Boiling Point Models of Alkanes

Kimberly Jordan Burch* Department of Mathematical Sciences,

Montclair State University, Upper Montclair, NJ 07043, USA

Diane Karen Wakefield, Department of Chemistry,

Chatham College, Pittsburgh, PA 15232, USA

Earl Glen Whitehead, Jr., Department of Mathematics,

University of Pittsburgh, Pittsburgh, PA 15260, USA

Abstract

We developed single variable models to calculate the boiling points of special families of alkanes and multivariable boiling point models for all alkane isomers up to and including 12 carbon atoms. Our models are based on the molecular weight, the number of carbon atoms, the Hosoya index, the Wiener number, the Wiener Path numbers $^1P, {}^2P, \ldots, {}^8P$, and the methyl and ethyl indices. We introduce the methyl and ethyl indices. We used nonlinear regression techniques using web based submission NEOS solvers to determine the coefficients and exponents of the indices in our models. A standard deviation of 2.7 °C was obtained from our best model which includes 187 of the alkane isomers from 6 to 12 carbon atoms for which experimental boiling point data was available. This model was used to predict the boiling points for a set of test data from 13 to 22 carbon atoms.

1 Introduction

The molecules called alkanes have the chemical formula $C_n H_{2n+2}$ where C represents a carbon atom and H represents a hydrogen atom. The alkanes are also called acyclic saturated hydrocarbons, paraffins, and the methane series. Each carbon atom C has four chemical bonds $(deg(C) = 4)$ and each hydrogen atom H has one chemical bond $(deg(H) = 1)$. Because of these bonding properties of carbon and hydrogen, it is known that each n-carbon alkane is a tree which is an acyclic connected graph containing $3n + 1$ edges (chemical bonds).

Cayley explained his method for enumerating the alkanes in an address to the British Association at their meeting in Bristol in 1875. See [3]. Cayley's interest in the alkanes came from the fact that the alkanes can be represented by graphs known as trees. By ignoring the hydrogen atoms, Cayley obtained a tree consisting entirely of carbon atoms and the carbon-carbon bonds. Cayley succeeded in correctly finding the number of these "carbon trees" with n vertices (carbon atoms) up to $n = 11$. Cayley made errors in the cases $n = 12$ and $n = 13$.

In [21], Rains and Sloane reported that they were unable to find any record of subsequent work where the alkanes were correctly enumerated by Cayley's method. (They cite numerous references of the correct enumeration of the alkanes by other methods.) Rains and Sloane correctly enumerated the alkanes by Cayley's method. They give a table for $1 \le n \le 22$. Table 1 gives the number of alkanes for $1 \le n \le 20$. This table shows the rapid growth of the number of alkanes as the number of carbons increases.

The various alkanes can be characterized by their physical properties including boiling

Table I: Number of alkanes with n carbon atoms

n	alkanes	n	alkanes	n	alkanes	n	alkanes
1	1	6	5	11	159	16	10,359
2	1	7	9	12	355	17	24,894
3	1	8	18	13	802	18	60,523
4	2	9	35	14	1858	19	148,284
5	3	10	75	15	4347	20	366,319

points. Experimental boiling point data is available for most of the alkanes with $n \leq 10$ [2, 19]. It is very unlikely that complete experimental boiling point data will soon be available for all the alkanes with n carbon atoms for any $n \geq 10$. This lack of data is a motivation for constructing boiling point models which can be used to estimate the boiling points of alkanes for which no boiling point data is available. Indices from graph theory and from geometric structure, provide a basis for developing boiling point models for alkanes.

The models presented in this paper are nonlinear models in that the indices are raised to powers different than 1. We decided to use nonlinear models because natural laws are often nonlinear. Light intensity, gravitational force, electric fields, and sound intensity all satisfy inverse square laws. It is interesting to note that Wiener's model (Model 6.1) has a term where n, the number of carbon atoms, appears as an inverse square. In searching for nonlinear models, a linear model would be found if it optimized the objective function.

2 Indices and Error Vector

In this section, we describe the indices and the error vector which are used in the models in the following sections. Our notation for the indices is in agreement with the notation used by Todeschini and Consonni [28].

Molecular weight (MW) is the weight in atomic mass units of all the atoms in a given molecule. Let n denote the number of carbons in the given alkane, the molecular weight is given by the formula $MW(n) = 12.01115\,n + 1.00797(2n + 2)$.

For example, the molecular weight of 2-methylbutane, shown in Figure 1, is 72.1514 amu.
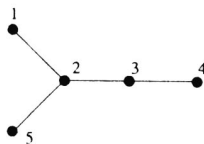

Figure 1: 2-methylbutane

Hosoya index (Z) is the sum of the coefficients of the simple matching polynomial. As defined in Farrell's paper [8], a *matching* is a spanning subgraph consisting of vertices and edges only. A *k-matching* is a matching which has exactly k edges. Letting a_k denote the number of k-matchings, the *simple matching polynomial* of a graph with p vertices is given by the formula $\sum_{k=0}^{[p/2]} a_k \omega^{p-k}$.

For example, the simple matching polynomial for 2-methylbutane is $\omega^5 + 4\omega^4 + 2\omega^3$. The Hosoya index is 7.

Wiener number (W) is the sum of the distances between all pairs of vertices in a graph. It can also be defined as

$$w = \frac{1}{2} \sum_i \sum_j d_{ij}$$

where d_{ij} is the ij^{th} entry in the distance matrix, D(G). See [24].

For example, the distance matrix for 2-methylbutane is $\begin{pmatrix} 0 & 1 & 2 & 3 & 2 \\ 1 & 0 & 1 & 2 & 1 \\ 2 & 1 & 0 & 1 & 2 \\ 3 & 2 & 1 & 0 & 3 \\ 2 & 1 & 2 & 3 & 0 \end{pmatrix}$.

The Wiener number w is 18.

Wiener Path numbers $(^1P, {}^2P, \ldots, {}^9P)$ are the numbers ^iP where ^iP is the number of pairs of vertices in the graph separated by i edges (chemical bonds). Again using the distance matrix, ^iP can be obtained by counting the number of times i appears in the upper triangular part of the matrix. For example, 2-methylbutane has

$$^1P = 4,\ ^2P = 4,\ ^3P = 2,\ ^4P = 0,\ ^5P = 0, \ldots, {}^9P = 0.$$

Methyl (Mth) is defined to be the number of degree one vertices which are adjacent to a vertex of degree three or greater. This is not equivalent to the number of methyls in the IUPAC name.

For example, 2,2-dimethylbutane has $Mth = 3$. The three methyls are edges (1,2), (2,5), and (2,6) as seen in Figure 2.

Since methyl is an electron donor group, the methyl hydrogens have a slight positive charge and the carbon atoms attached to the methyl group will be slightly negative.
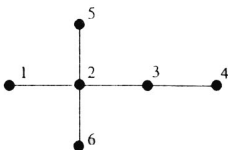
Figure 2: 2,2-dimethylbutane

This polarity will influence the boiling point.

Ethyl (*Eth*) is defined to be the number of degree one vertices that are adjacent to a degree two vertex which is adjacent to a vertex of degree three or higher. This is not equivalent to the number of ethyls in the IUPAC name.

For example, 2,2-dimethylbutane has $Eth = 1$. The ethyl is the P_3 consisting of vertices 2, 3, and 4 shown in Figure 2.

We introduced Ethyl as a companion index to Methyl. Our models indicate that Ethyl is not as useful an index as Methyl.

Error Vector (r^2, number of data points, standard deviation) where r^2 is the coefficient of determination as given in [4],the number of data points is the number of boiling point data in the model, and standard deviation, denoted σ, is the square root of the sum of the squares divided by the number of data points minus the number of indices used.

3 Correlations amongst the Wiener Path Numbers

In [29], Wiener discusses a theory of alternate polarities. He claims that the "total inductive attraction force between atoms of opposite polarities" is given by an expression of the form $k_1\,{}^1P + k_3\,{}^3P + k_5\,{}^5P + \dots$. He then drops ${}^1P$ because it is equal to the number of carbon atoms minus one. He drops ${}^5P$ claiming that k_5 is small when compared to k_3.

In this paper, we use the Wiener path numbers ${}^1P, {}^2P, \dots, {}^9P$. To justify the use of all of these Wiener path numbers, we present the following table of Pearson's sample correlation coefficients between pairs of these Wiener path numbers for the six to twelve

carbon alkanes. Devore and Peck [4] define the Pearson's sample correlation coefficient. This table shows that most of these correlation coefficients are relatively small.

Table II: Wiener Path Number Correlations for C6 to C12

	^2P	^3P	^4P	^5P	^6P	^7P	^8P	^9P
^1P	0.747	0.665	0.764	0.730	0.500	0.330	0.210	0.122
^2P		0.791	0.692	0.349	0.037	-0.110	-0.165	-0.145
^3P			0.764	0.193	-0.166	-0.257	-0.237	-0.165
^4P				0.429	-0.005	-0.148	-0.151	-0.108
^5P					0.597	0.306	0.153	0.065
^6P						0.689	0.421	0.227
^7P							0.692	0.402
^8P								0.653

4 Regression Techniques

The objective functions for our models were written in AMPL [10]. AMPL is a modelling language for mathematical programming. Problems written in AMPL can be solved in two ways:

- You can solve problems locally using packages included with the AMPL software such as MINOS.

- You can solve problems over the Internet using NEOS solvers at Argonne National Laboratory [18].

Due to the size of our data sets, we used the web based submission NEOS solvers. We first tried the solver SNOPT [11], which did not work. SNOPT converged to a point with a much worse objective function than we were obtaining by our initial attempts using Mathematica. We changed to FILTER [9], another NEOS solver. Both of these are sequential quadratic programming (SQP) algorithms [9, 11]. SNOPT uses only the function values and first derivative information from the objective function and constraints

whereas FILTER uses second derivative information from the objective function and constraints in a meaningful way. As a result, FILTER is less prone to converge to shallow local minima.

Nonlinear program algorithms such as FILTER do not stay in the feasible region when optimizing the objective function. This means the resulting powers of the objective function may be slightly negative. This causes problems when data values are zero. Many of the Wiener path numbers, the methyl index, and the ethyl index were zero for the alkanes we modelled. We therefore replaced these zero data values with 10^{-4}. We then restored the zero data values and used the powers FILTER computed to determine the actual models using the Regress command in Mathematica.

5   One Variable Models for Families of Alkanes

All of the models below are used to predict the boiling points of various families of alkanes. Boiling points are a measure of the forces of attraction between like molecules. For essentially nonpolar compounds such as alkanes these forces are London dispersion forces due to instantaneous dipole-induce dipole attractions. Dispersion forces are very short range forces which increase with the number of electrons which is proportional to the MW for the alkanes. The alkane boiling point should depend on the MW and on how well the molecules pack together which is related to the geometry of the molecule. The dependence on the geometry is complex but the boiling point should decrease in a general way as the compactness of the molecule increases if the MW stays the same. Balaban noted that for the same molecular weight, the boiling point decreases with increasing branching [1].

Models 5.1, 5.2, and 5.3 correlate boiling points with MW for alkanes with similar geometric structure. Model 5.4 correlates boiling points with the Hosoya index which indicates some facet of structure as well as the number of carbons. All data and models are in degrees Celsius ($^{\circ}C$).

Model 5.1 *This model uses the $MW^{0.345}$ of the n-alkanes (branchless chains) where $1 \leq n \leq 25$. The formula is $f_{5.1}(MW) = -464.424 + 115.349 MW^{0.345}$. The error vector is $(0.999794, 25, 2.3)$.*

The standard deviation in Model 5.1 is greatly reduced in Model 5.2 by eliminating the first four alkanes. It should be noted that Models 5.1 and 5.2 use similar equations.

Model 5.2 *This model uses the $MW^{0.274}$ of the n-alkanes (branchless chains) where $5 \leq n \leq 25$. The formula is $f_{5.2}(MW) = -642.362 + 210.137 MW^{0.274}$. The error vector is $(0.999863, 21, 1.3)$.*

In order to fit the data beginning with $n = 1$, an equation using the MW in three terms with variable powers was used. Model 5.3 provides a much better fit for the n-alkanes from $n = 1$ up to and including $n = 25$.

Model 5.3 *This model uses the molecular weight, MW, of the n-alkanes where $1 \leq n \leq 25$.*

The formula is $f_{5.3}(MW) = -240.103 - 330.713 MW^{-0.797} + 43.9249 MW^{0.793} - 26.3154 MW^{0.854}$. The error vector is $(0.999925, 25, 1.4)$.
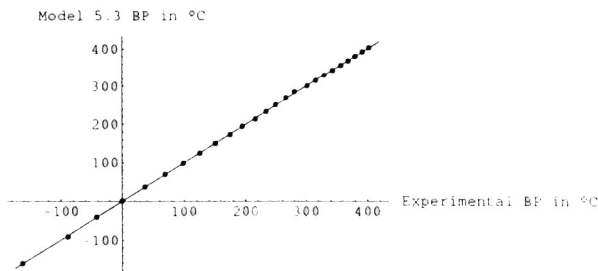

Figure 3: Experimental BP vs. Model BP for Model 5.3

Figure 3 shows a plot of the experimental boiling points versus the model boiling points determined by Model 5.3. The straight line represents an exact prediction.

Model 5.4 *This model uses the Hosoya index, Z, of the normal-alkanes where $1 \leq n \leq 25$. The formula is $f_{5.4}(Z) = -163.231 + 1202.27(\ln Z)^{1.285} - 1800.57(\ln Z)^{1.385} + 707.922(\ln Z)^{1.473}$. The error vector is $(0.999879, 25, 1.8)$.*

In Model 5.4, the carbon trees of the normal-alkanes are paths on n vertices. The Hosoya index on paths is well known to generate the Fibonacci sequence [13, 15]. Plotting the experimental boiling points versus the model boiling points determined by Model 5.4 produces a graph very similar to Figure 3.

If you plot the boiling point for the first few normal-alkanes against MW, it rapidly increases because the London forces increase with MW and these are the forces holding the molecules together in a liquid. This is seen in Figure 4b. But, as the chain gets longer the molecule can fold back on itself and some of the London forces are directed towards itself so the dependence of boiling point on MW diminishes, see Figure 4a. Since the way in which the molecule wraps around itself may be complex and involve significant differences depending on the number of carbons, it is not surprising that the nonlinear function $f_{5.3}(MW)$ fits the data points so well as shown in Figure 3.
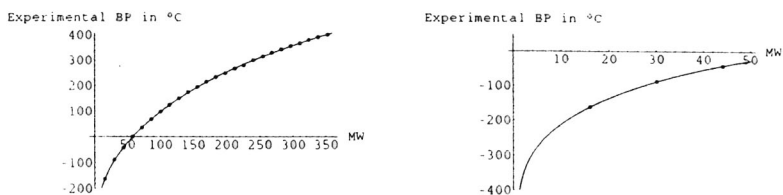

Figure 4: a: MW vs. Experimental BP for $1 \leq n \leq 25$, b: MW vs. Experimental BP for the first three alkanes

6 Variations on Wiener's Model for Six to Ten Carbon Alkanes

Wiener stated that the equation in Model 6.1 was obtained by the method of least squares fitting to the boiling point data of 37 alkanes from C_4H_{10} to C_8H_{18} [29]. In this section we use the boiling data from Dreisbach [6], for the 142 alkanes having n carbon atoms where $6 \leq n \leq 10$. In Models 6.1 and 6.2, the normal alkanes are not fitted because $\Delta t = \Delta W = \Delta P = 0$.

Model 6.1 *This model fits Δt using the change in Wiener index ΔW, the number of carbon atoms n, and the change in the Wiener ^3P number, ΔP [29], [30]. The formula is*

$$f_{6.1}(\Delta W, n, \Delta P) = 98\frac{\Delta W}{n^2} + 5.5\Delta P$$

where $\Delta t = t_{normal} - t_{isomer}$, $\Delta W = W_{normal} - W_{isomer}$, $\Delta P = P_{normal} - P_{isomer}$, and P is an abbreviation for ^3P. The error vector is $(0.76492, 138, 3.1)$.

Model 6.2 was obtained from 6.1 by changing the power on n from 2 to 2.21 and by allowing a nonzero intercept, namely 1.79342. A manual approach using *Mathematica* was used to obtain the exponent. Examining the standard deviations of Model 6.2 and Model 6.1, we see that Model 6.2 with a standard deviation of 2.3 fits the data significantly better than Model 6.1, which has a standard deviation of 3.1.

Model 6.2 *This model fits Δt using the change in Wiener index ΔW, the number of carbon atoms n, and the change in the Wiener ^3P number, ΔP. The formula is*

$$f_{6.2}(\Delta W, n, \Delta P) = 1.79342 + 136.908\frac{\Delta W}{n^{2.21}} + 5.30056\Delta P$$

where $\Delta t = t_{normal} - t_{isomer}$, $\Delta W = W_{normal} - W_{isomer}$, $\Delta P = P_{normal} - P_{isomer}$, and P is an abbreviation for ^3P. The error vector is $(0.86914, 138, 2.3)$.

The transition from Model 6.2 to Model 6.3 involves some algebra. The equation in Model 6.2 can be written as follows.

$$\Delta t \approx C_1 + C_2\frac{\Delta W}{n^{2.21}} + C_3\Delta P$$

Next, we replace Δt by $t_{normal} - t$ where t means t_{isomer}, ΔW by $W_{normal} - W$ where W means W_{isomer}, and ΔP by $P_{normal} - P$ where P means P_{isomer}. After making these replacements, we obtain the following.

$$t_{normal} - t \approx C_1 + C_2 \frac{W_{normal} - W}{n^{2.21}} + C_3(P_{normal} - P)$$

Solving for t we obtain the following.

$$t \approx C_4 + C_2 \frac{W}{n^{2.21}} + C_3 P + f(n)$$

Here $f(n)$ is represents all terms involving the normal data. In Model 6.3, we replaced $f(n)$ by $C_8(\ln n)^{1.15584}$. Also, we used FILTER to adjust the power on n from 2.21 to 2.43001 and the power on P from 1 to 1.00409. Thus, we have

$$t \approx C_5 + C_6 \frac{W}{n^{2.43001}} + C_7 P^{1.00409} + C_8(\ln n)^{1.15584}$$

Model 6.3 *This model uses the Wiener W number, the number of carbons n, and the Wiener ^3P number. The formula is*

$$f_{6.3}(W, n, P) = -180.043 + 225.64 \frac{W}{n^{2.43001}} + 5.24302 P^{1.00409} + 67.4542(\ln n)^{1.15584}$$

The error vector is $(0.99351, 142, 2.2)$.

Models 6.1 and 6.2 approximate Δt, while Model 6.3 approximates BP (t itself). Model 6.3 slightly improves the standard deviation of Model 6.2.

7   Multivariable Models

Multivariable boiling point models were developed to rationalize existing data and to predict new or unknown boiling point data. The boiling point models in this section use data from Beilstein and NIST [2, 19]. Indices based on the graphical properties of their carbon trees are used to construct these multivariable models. Similar work was done by Espinosa and others in which they used neural nets to predict the boiling points

for aliphatic hydrocarbons [7]. Their work was based on the 142 alkanes with six to ten carbons. Neural nets provide an approach to predict the boiling points of alkanes but do not provide equations for the models. By using the NEOS solver FILTER and Mathematica, we obtained equations for the boiling point models. These models were developed both to rationalize existing data and to predict new or unknown data.

Initially models were developed using the alkanes having one to twelve carbon atoms for which boiling point data was available and the normal alkanes with thirteen through twenty-two carbons. We included the normal alkanes having thirteen to twenty-two carbons because our test data was composed of miscellaneous compounds having thirteen to twenty-two carbons compiled from Beilstein and NIST [2, 19].

Model 7.1 was developed using the Wiener path numbers from one to eight and the Methyl and Ethyl indices. We used all the Wiener path numbers since there is not a significant correlation between most of these numbers as seen in Table II. The ethyl index was omitted because FILTER converged to a solution having a zero value for the coefficient of the ethyl index.

Model 7.1 *This model uses the Wiener ^iP numbers $(1 \leq i \leq 8)$, and the Mth index. The formula is*

$$f_{7.1}(^1P, ^2P, \ldots, ^8P, Mth) = -167.49997 + 84.28344(^1P)^{0.46072} + 15.94534(^2P)^{0.00348} +$$
$$17.42198(^3P)^{0.53517} + 11.16515(^4P)^{0.00164} + 4.74582(^5P)^{0.00089} + 5.23270(^6P)^{0.00143} +$$
$$6.67018(^7P)^{0.14687} + 5.27829(^8P)^{0.96677} - 5.53723Mth^{0.00072}.$$

The error vector is $(0.995322, 195, 4.3)$.

Table III gives the number and percentage of alkanes with the specified absolute boiling point deviations given by Model 7.1. The agreement between the boiling point calculated with a model and the experimental boiling point is given by the absolute boiling point deviation (BP Dev.), BP Dev. = |Model BP − Experimental BP|.

In an attempt to improve Model 7.1, we added the natural logarithm Hosoya index. This improvement can be seen in the drop of the standard deviation from 4.3 degrees for Model 7.1 to 3.0 degrees for Model 7.2.

Table III: Alkanes and their Corresponding Absolute BP Deviations for Model 7.1

BP Dev.	# alkanes	% alkanes
$0-1°$	49	25.1
$1-2°$	35	17.9
$2-4°$	52	26.7
$4-6°$	32	16.4
$6-9°$	22	11.3
$9-11°$	2	1.0
$>11°$	3	1.5

Model 7.2 *This model uses the Wiener ^iP numbers ($1 \leq i \leq 8$), the Mth index, and the Hosoya index. The formula is*

$$f_{7.2}(^1P, ^2P, \ldots, ^8P, Mth, Z) = -157.39801 - 1.71578(^1P)^{2.06741} + 0.38684(^2P)^{1.98738} -$$
$$3.53937(^3P)^{0.73904} - 4.69814(^4P)^{0.04224} - 2.31936(^5P)^{0.85500} - 0.00004(^6P)^{5.13758} +$$
$$1.57097(^7P)^{0.57937} + 0.23048(^8P)^{2.70854} + 1.20993Mth^{1.32159} + 108.36449(\ln Z)^{1.00972}.$$

The error vector is $(0.997648, 195, 3.0)$.

Table IV gives the number and percentage of alkanes with the specified absolute boiling point deviations given by Model 7.2.

Table IV: Alkanes and their Corresponding Absolute BP Deviations for Model 7.2

BP Dev.	# alkanes	% alkanes
$0-1°$	65	33.3
$1-2°$	54	27.7
$2-4°$	53	27.2
$4-6°$	18	9.2
$6-9°$	4	2.1
$9-10°$	0	0
$>10°$	1	0.5

Table V gives the boiling points as calculated by Models 7.1 and 7.2 for the first eight alkanes, $n = 1$ to $n = 5$. They do a fairly good job of predicting these boiling points, particularly Model 7.2.

Including the first eight alkanes in the models constrains the intercept and the coefficients and exponents of some of the indices, particularly ^1P, ^2P, and Z. This lead us to explore a model omitting the first eight alkanes having one to five carbon atoms.

Table V: C1-C5 BP determined by Models 7.1 and 7.2

Alkane	Exp BP	Model 7.1	Model 7.2
methane	-162.15	-167.5	-157.4
ethane	-88.55	-83.22	-84.27
propane	-42.15	-35.56	-45.04
butane	-0.15	5.72	-0.82
2-methylpropane	-11.15	-17.22	-14.72
n-pentane	36.05	44.55	32.24
2-methylbutane	27.03	27.86	27.9
2,2-dimethylpropane	9.45	2.64	8.85

The following model uses the same data set as Models 7.1 and 7.2 with the exception of the first eight alkanes having one to five carbons. The indices used in this model are the Wiener path numbers from one to six, the Hosoya index, and the Methyl index. The ethyl index, ^7P, and ^8P were omitted from this model because FILTER converged to a solution having a zero value for their coefficients.

Model 7.3 *This model uses the Wiener ^iP numbers $(1 \leq i \leq 6)$, the Mth index, and the Hosoya index. The formula is*

$$f_{7.3}(^1P, ^2P, \ldots, ^6P, Mth, Z) = 847.41474 + 221.61698(^1P)^{0.49420} - 1182.20853(^2P)^{0.03689} +$$
$$0.00125(^3P)^{3.39724} - 3.02445(^4P)^{0.93751} - 2.16070(^5P)^{1.01631} - 0.56366(^6P)^{1.38233} -$$
$$2.10575Mth^{0.5695} - 9.61075Z^{0.19907}.$$

The error vector is $(0.997068, 187, 2.7)$.

Table VI gives the number and percentage of alkanes with the specified absolute boiling point deviations given by Model 7.3. This model fits the data significantly better with 70% of the alkanes having an error less than two degrees.

Table VI: Alkanes and their Corresponding Absolute BP Deviations for Model 7.3

BP Dev.	# alkanes	% alkanes
$0 - 1°$	84	44.9
$1 - 2°$	47	25.1
$2 - 4°$	46	24.6
$4 - 6°$	5	2.7
$6 - 9°$	3	1.6
$9 - 10°$	1	0.5
$> 10°$	1	0.5

Figure 5 shows a plot of the experimental boiling points versus the model boiling points determined by Model 7.3. The straight line represents an exact prediction.
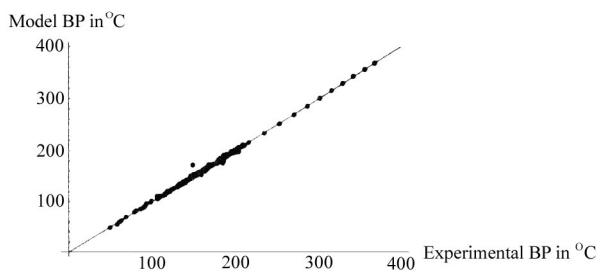
Figure 5: Experimental BP vs. Model BP for Model 7.3

8 Using Multivariable Models to Predict New Data

Using both the Beilstein Crossfire database [2] and the NIST Chemistry WebBook [19], we gathered boiling point data for 52 additional alkanes having between 13 and 22 carbons. We then examined how well the models predicted the boiling points of these 52 new alkanes. One of the alkanes used in testing the predictive ability of our models had conflicting experimental boiling point data. 2,4-dimethyltetradecane is reported as 266 degrees in Beilstein Crossfire and is reported as 247.85 degrees in the NIST Chemistry WebBook. When there was a discrepancy between the two sources, Beilstein was used except in the case of 4-methyl-6-propylundecane.

Figure 6 shows a plot of the experimental boiling points for the 52 additional alkanes versus their model boiling points determined by Model 7.3. The straight line represents an exact prediction.

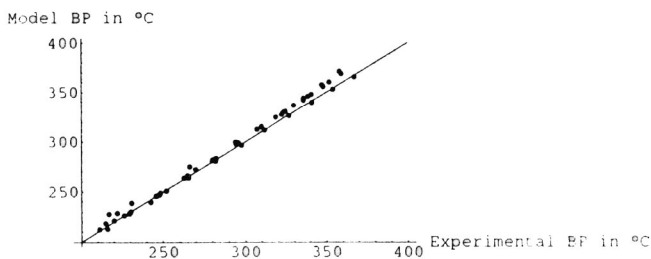Figure 6: Experimental BP vs. Model BP for Model 7.3

Table VII gives the experimental boiling points of these 52 alkanes, the boiling points predicted by Model 7.2 and Model 7.3, and the corresponding boiling point deviations.

Table VII: Predictive Ability of Model 7.2 and Model 7.3

Alkane	Exp BP	Model 7.2		Model 7.3	
		Calc BP	BP Dev.	Calc BP	BP Dev.
2,5,9-trimethyldecane	211.1	210.2	0.9	212.7	1.6
2,5-dimethylundecane	215.0	214.5	0.5	218.4	3.4
3,3-diethylnonane	222.1	222.5	0.4	228.9	6.8
2-methyldodecane	229.5	227.2	2.3	228.2	1.3
2,3-dimethylundecane	226.3	228.3	2.0	226.7	0.4
3-methyldodecane	230.2	231.0	0.8	230.0	0.2
2,2-dimethylundecane	220.0	222.3	2.3	221.4	1.4
4,5-diethyl-2,7-dimethyloctane	215.9	202.2	13.6	212.9	2.9
2,4,4,5,5,7-hexamethyloctane	216.9	216.2	0.6	228.1	11.2
4-propylundecane	231.0	242.8	11.9	239.0	8.0
2-methyltridecane	247.9	245.5	2.4	246.6	1.3
3-methyltridecane	248.4	249.3	0.9	248.7	0.3
4-methyltridecane	246.5	246.7	0.2	246.4	0.1
2,3-dimethyl-dodecane	245.9	247.4	1.5	246.0	0.1
2,6,10-trimethyldodecane	252.0	234.7	17.2	251.2	0.7
4-methyl-6-propylundecane	242.6	220.7	21.9	239.8	2.8
2,3-dimethyltridecane	263.7	266.0	2.3	264.5	0.8
4-methyltetradecane	262.7	264.5	1.8	264.2	1.5
3-methyltetradecane	265.0	267.2	2.2	266.5	1.5
2-methyltetradecane	265.4	263.3	2.1	264.1	1.3
7,8-dimethyltetradecane	269.9	273.0	3.2	272.3	2.5
3-methylpentadecane	282.0	284.5	2.5	283.4	1.4
2,4-dimethyltetradecane	266.0	278.5	12.5	275.1	9.1
2-methylpentadecane	281.9	280.6	1.3	280.7	1.2
4-methylpentadecane	279.7	281.8	2.1	281.1	1.4
2,3-dimethyltetradecane	280.2	284.1	3.9	282.1	1.9
2,3-dimethylpentadecane	295.5	301.7	6.2	298.9	3.4

Table VII cont.

Alkane	Exp BP	Model 7.2 Calc BP	BP Dev.	Model 7.3 Calc BP	BP Dev.
4-methylhexadecane	294.3	298.7	4.4	297.3	3.0
3-methylhexadecane	294.0	301.4	7.4	299.5	5.5
2-methylhexadecane	297.7	297.4	0.3	296.4	1.3
2-methylheptadecane	312.0	313.8	1.8	311.5	0.5
3-methlheptadecane	310.0	317.7	7.7	314.9	4.9
4-methylheptadecane	307.1	315.1	8.0	312.8	5.7
2,3-dimethyl-hexadecane	310.0	318.8	8.8	315.2	5.2
2,3-dimethyl-heptadecane	324.9	335.2	10.3	330.8	5.9
4-methyloctadecane	322.5	330.7	8.2	327.6	5.1
3-methyloctadecane	323.5	333.4	9.9	329.8	6.3
2-methyloctadecane	327.0	329.4	2.4	325.9	1.1
3-ethyloctadecane	340.9	349.6	8.8	347.9	7.1
2,6,11,15-tetramethylhexadecane	318.9	257.2	61.6	324.6	5.7
2,2-dimethyloctadecane	330.0	349.1	19.1	336.7	6.7
2,3-dimethyloctadecane	338.8	350.5	11.7	345.9	7.1
4-methylnonadecane	336.0	345.2	9.2	341.9	5.9
3-methylnonadecane	336.0	347.9	11.9	344.0	8.0
2-methylnonadecane	341.0	343.9	2.9	339.6	1.4
2-methyleicosane	354.0	356.9	2.9	352.8	1.2
3-methyleicosane	347.0	360.9	13.9	357.6	10.6
4-methyleicosane	348.0	358.3	10.3	355.6	7.6
2,3-dimethylnonadecane	351.8	364.2	12.4	360.4	8.6
2-methylheneicosane	367.0	367.8	0.8	365.4	1.6
3-methylheneicosane	358.0	371.8	13.8	370.8	12.8
4-methylheneicosane	359.0	369.2	10.2	368.8	9.8

9   Conclusions

Single variable models including the lower alkanes as data have a low standard deviation when the equation includes the selected index in multiple terms with variable powers. In an attempt to improve Wiener's model, Model 6.1, we see that Model 6.2 with a standard deviation of 2.3 fits the data significantly better than Model 6.1, which has a standard deviation of 3.1.

Model 7.3 does an outstanding job of predicting the fifty-two additional alkanes listed in Section 8. The standard deviation of Model 7.3 for this set of data is 4.1 degrees, the largest individual error being 12.8 degrees. Many of the boiling points are predicted extremely well: twenty-four of the fifty-two alkanes have an absolute boiling point deviation less than 2 degrees.

Rücker and Rücker [26] compiled and weighted the boiling point data for the saturated hydrocarbons for $1 \leq n \leq 10$, for which boiling point data was available. They assigned a weight to each alkane based on the reliability of the boiling point data. Appendix A lists the boiling point data we used from Beilstein and NIST, the boiling point data from Rücker and Rücker and their corresponding weights, and the calculated boiling points from Models 7.2 and 7.3. Our models predict the Rücker and Rücker boiling point data somewhat better than the Beilstein and NIST boiling point values we used to create the models.

A   Beilstein/NIST vs. Rücker and Rücker BP's for Models 7.2 and 7.3

Alkane	Beilstein/NIST BP	Rücker BP	Weight	Model 7.2 BP	Model 7.3 BP
methane	-162.2	-161.5	2	-157.4	
ethane	-88.6	-88.6	2	-84.3	
propane	-42.2	-42.1	2	-45.0	
butane	-0.1	0.5	2	-0.8	
2-methylpropane	-11.2	-11.7	2	-14.7	
pentane	36.1	36.0	2	32.2	
2-methylbutane	27.0	27.8	2	27.9	
2,2-dimethylpropane	9.5	9.5	2	8.8	
hexane	68.8	68.7	2	66.3	70.2
2-methylpentane	60.9	60.3	2	56.5	59.5
3-methylpentane	63.3	63.3	2	62.6	63.1
2,2-dimethylbutane	49.8	49.7	2	50.4	49.4
2,3-dimethylbutane	58.1	58.0	2	57.7	55.7
heptane	98.5	98.5	2	96.9	99.2
3-ethylpentane	93.5	93.5	2	96.4	96.3
3-methylhexane	91.8	92	2	91.3	91.9
2-methylhexane	90.1	90	2	87.0	88.7
2,3-dimethylpentane	89.8	89.8	2	88.6	88.4
3,3-dimethylpentane	86.1	86.1	2	85.5	85.9
2,4-dimethylpentane	80.6	80.5	2	80.4	82.4
2,2-dimethylpentane	79.2	79.2	2	76.1	79.7
2,2,3-trimethylbutane	81.0	80.9	2	80.7	81.9
octane	125.6	125.7	2	126.2	125.9
3-methylheptane	118.9	118.9	2	118.8	118.1
3-ethylhexane	118.6	118.5	2	121.2	120.7

Alkane	Beilstein/NIST BP	Rücker BP	Weight	Model 7.2 BP	Model 7.3 BP
3-ethyl-3-methylpentane	118.4	118.2	2	120.9	118.1
3,4-dimethylhexane	117.9	117.7	2	119.4	117.0
4-methylheptane	117.8	117.7	2	116.9	117.9
2-methylheptane	117.6	117.6	2	114.9	116.2
3-ethyl-2-methylpentane	115.7	115.6	2	117.7	116.7
2,3-dimethylhexane	115.9	115.6	2	113.3	113.7
2,3,3-trimethylpentane	114.7	114.8	2	112.7	114.0
2,3,4-trimethylpentane	113.7	113.5	2	114.2	112.2
3,3-dimethylhexane	112.0	112.0	2	110.8	111.5
2,2,3-trimethylpentane	110.0	109.8	2	109.2	110.7
2,4-dimethylhexane	108.9	109.4	2	110.6	110.9
2,5-dimethylhexane	109.0	109.1	2	107.0	108.2
2,2-dimethylhexane	106.9	106.8	2	103.4	105.9
2,2,3,3-tetramethylbutane	106.5	106.5	2	105.8	110.9
2,2,4-trimethylpentane	99.3	99.2	2	100.2	101.7
nonane	150.7	150.8	2	151.6	150.7
3-ethylheptane	143.3	143.0	2	145.2	144.1
4-ethylheptane	141.3	142.1	2	143.0	142.6
3-methyloctane	143.9	144.0	2	145.2	143.1
4-methyloctane	141.9	142.4	2	142.9	141.7
3,3-diethylpentane	146.2	145.0	1	151.6	148.6
4-ethyl-3-methylhexane	140.4	140.4	1	144.3	141.9
2-methyloctane	143.0	142.8	2	142.5	141.9
3,4-dimethylheptane	140.5	140.6	2	141.2	139.8
3-ethyl-2-methylhexane	138	138.0	1	138.7	137.9

Alkane	Beilstein/NIST BP	Rücker BP	Weight	Model 7.2 BP	Model 7.3 BP
3,5-dimethylheptane	135.6	136.0	1	138.6	136.9
3-ethyl-3-methylhexane	140.6	140.6	1	141.8	139.7
2,3-dimethylheptane	139.9	140.5	1	138.2	137.8
4-ethyl-2-methylhexane	133.8	133.8	1	136.0	135.0
2,5-dimethylheptane	134.9	136.0	2	135.7	135.1
2,3,4-trimethylhexane	138.9	140.5	1	140.3	137.5
3,3-dimethylheptane	136.9	137.3	2	135.4	134.8
2,4-dimethylheptane	132.9	133.5	2	133.6	134.8
3-ethyl-2,3-dimethylpentane	141.6	142.0	1	144.5	143.4
2,6-dimethylheptane	134.9	134.0	2	133.2	134.2
3,3,4-trimethylhexane	140.2	140.5	1	140.7	139.8
3-ethyl-2,4-dimethylpentane	136.7	136.7	1	138.9	137.0
4,4-dimethylheptane	134.9	135.2	1	133.2	134.4
2,2-dimethylheptane	131.9	132.7	2	129.3	131.5
2,3,5-trimethylhexane	130.9	131.3	1	131.1	131.1
2,3,3-trimethylhexane	137.1	137.7	1	133.8	136.0
3-ethyl-2,2-dimethylpentane	133.8	133.8	1	135.5	136.0
2,2,3-trimethylhexane	133.1	133.6	1	130.4	133.0
2,4,4-trimethylhexane	130.9	130.7	1	130.3	129.5
2,3,3,4-tetramethylpentane	141.5	141.5	1	139.5	140.6
2,2,4-trimethylhexane	126.5	126.5	2	126.8	126.8
2,2,5-trimethylhexane	123.9	124.0	1	122.9	124.3
2,2,3,4-tetramethylpentane	133	133.0	1	135.3	133.1
2,2,3,3-tetramethylpentane	140.3	140.2	1	135.6	140.4
2,2,4,4-tetramethylpentane	122.3	122.3	2	122.3	120.1

Alkane	Beilstein/NIST BP	Rücker BP	Weight	Model 7.2 BP	Model 7.3 BP
decane	174.2	174.1	2	174.4	173.7
2-methylnonane	166.9	166.8	2	166.2	165.7
3-methylnonane	167.8	167.8	2	168.7	167.0
4-methylnonane	166.9	165.7	2	165.2	164.6
5-methylnonane	165.1	165.1	2	165.1	163.5
3-ethyloctane	166.0	168.0	1	169.1	167.1
4-ethyloctane	163.4	164.0	1	165.5	163.7
2,2-dimethyloctane	154.0	155.0	1	155.6	156.4
2,3-dimethyloctane	164.2	164.0	2	162.7	161.5
2,4-dimethyloctane	155.9	153.0	1	157.8	157.3
2,5-dimethyloctane	156.9	157.0	2	157.9	157.4
2,6-dimethyloctane	158.9	158.2	2	160.9	159.5
2,7-dimethyloctane	160.1	160.0	2	159.5	159.9
3,3-dimethyloctane	161.2	161.2	1	159.6	158.1
3,5-dimethyloctane	158.9	159.0	1	159.8	158.5
4,4-dimethyloctane	157.5	160.0	1	156.1	155.6
4,5-dimethyloctane	162.4	162.4	1	161.4	160.5
4-propylheptane	161.9	161.8	1	161.2	161.3
4-isopropylheptane	158.8	159.5	1	156.8	156.7
3-ethyl-3-methylheptane	163.9	163.8	1	163.0	160.7
5-ethyl-3-methylheptane	158.3	160.0	1	161.0	159.0
4-ethyl-4-methylheptane	159.0	166.0	1	159.8	159.1
2,2,4-trimethylheptane	148.9	147.7	1	147.7	148.7
2,2,5-trimethylheptane	146.0	147.0	1	149.3	149.2
2,2,6-trimethylheptane	148.3	148.2	1	147.7	148.8

48

Alkane	Beilstein/NIST BP	Rücker BP	Weight	Model 7.2 BP	Model 7.3 BP
2,3,3-trimethylheptane	160.1	160.0	1	155.8	157.4
2,3,4-trimethylheptane	155.9	163.0	1	159.5	158.4
2,3,6-trimethylheptane	149.9	155.7	1	154.8	154.0
2,4,4-trimethylheptane	149.6	152.0	1	150.0	150.3
2,4,6-trimethylheptane	146.8	145.0	1	149.9	150.7
2,5,5-trimethylheptane	152.8	152.8	1	151.8	150.9
3,3,5-trimethylheptane	155.7	155.7	1	154.8	152.8
3,3-diethylhexane	164.5	166.3	1	168.2	166.9
3,4-diethylhexane	162.8	162.0	1	165.8	164.6
2,2-dimethyl-4-ethylhexane	147.1	147.0	1	148.9	147.9
2,4-dimethyl-4-ethylhexane	158.0	158.0	1	156.8	154.3
2,2,3,3-tetramethylhexane	158.9	160.0	1	153.1	159.6
2,2,3,4-tetramethylhexane	156.7	155.0	1	157.7	155.9
2,2,3,5-tetramethylhexane	148.7	148.7	1	147.5	148.7
2,2,4,5-tetramethylhexane	148.0	147.9	1	146.6	145.4
2,2,5,5-tetramethylhexane	137.2	137.0	1	140.1	139.3
2,3,3,4-tetramethylhexane	164.6	164.0	1	162.8	164.4
2,3,3,5-tetramethylhexane	152.9	153.0	1	151.4	151.9
2,3,4,4-tetramethylhexane	162.3	162.0	1	162.1	159.8
2,3,4,5-tetramethylhexane	155.5	158.0	1	159.5	156.5
3,3,4,4-tetramethylhexane	170.3	170.5	1	164.8	169.0
2,4-dimethylpentane-3-isopropyl	157.0	157.0	0	159.0	155.8
2,2,4-trimethyl-3-ethylpentane	155.3	155.3	1	157.1	155.3
2,3,4-trimethyl-3-ethylpentane	169.4	169.4	0	167.7	168.8
2,2,3,3,4-pentamethylpentane	166.1	166.0	1	163.0	167.0
2,2,3,4,4-pentamethylpentane	159.3	159.3	1	158.9	153.1

Alkane	Beilstein/NIST BP	Rücker BP	Weight	Model 7.2 BP	Model 7.3 BP
2,3-dimethyl-3-ethylhexane	169	169	1	163.8	163.8
2,3-dimethyl-4-ethylhexane	164	164	1	161.4	161.5
2,4-dimethyl-3-ethylhexane	164	164	1	161.2	161.4
2,4-dimethyl-4-ethylhexane	158	158	1	157.3	157.4
2,5-dimethyl-3-ethylhexane	157	157	1	154.7	154.7
3,3-dimethyl-4-ethylhexane	165	165	1	164.3	164.5
3,4-dimethyl-3-ethylhexane	170	170	1	167.1	167.4
2,2,3,3-tetramethylhexane	160.3	160	1	162.0	161.9
2,2,3,4-tetramethylhexane	154.9	155	1	158.0	158.0
2,2,3,5-tetramethylhexane	148.4	148.7	1	151.9	151.6
2,2,4,4-tetramethylhexane	153.3	153	1	148.0	148.2
2,2,4,5-tetramethylhexane	147.9	147.9	1	148.4	148.1
2,2,5,5-tetramethylhexane	137.5	137	1	141.8	141.9
2,3,3,4-tetramethylhexane	164.6	164	1	164.3	164.4
2,3,3,5-tetramethylhexane	153	153	1	154.7	154.4
2,3,4,4-tetramethylhexane	162.2	162	1	160.6	160.8
2,3,4,5-tetramethylhexane	161	158	1	157.8	157.7
3,3,4,4-tetramethylhexane	170.0	170.5	1	167.8	168.3
2,4-dimethyl-3-isopropylpentane	157.0	157	0	156.4	156.7
2-methyl-3,3-diethylpentane	174	174	1	169.3	169.9
2,2,3-trimethyl-3-ethylpentane	168	168	1	167.1	167.7
2,2,4-trimethyl-3-ethylpentane	155.3	155.3	1	156.7	157.1
2,3,4-trimethyl-3-ethylpentane*	169.4	169.4	0	166.2	166.6
2,2,3,3,4-pentamethylpentane	166.1	166	1	164.7	165.2
2,2,3,4,4-pentamethylpentane	159.3	159.3	1	154.0	154.7

*2,3-dimethyl-3-isopropylpentane

References

[1] Balaban, Alexandru T., A Personal View About Topological Indices for QSAR / QSPR, *QSPR / QSAR Studies by Molecular Descriptors*, Nova Science Publishers, Huntington, NY, 2001, 1-30.

[2] Beilstein Crossfire database, http://www.beilstein.com/products/xfire/.

[3] Biggs, Norman L., Lloyd, E. Keith and Wilson, Robin J., *Graph Theory: 1736-1936*, Oxford Univ. Press, London, 1976.

[4] Devore, Jay L. and Peck, Roxy, *Statistics: the exploration and analysis of data*, Third edition, Duxbury Press, Belmont, CA, 1997.

[5] Doss, Milburn Price, *Physical constants of the principal hydrocarbons*, Fourth edition, The Texas Co., New York, 1943.

[6] Dreisbach, Robert Richert, *Physical properties of chemical compounds - II*, Advances in chemistry series, no. 22, Amer. Chem. Soc., Washington, D. C., 1959.

[7] Espinosa, Gabriela, Yaffe, Denise, Cohen, Yoram, Arenas, Alex, and Giralt, Francesc, Neural Network Based Quantitative Structural Property Relations (QSPRs) for Predicting Boiling Points of Aliphatic Hydrocarbons, *J. Chem. Inf. Comput. Sci.* **40** (2000) 859-879.

[8] Farrell, Edward J., An introduction to matching polynomials, *J. Combinat. Theory Ser. B* **27** (1979) 75-86.

[9] Fletcher, Roger and Leyffer, Sven, Nonlinear programming without a penalty function, *Math. Program. Ser. A*, **91** (2002) 239-269.

[10] Fourer, Robert, Gay, David M., and Kernighan, Brian W., *AMPL: A Modeling Language for Mathematical Programming*, Boyd and Fraser Publishing Co., Danvers, MA, 1993.

[11] Gill, Philip E., Murray, Walter, and Saunders, Michael A., SNOPT: An SQP algorithm for large-scale constrained optimization, *SIAM Journal on Optimization* **12** (2002) 979-1006.

[12] Gutman, Ivan and Miljković, Olga, Molecules with smallest connectivity indices, *MATCH - Commun. Math. Comput. Chem.* **41** (2000) 57-70.

[13] Gutman, Ivan and Polansky, Oskar E., *Mathematical concepts in organic chemistry*, Springer-Verlag, Berlin, 1986.

[14] Hall, Kenneth R. [and others], Selected Values of Properties of Hydrocarbons and Related Compounds, *formerly API Research Project 44: v. A-78*, Texas A&M University, 1979.

[15] Holland, Jonathan Earl and Whitehead, Earl Glen Jr, On matching equivalent graphs, *J. Combin. Math. & Combin. Computing* **28** (1998) 227-235.

[16] Kinney, Corliss R. and Spliethoff, William L., A system correlating molecular structure of organic compounds with their boiling points. VII. new boiling points for certain paraffins and olefins, *J. Org. Chem.* **14** (1949) 71-78.

[17] Konstantinova, Elena V. and Skorobogatov, Vladimir A., Application of hypergraph theory in chemistry, *Discrete Math.* **235** (2001) 365-383.

[18] The NEOS Guide, Available online at http://www-fp.mcs.anl.gov/otc/Guide/.

[19] NIST Chemistry WebBook, Available online at http://webbook.nist.gov/chemistry/.

[20] Pomerantz, Philip [and others], Synthesis and physical properties of several aliphatic and alicyclic hydrocarbons, *J. Res. Natl. Bur. Stand.* **52** (1954) 59-65.

[21] Rains, Eric M. and Sloane, Neil J. A., On Cayley's enumeration of alkanes (or 4-valent trees), *J. Integer Sequences* **2** (1999), Article 99.1.1.

[22] Rossini, Frederick Dominic [and others], *Selected values of physical and thermodynamic properties of hydrocarbons and related compounds*, Carnegie Press, Carnegie Inst. Tech., Pittsburgh, 1953.

[23] Rossini, Frederick Dominic [and others], *Selected Values of Properties of Hydrocarbons*, Circular of the Natl. Bur. Stand. 461, U. S. Government Printing Office, Washington, DC (1947).

[24] Rouvray, Dennis H., *The role of the topological distance matrix in chemistry,* Mathematics and computational concepts in chemistry (N. Trinajstić, ed.), Ellis Horwood, Chichester, England, 1986, 295-306.

[25] Rouvray, Dennis H., *Combinatorics in chemistry,* Handbook of combinatorics, vol. 2, Elsevier, Amsterdam, 1995, 1955-1981.

[26] Rücker, Gerta and Rücker, Christoph, On Topological Indices, Boiling Points, and Cycloalkanes, *J. Chem. Inf. Comput. Sci,* **39**, (1999) 788-802.

[27] Seifer, A. L. and Smolenskii, E. A., A formal method for correlating the properties of alkanes: I. normal boiling points of monoalkylalkanes, *Russ. J. Phys. Chem.* (English transl.) **37** (1963) 1438-1441.

[28] Todeschini, Roberto and Consonni, Vivianna, *Handbook of Molecular Descriptors,* Wiley-VCH, Weinheim, Germany, 2000.

[29] Wiener, Harry, Structural determination of paraffin boiling points, *J. Amer. Chem. Soc.* **69** (1947) 17-86.

[30] Wiener, Harry, Vapor pressure – temperature relationships among branched paraffin hydrocarbons, *J. Chem. Phys.* **52** (1948) 425-430.