BR} + 4.35N_{AA} + 29.87N_R$$



**RMS Error = 8.27**

**R = 0.990**

**Figure 2.** Graphical comparison of the values of modelled and experimental values of BP.

### ACKNOWLEDGEMENTS

We would like to thank the Bulgarian National Fund for Scientific Research NFNI (project VUH-17/05) and Plovdiv University Scientific Fund (project 05-X-64) for supporting this scientific work.

### REFERENCES

1. Dearden J. C. Quantitative Structure–Property Relationships For Prediction Of Boiling Point, Vapor Pressure, And Melting Point, *Environmental Toxicology and Chemistry*, Vol. 22, No. 8, pp. 1696–1709 (2003).
2. H. Wiener, Structural Determination of Paraffin Boiling Points *J. Am. Chem. Soc.* **69** 17–20 (1947).
3. Diudea M.V., Gutman I. Wiener-Type Topological Indices, *Croatica Chemica Acta*, **71**(1)21-51 (1998).