Multicenter Wiener indices and their applications

IVAN GUTMAN[1,2,*], BORIS FURTULA[1] and XUELIANG LI[3]

[1]*Faculty of Science, University of Kragujevac, P. O. Box 60, 34000 Kragujevac, Serbia,*
[2]*State University of Novi Pazar, Novi Pazar, Serbia*
[3]*Center for Combinatorics, Nankai University, Tianjin, 300071, China*

Abstract. The Wiener index W can be viewed as a molecular structure descriptor composed of increments representing interactions between pairs of atoms. A generalization of the W are the Steiner-Wiener indices W_k, k=3,4,.... In the quantity W_k, interactions between k-tuples of atoms play role, based on the concept of Steiner distance. It is shown that the term $W + \lambda W_k$ provides an approximation for the boiling points of alkanes better than W itself. The best such approximation is obtained for $k = 7$.

INTRODUCTION

The Wiener index (W) is one of the oldest and most examined graph-based molecular structure descriptors. For details on its mathematical properties and chemical applications see in the reviews,[1-4] the recent research papers,[5-7] and the references cited therein. On occasion of the fiftieth anniversary of the Wiener index, three special journal issues were published.[8-10] Additional historical data on W can be found in the survey.[11]

The Wiener index is defined in the following manner. Let G be a molecular graph and $v_1, v_2, \ldots, v_n$ be its vertices. The distance between the vertices v_i and v_j, denoted by $d(v_i, v_j)$, is the number of edges in (= the length of) a shortest path that connects v_i and v_j. Then,

$$W = W(G) = \sum_{i<j} d(v_i, v_j) \tag{1}$$

with the summation embracing all pairs of vertices (v_i, v_j) of the molecular graph *G*.

---

*Corresponding author. E-mail: gutman@kg.ac.rs

Bearing in mind that each vertex of the molecular graph represents an atom of the underlying molecule,[12] the quantity W, defined by means of Eq. (1), may be viewed as a sum of structural increments representing *pairs* of atoms, i.e., *two-center* interatomic interactions. From this point of view, one could think of three-center, four-center, etc. interactions that would lead to the following evident multicenter extension of the Wiener-index concept:

$$W_3 = W_3(G) = \sum_{i<j<k} d(v_i, v_j, v_k) \tag{2}$$

$$W_4 = W_4(G) = \sum_{i<j<k<l} d(v_i, v_j, v_k, v_l) \tag{3}$$

$$W_5 = W_5(G) = \sum_{i<j<k<l<m} d(v_i, v_j, v_k, v_l, v_m) \tag{4}$$

etc. In formulas (2)-(4), the meaning of the three-, four-, five-vertex distances needs to be clarified. In fact, long time ago, such a multi-vertex distance was introduced in graph theory by Chartrand et al.[13] and was eventually much studied under the name "*Steiner distance*".[14] Its definition is given in the subsequent section.

The multicenter Wiener-type indices, based on Steiner distance, will be referred to as "*Steiner-Wiener indices*" and are also defined in the subsequent section. The present work is aimed at establishing their chemical applicability.

STEINER DISTANCE AND STEINER-WIENER INDEX

Let G be a connected graph with n vertices. Let $S = \{v_{i_1}, v_{i_2}, \ldots, v_{i_k}\}$ be a set of k distinct vertices of G. Then the *Steiner tree* $T(S)$ is a tree (= connected acyclic graph) that is a subgraph of G, containing all vertices of S, and possessing minimal number of edges. The number of edges of $T(S)$ is the Steiner distance of the vertices $v_{i_1}, v_{i_2}, \ldots, v_{i_k}$. For details on the Steiner-distance concept see elsewhere.[15,16]

For $k = 2, 3, \ldots, n$, the k-th Steiner-Wiener index of the (molecular) graph G is defined as

$$W_k = W_k(G) = \sum_S d(v_{i_1}, v_{i_2}, \ldots, v_{i_k}) \tag{5}$$

where the summation goes over all k-element subsets $S = \{v_{i_1}, v_{i_2}, \ldots, v_{i_k}\}$ of the vertex set of G.

Steiner-Wiener indices W_k were recently considered,[17] and their basic mathematical properties determined. Some of these are the following:

1. $W_2(G)$ coincides with the ordinary Wiener index $W(G)$, Eq. (1).

2. Eqs. (2), (3), and (4) are special cases of Eq. (5), for $k = 3$, $k = 4$, and $k = 5$, respectively.

3. For a graph G with n vertices, if $k = n$, then $W_k(G) = n - 1$.

4. For a graph G with n vertices, if $k > n$, then $W_k(G) = 0$.

5. For a tree T, and for all $k = 2, 3, \ldots, n$,

$$W_k(T) = \sum_e \sum_{i=1}^{k-1} \binom{n_1(e)}{i} \binom{n_2(e)}{k-i} \tag{6}$$

where $n_1(e)$ and $n_2(e)$ are the number of vertices lying on the two sides if the edge e, and where the first summation goes over all edges of T. For all edges e of the tree T, $n_1(e) + n_2(e) = n$.

Note that for $k = 2$, formula (6) reduces to the expression (7), discovered by Wiener himself as early as in 1947:[5,12,18]

$$W(T) = \sum_e n_1(e) n_2(e). \tag{7}$$

STEINER-WIENER INDICES AND BOILING POINTS OF ALKANES

The first chemical application of the Wiener index was its usage for the prediction of the normal boiling points of alkanes.[18] Eventually, correlations with boiling points became a standard test for the quality of topological indices.[19-22] In view of this, we also used this physico-chemical parameter in our studies of the Steiner-Wiener indices.

In Fig. 1 is reproduced the well known[21] plot of the normal boiling points vs. the Wiener index. The curve passing through the data-points is of the form

$$bp \approx bp_{calc}(W^*) = \frac{a + bW^*}{1 + cW^*} \tag{8}$$

where $W^* = W$ and where a, b, c are fitting parameters. The correlation between the experimental and calculated boiling points (i.e., between bp and bp_{calc}, cf. Eq. (8)) are shown in Fig. 2.

Figs. 1 and 2 come about here

The most obvious idea for testing the Steiner-Wiener indices would be to set $W^* = W_k$ into Eq. (8). This, however, did not yield any improvement, and thus had to be abandoned. A better option was to modify the Wiener index as

$$W^* = W + \lambda W_k \tag{9}$$

and to use variable W^* in combination with Eq (8) . For each fixed choice of k, k=3,4,...8, the parameter λ was varied, and its value determined so as to maximize the correlation coefficient for the linear correlation between bp and $bp_{calc}(W^*)$. In all the studied cases, there exists an optimal value for λ at which the correlation coefficients attains a maximum; a characteristic example is shown in Fig. 3.

Fig. 3 comes about here

The results thus obtained are presented in Tables I and II, and in Figs. 4 and 5.

Tables I and II and Figs. 4 and 5 come about here

DISCUSSION AND CONCLUDING REMARKS

Viewing at the Wiener index as a structure descriptor based on two-center interatomic interactions, one could expect that the next-important structural feature will be three-center interactions. In the case of Steiner-Wiener index applied to alkanes, this certainly cannot be the case, since for trees the following identity holds:

$$W_3(T) = \frac{n-2}{2} W(T).\tag{10}$$

Therefore W_3 contains the exactly same structural information as the ordinary Wiener index W.

Relation (10) is deduced from Eq. (6) as follows. For $k = 3$, Eq. (6) has the form

$$W_3(T) = \sum_e \left[\binom{n_1(e)}{1}\binom{n_2(e)}{2} + \binom{n_1(e)}{2}\binom{n_2(e)}{1} \right]$$

which, bearing in mind that $n_1(e) + n_2(e) = n$, is transformed into

$$W_3(T) = \sum_e \left[n_1(e)\frac{n_2(e)[n_2(e)-1]}{2} + n_2(e)\frac{n_1(e)[n_1(e)-1]}{2} \right]$$
$$= \sum_e n_1(e)n_2(e)\frac{n_1(e)+n_2(e)-2}{2} = \frac{n-2}{2}\sum_e n_1(e)n_2(e)$$

Formula (10) is now obtained from Eq. (7).

Our calculations fully agree with the above argument: The accuracy of our models for $k = 2$ and $k = 3$ is same, see Table I.

If $k > 3$, because of the very large number of k-tuples of vertices, the calculation of the Steiner-Wiener index W_k, based on its definition (5), becomes extremely cumbersome. In the case of acyclic systems (such as the molecular graphs of alkanes), instead of Eq. (5) the calculations can be done by using Eq. (6), which is significantly easier. In fact, by means of Eq. (6) any Steiner-Wiener index W_k can be calculated equally easily as the ordinary Wiener index W.

The fact that the accuracy of the approximations based on the indices $W^* = W + \lambda W_k$ increases with k, and reaches its maximum at $k = 7$, is somewhat unexpected. It may be that this is a statistics-based artifact of the models considered. Anyway, this phenomenon deserves to be further examined.

The results of the present study may be considered from a pessimistic and from an optimistic point of view. A pessimist would say that there is very little difference between the Figs. 1 and 4, as well as between Figs. 2 and 5. An optimist would point at the fact that the

average and maximal errors of our best model (based on $W + \lambda W_7$) are, respectively, by 20% and 50% smaller than those of the starting model (based solely on W). In view of this, we may conclude that by adding multicenter distance-contributions to the Wiener index, its applicability to model physico-chemical properties of alkanes is improved, but only to a limited extent.

I Z V O D

VIŠECENTRIČNI VINEROVI INDEKSI I NJIHOVE PRIMENE

IVAN GUTMAN[1,2], BORIS FURTULA[1] i XUELIANG LI[3]

[1]*Prirodno-matematički fakultet Univerziteta u Kragujevcu, Srbija,*
[2]*Državni univerzitet u Novom Pazaru,*
[3]*Center for Combinatorics, Nankai University, Tianjin, China*

Vienerov indeks W se može posmatrati kao molekulski strukturni deskriptor sastavljen od sabiraka koji reprezentuju interakcije izmedu parova atoma. Jedna generalizacija Vinerovog indeksa su Štajner-Vinerovi indeksi W_k , $k=3,4,....$ U indeksu W_k se vodi računa o interakcijama k atoma, zasnovanih na pojmu Štajnerovog rastojanja. Pokazano je da formula $W + \lambda W_k$ omogućava aproksimativno izračunavanje tačke ključanja alkana bolje nego sam Vinerov indeks. Najbolja takva aproksimacije je za $k=7$.

(Primljeno 26. januara 2015)

REFERENCES

1 I. Gutman, Y. N. Yeh, S. L. Lee, Y. L. Luo, *Indian J. Chem.* **32A** (1993) 651
2 S. Nikolić, N. Trinajstić, Z. Mihalić, *Croat. Chem. Acta* **68** (1995) 105
3 D. H. Rouvray, in *Topology in Chemistry - Discrete Mathematics of Molecules*, D. H. Rouvray, R. B. King, Eds., Horwood, Chichester, 2002, pp. 16-37
4 K. Xu, M. Liu, K. C. Das, I. Gutman, B. Furtula, *MATCH Commun. Math. Comput. Chem.* **71** (2014) 461
5 R. Škrekovski, I. Gutman, *MATCH Commun. Math. Comput. Chem.* **72** (2014) 295

6   H. Lin, *MATCH Commun. Math. Comput. Chem.* **72** (2014) 783

7   K. Hrinakova, M. Knor, R. Škrekovski, A. Tepeh, *MATCH Commun. Math. Comput. Chem.* **72** (2014) 791

8   I. Gutman, S. Klavžar, B. Mohar, Eds., 50th Anniversary of the Wiener Index, *Discr. Appl. Math.* **80** (1997), pp. 1-113

9   I. Gutman, S. Klavžar, B. Mohar, Eds., Fifty Years of the Wiener Index, *MATCH Commun. Math. Comput. Chem.* **35** (1997), pp. 1-259

10 I. Gutman, Ed., Fifty Years of the Wiener Index, *J. Serb. Chem. Soc.* **62** (1997), pp. 185-294

11 D. H. Rouvray, in *Topology in Chemistry - Discrete Mathematics of Molecules*, D. H. Rouvray, R. B. King, Eds., Horwood, Chichester, 2002, pp. 1-15

12 I. Gutman, O. E. Polansky, *Mathematical Concepts in Organic Chemistry*, Springer, Berlin, 1986

13 G. Chartrand, O. R. Oellermann, S. Tian, H. B. Zou, *Časopis Pest. Mat.* **114** (1989) 399

14 Jakob Steiner (1795-1863), Swiss mathematician

15 F. K. Hwang, D. S. Richards, P. Winter, *The Steiner Tree Problem*, North-Holland, Amserdam, 1992

16 W. Goddard, O. R. Oellermann, in *Structural Analysis of Complex Networks*, M. Dehmer, Ed., Birkhäuser, Dordrecht, 2011, pp. 49-72

17 X. Li, Y. Mao, I. Gutman, The Steiner-Wiener index of a graph, submitted for publication

18 H. Wiener, *J. Am. Chem. Soc.* **69** (1947) 17

19 D. H. Rouvray, *J. Comput. Chem.* **8** (1987) 470

20 P. G. Seybold, M. May, U. A. Bagal, *J. Chem. Educ.* **64** (1987) 575

21 Z. Mihalić, N. Trinajstić, *J. Chem. Educ.* **69** (1992) 701

22 G. Rücker, C. Rücker, *J. Chem. Inf. Comput. Sci.* **39** (1999) 788

23 R. L. Brown, S. E. Stein, in *NIST Standard Reference Database No. 69*, P. J. Linstrom, W. G. Mallard, Eds., Gaithersburg, 2014, http://webbook.nist.gov

24 J. Nocedal, S. J. Wright, *Numerical Optimization*, Springer, New York, 2006.

Table and Figure Captions

TABLE I. Statistical data for the correlations between boiling points and the topological indices $W^* = W + \lambda W_k$, k=2,3,...,9; R = correlation coefficient, ARE = average relative error (in %), MRE = maximal observed relative error (in %). By boldface are indicated the data for the best approximation. For details, see Eqs. (8) and (9) and the text.

TABLE II. Fitting parameters in formulas (8) and (9), for k=2,3,...,9. The (a,b,c)-values were obtained by means of the scaled Levenberg-Marquardt algorithm.[24] The λ-values are those for which the respective correlation coefficients are maximal, cf. Fig. 3. By boldface are indicated the parameters for the best approximation.

Fig. 1. Correlation between normal boiling points ($bp / {}^{\circ}C$) and Wiener index (W) for the set of all isomeric alkanes with 2 to 9 carbon atoms (74 compounds).[23] The curve passing through the data-points is specified by Eq. (8). Statistical data pertaining to this correlation are found in Tables I and II, for $k = 2$.

Fig. 2. Correlation between the calculated boiling points (bp_{calc}, according to Eq. (8), $W^* = W$) and the experimental boiling points (bp) for the same compounds as in Fig. 1. Statistical data pertaining to this correlation are found in Tables I and II, for $k = 2$.
.

Fig. 3. The λ-dependence of the correlation coefficient R for the correlation between bp and bp_{calc} for the case $k = 5$. The maximum is attained at $\lambda = 0.063$, cf. Table II.

Fig. 4. Normal boiling points ($bp / {}^{\circ}C$) vs. $W + \lambda W_7$ for the same alkanes as in Fig. 1. As the data in Tables I and II. show, the choice $k = 7$ provides the best agreement between bp and bp_{calc}, cf. Eqs. (8) and (9).

Fig. 5. The best correlation between bp and bp_{calc}, obtained by Eqs. (8) and (9) for $k = 7$. Statistical data pertaining to this correlation are found in Tables I and II.

TABLE I. Statistical data for the correlations between boiling points and the topological indices $W^* = W + \lambda W_k$, k=2,3,...,9; R = correlation coefficient, ARE = average relative error (in %), MRE = maximal observed relative error (in %). By boldface are indicated the data for the best approximation. For details, see Eqs. (8) and (9) and the text.

k	R	ARE	MRE
2	0.98954	1.45	8.42
3	0.98957	1.45	8.83
4	0.99018	1.41	9.46
5	0.99135	1.33	8.58
6	0.99256	1.23	6.80
7	**0.99323**	**1.18**	**4.63**
8	0.99273	1.23	4.19
9	0.99149	1.33	6.98

TABLE II. Fitting parameters in formulas (8) and (9), for k=2,3,...,9. The (*a,b,c*)-values were obtained by means of the scaled Levenberg-Marquardt algorithm.[24] The λ-values are those for which the respective correlation coefficients are maximal, cf. Fig. 3. By boldface are indicated the parameters for the best approximation.

k	*a*	*b*	*c*	λ
2	191.328	15.104	0.031	-
3	192.480	14.547	0.031	0.023
4	193.704	14.820	0.032	0.044
5	191.287	16.476	0.037	0.063
6	186.764	18.773	0.043	0.127
7	**181.255**	**21.421**	**0.049**	**0.392**
8	180.547	20.834	0.047	0.802
9	187.788	16.790	0.036	1.400

Fig. 1. Correlation between normal boiling points (bp / ^{o}C) and Wiener index (W) for the set of all isomeric alkanes with 2 to 9 carbon atoms (74 compounds).[23] The curve passing through the data-points is specified by Eq. (8). Statistical data pertaining to this correlation are found in Tables I and II, for $k = 2$.
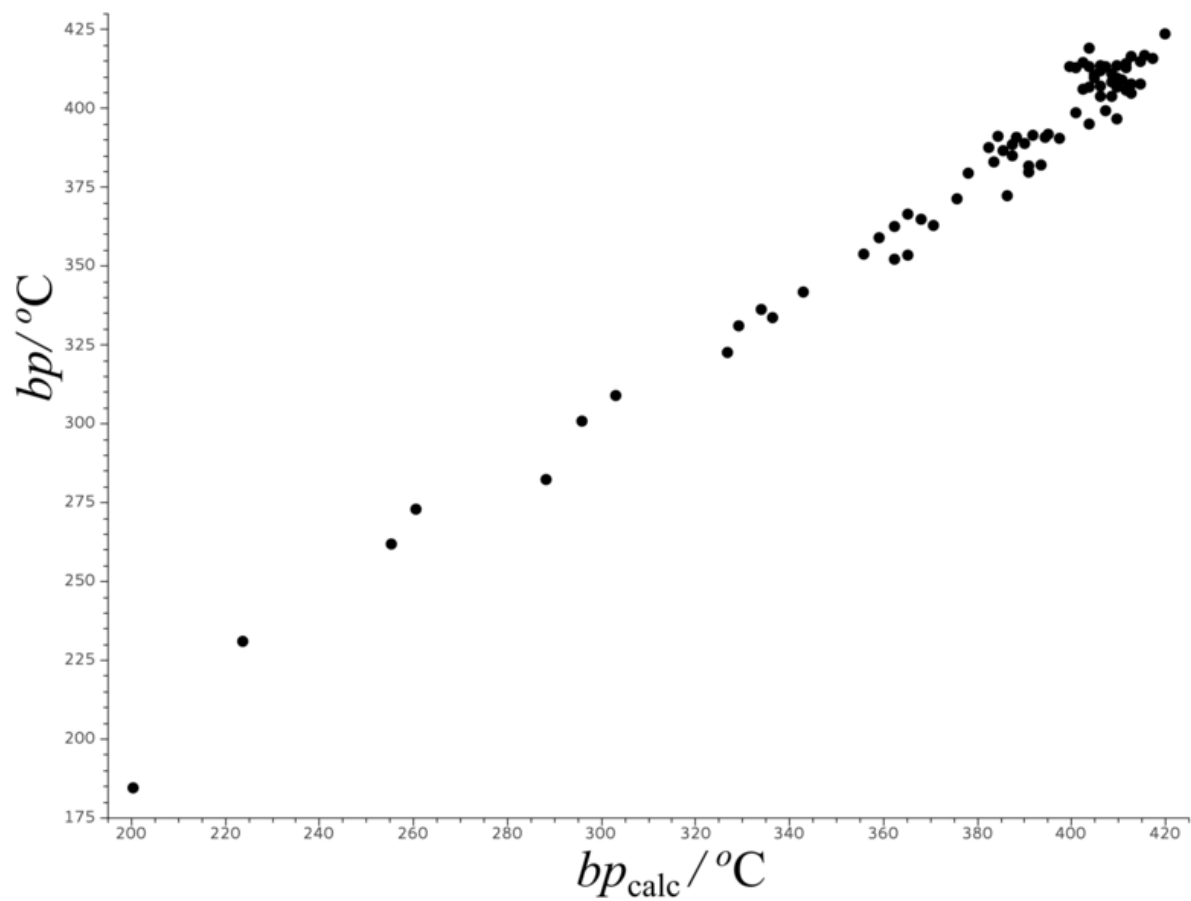
Fig. 2. Correlation between the calculated boiling points (bp_{calc}, according to Eq. (8), $W^* = W$) and the experimental boiling points (bp) for the same compounds as in Fig. 1. Statistical data pertaining to this correlation are found in Tables I and II, for $k = 2$.
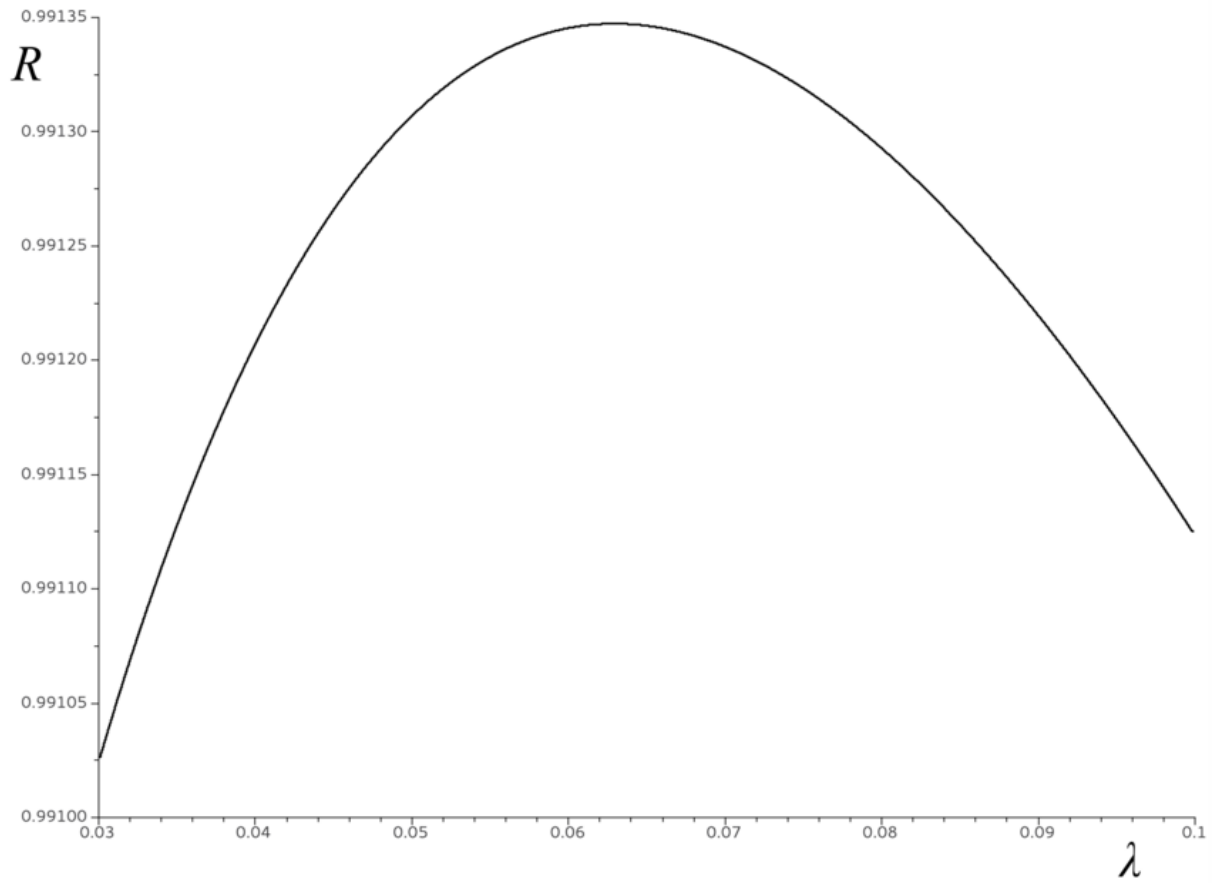
Fig. 3. The λ-dependence of the correlation coefficient R for the correlation between bp and bp_{calc} for the case $k = 5$. The maximum is attained at $\lambda = 0.063$, cf. Table II.
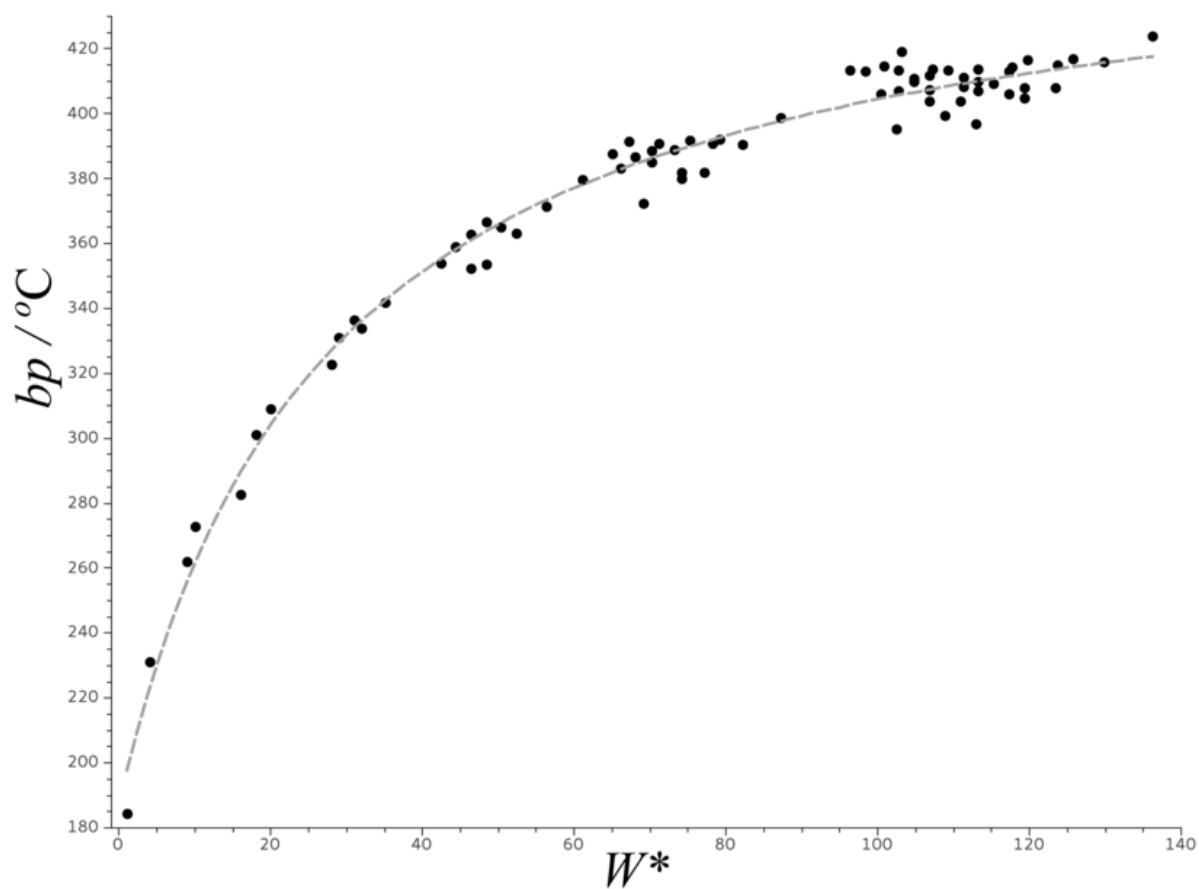
Fig. 4. Normal boiling points ($bp\,/\,{}^oC$) vs. $W + \lambda W_7$ for the same alkanes as in Fig. 1. As the data in Tables I and II. show, the choice $k = 7$ provides the best agreement between bp and bp_{calc}, cf. Eqs. (8) and (9).
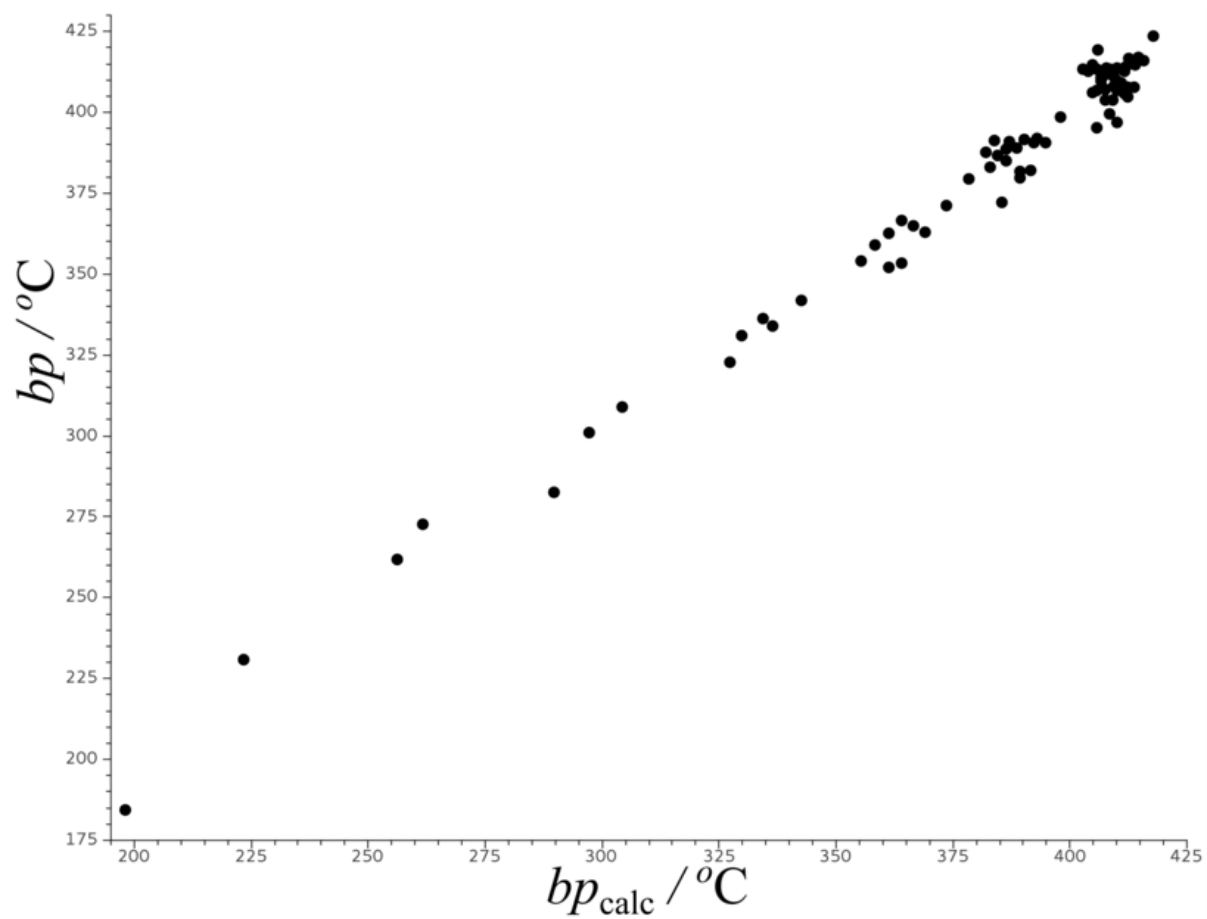
Fig. 5. The best correlation between bp and bp_{calc}, obtained by Eqs. (8) and (9) for $k = 7$. Statistical data pertaining to this correlation are found in Tables I and II.