

Position Group Contribution Method for Predicting the Normal Boiling Point of Organic Compounds

WANG Qiang (王强)^{1,2}, MA Peisheng (马沛生)², WANG Chang (王昶)¹ and XIA Shuqian (夏淑倩)²

¹ School of Material Science and Chemical Engineering, Tianjin University of Science and Technology, Tianjin 300457, China

² School of Chemical Engineering, Tianjin University, Tianjin 300072, China

Abstract A new position group contribution model is proposed for the estimation of normal boiling data of organic compounds involving a carbon chain from C₂ to C₁₈. The characteristic of this method is the use of position distribution function. It could distinguish most of isomers that include *cis*- or *trans*-structure from organic compounds. Contributions for hydrocarbons and hydrocarbon derivatives containing oxygen, nitrogen, chlorine, bromine and sulfur, are given. Compared with the predictions, results made use of the most common existing group contribution methods, the overall average absolute difference of boiling point predictions of 417 organic compounds is 4.2 K; and the average absolute percent derivation is 1.0%, which is compared with 12.3 K and 3.2% with the method of Joback, 12.1 K and 3.1% with the method of Constantinou-Gani. This new position contribution groups method is not only much more accurate but also has the advantages of simplicity and stability.

Keywords normal boiling point, prediction, position group contribution

1 INTRODUCTION

Boiling point is one of the most important thermal properties. Almost other thermo chemical properties are predictable from boiling point and critical constants [1–5], so the precise forecast of boiling point is much needed. Joback and Reid [6] proposed a group contribution method that gives an approximate value of the boiling point of aliphatic and aromatic hydrocarbons. The boiling point is estimated with the sum of contributions of all structural groups found in the molecule. Joback tested this method on 353 compounds. The average error was 16.8 K or 5.0% with standard a deviation of 17.9 K.

$$T_b = 198.2 + \sum_k N_k (t_{b,k}) \quad (1)$$

where T_b is normal boiling point, K; N_k is number of k groups in molecule, $(t_{b,k})$ is Joback contribution of group k to T_b . Boiling point is used in the prediction of many other properties such as critical temperature. The errors in T_b will be propagated to other properties calculated with it. It is recommended to get true value for this very important and fundamental property. Devotta and Pendyala [7] modified the Joback method to treat T_b of halogenated compounds much more accurately. They report that the average percent deviations for refrigerants and other substances was 12% in the original method.

Constantinou and Gani [8, 9] developed an advanced group contribution method based on the UNIFAC groups but were enhanced by allowing more sophisticated functions of the desired properties and by providing contributions at a “Second order” level for T_b , the C-G equations are

$$T_b = 204.359 \ln \left[\sum_k N_k (t_{b1,k}) + W \sum_j M_j (t_{b2,j}) \right] \quad (2)$$

Here, $t_{b1,k}$, $t_{b2,j}$ are contribution of group k and j to T_b .

The first order results are generally good at boiling but not at melting. The second order contributions improve the agreement for all but T_{fp} for the dimethylpentanes, where the correction goes in the wrong direction.

Marrero-Morejon and Pardillo-Fontdevila [10] give two equations for estimating T_b . They call their preferred method a group interaction contribution technique. It can also be considered as a method of bond contributions. They tabulate contributions from 167 pairs of atoms alone or with hydrogen attached. For T_b , their basic equation is

$$T_b = M_w^{-0.404} \sum N_k (t_{bbk}) + 156.00 \quad (3)$$

where M_w is the molecular weight, and N_k is the number of atoms of type k with contributions t_{bbk} .

Yalkowsky and coworkers [11–13] (Yalkowsky *et al.*, 1994; Krzyzaniak *et al.*, 1995; Zhao and Yalkowsky, 1999) have investigated connections between T_{fp} and T_b , as well as having proposed correlations for T_{fp} . This is part of an extensive program to correlate many pure components and mixture properties of complex substances (Yalkowsky *et al.*, 1994). The method consists of both group contributions, which are additive, and molecular descriptors, which are not additive. The main properties of the latter are the symmetry number, σ , and the flexibility number, Φ . The former is similar to that used for ideal gas properties, and the latter is strictly defined as the inverse of the probability that a molecule will be in the conformation of the solid crystalline phase of interest. It is argued that flexibility affects both melting and boiling, whereas symmetry

only affects melting. Using this methodology, molecular properties can be easily determined to estimate the flexibility contribution. Whereas values of σ are to be obtained directly from molecular structure for the Benson group method for gas properties. Thus, Krzyzaniak *et al.* [11, 13] use

$$T_b = \sum_k N_k (b_k) / (86 + 0.4\tau) \quad (4)$$

Table 1 Summary of comparison of estimation methods for T_b [14]

	Substances	$\bar{\varepsilon}$	$\bar{\delta}$ /%	$\delta > 10\%$	$\delta < 5\%$
Joback	353	16.8	5.0	42	242
Constantinou-Gani	341	13.4	4.0	39	279
Marrero-Pardillo	347	7.5	2.0	10	318
Katritzky	175	9.2	2.7	6	154

Note: $\bar{\varepsilon}$ is mean absolute error in the property; $\bar{\delta}$ is mean absolute percentage error; $\delta > 10\%$ is the number of substances for which the absolute percentage error was greater than 10%; $\delta < 5\%$ is the number of substances for which the absolute percentage error was less than 5%.

The advantages of group contribution methods are simple and general. Since organic compounds used in chemical industry consist about only 100 groups, group contribution methods can be applied to a great real number of substances. However, most of group contribution methods have serious problems that they cannot distinguish from structural isomers because the isomers have the same number and kind of groups so that the calculated results are inevitably the same.

This work proposes a new method for estimating the physical properties such as normal boiling point, and defining and utilizing the position parameter of group in organics structure that the most accurate existing method used for predicting the normal boiling point parameters of formation.

2 EXPERIMENTAL DATA

Total of 417 compounds containing carbon, hydrogen, oxygen, nitrogen, chlorine, bromine, and sulfur were used for the determination of group contributions. The complete list is given in the following table and includes linear and branched alkanes (146), cycloalkanes (26), alkenes (32), aromatics (26), alcohols (27), aldehydes and ketones (31), acids (12), phenols and ether oxides (14), esters (22), chloro and bromoalkanes (19), amines (27), nitriles (9), pyridines (10), thiols (14), and thioethers (17). The experiment data from the TRC Thermodynamic Tables [15] priority are used in calculation, which give the critical properties, normal boiling points, and melting point for a large number of hydrocarbons and derivatives; Ma and Poling's handbooks [14, 16] provide a abundance of properties data about organic compound.

3 COMPUTATIONS

3.1 Calculated method

The first step consists testing correlations to represent the properties. Only one-parameter contribution was considered for each group. The normal boiling point function is constructed by all groups' contribution, as well as position correction. Benson's second order groups [17] and a few of third order groups [14] were applied into these works. Here, normal boiling point is expressed as follows [18–20]

$$T_b - T_{b0} = \sum_i A_i N_i + \sum_j A_j \tanh(N_j / N) + \sum_k A_k P_k + a_1 \exp(1/M_w) + a_2 \exp(1/N) \quad (5)$$

$$N = \sum_i N_i + \sum_j N_j \quad (6)$$

Parameter A_i or A_j reveals i or j group contributions, and the set of contributions that allowed minimizing the residual estimation error were then computed by regressive simulation. N_i represents the number of groups that carbon element forms the center of the group, N_j represents the number of groups that non-carbon element forms the center. N is total number of groups. P_k characterizes position factor [18]. T_{b0} is 4775.5907 K, M_w is molecular weight.

Not only were group contributions considered here but also corrections used by position to take into account longer distance interactions were considered. Some corrections for interactions through benzene or pyridine rings were obtained, resulting in a better distinction between series of chain-branched aromatic isomers. Parameters for five-membered and six-membered saturated hydrocarbon rings were also determined. We retained the correlations that were found to statistically give the best estimations.

3.2 Application samples

The examples of T_b computation for the organics is given as follows.

Example 1 Estimation of the boiling point of 2,3,4,4-tetramethylhexane:



This compound is decomposed in position groups as follows:

3 C—(CH)(H)₃; 2 C—(C)(H)₃; 1 C—(CH₂)(H)₃;
1 C—(C)₂(H)₂; 2 C—(C)₃(H); 1 C—(C)₄

Total number of group: $N = 10$

The position factor: position of (CH) group 2, 3, the compensated factor is $2 + 3 = 5$; position of (C) group 4, the compensated factor is 4

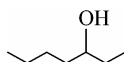
Molecular weight: 142.285

From the contributions in Table 2, the normal boiling point is estimated by Eq. (5):

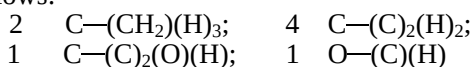
$$T_b = 0.2688 \times 3 - 1.3342 \times 2 + 1.4758 \times 1 + 21.2521 \times 2 + 38.4077 \times 1 + 53.6393 \times 1 + 17.3615 \times \exp(1/10) - 4491.3432 \times \exp(1/142.285) - 0.3567 \times 5 + 2.2817 \times 4 + 4775.5907 = 430.422 \text{ K}$$

Hence, the calculated result is 430.4 K, whereas the experimental boiling point is 434.75 K.

Example 2 Estimation of the boiling point of 3-heptanol:



This compound is decomposed in position groups as follows:



Total number of groups: $N = 8$

The position of (OH) group: the compensated factor is 3

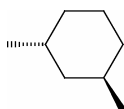
Molecular weight: 116.203

From the contributions in Table 2, the normal boiling point is calculated by Eq. (5):

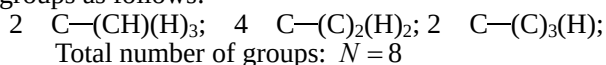
$$T_b = 1.4758 \times 2 + 21.2521 \times 4 + 49.1420 \times 1 + 340.1722 \tanh(1/8) + 17.3615 \times \exp(1/8) - 4491.3422 \times \exp(1/116.034) - 4.1845 \times 3 + 4775.5907 = 431.953 \text{ K}$$

Hence, the calculated result is 431.95 K, whereas the experimental boiling point is 425.15 K.

Example 3 Estimation of the boiling point of *trans*-1,3-dimethylcyclohexane



This compound could be decomposed in position groups as follows:



Total number of groups: $N = 8$

The position factor: position of (CH₃) group 1, 3, the compensated factor is $1 + 3 = 4$

position of *trans*-structure compensated factor is -1
1 cyclohexane correction

Molecular weight: $M_w = 112.215$

From the contributions in Table 1, the critical temperature is estimated by Eq. (5):

$$T_b = 0.2688 \times 2 + 21.2521 \times 4 + 38.4077 \times 2 - 0.3567 \times 4 + 0.7583 \times (-1) - 26.7179 \times 1 + 17.3615 \times \exp(1/8) - 4491.342 \times \exp(1/112.215) + 4775.5907 = 397.18 \text{ K}$$

Hence, the calculated result is 397.18 K, whereas

the experimental boiling point is 397.61 K.

4 RESULTS AND DISCUSSION

The results obtained using the new position group contribution method were compared with experimental boiling points and values predicted using the method of Joback, which is the most extended existing method, as well as the more precise method of Constantinou and Gani. Table 3 summarizes the mean absolute errors ($\bar{\varepsilon}$) or mean absolute percentage error ($\bar{\delta}$) between experimental and predicted values:

$$\varepsilon(\text{K}) = |T_{b,\text{exp}} - T_{b,\text{pred}}| \quad (7)$$

$$\bar{\varepsilon}(\text{K}) = \frac{1}{n} \sum |T_{b,\text{exp}} - T_{b,\text{pred}}| \quad (8)$$

$$\delta(\%) = \left| \frac{T_{b,\text{exp}} - T_{b,\text{pred}}}{T_{b,\text{exp}}} \right| \times 100 \quad (9)$$

$$\bar{\delta}(\%) = \frac{1}{N} \sum_n \left| \frac{T_{b,\text{exp}} - T_{b,\text{pred}}}{T_{b,\text{exp}}} \right| \times 100 \quad (10)$$

The excellent results were obtained for all the chemical families that were under investigation. The mean estimation error being of 4.2 K *versus* 12.3 K obtained the use of Joback's correlation. This confirms the greater precision of methods based on second order Benson's groups for thermochemical data prediction. The second order method of Constantinou and Gani resulted in a mean error of 12.1 K, slightly higher than the accuracy claimed by these authors. It should also be pointed out that Constantinou's T_b estimation method gives higher discrepancies than Joback's for the pyridine derivatives, the amines, and the nitriles, indicating that some contributions need to be revised.

The prediction value in Table 3 indicates that the normal boiling temperature can be described by Eq. (5) very well. Its relation coefficient is 0.9935, and then, we can obtain confident level from incomplete beta function, which can be calculated from incomplete Gamma function. The confidence level is as high as 0.9990, which is the most significance level.

One complementary sets of position group contributions were developed for the predictive estimation of normal boiling point of organic compounds. The second order groups is defined by Benson, whose ideas have been recognized for a long time as more accurate than first order groups for thermochemical predictions, which were used for this purpose. Position factor could distinguish most isomers that include *cis*- or *trans*-structure of organic compounds for their properties. Contributions for compounds containing carbon, hydrogen, oxygen, nitrogen, sulfur, chlorine, and bromine were reported, and results were evaluated with regard to predictive methods developed previously. Normal boiling point predictions exhibit an average

Table 2 Position group contributions for the prediction of T_b

Group	A/K	Group	A/K
C—(CH ₃)(H) ₃	11.7763	C—(C)(Br)(H) ₂	52.5797
C—(CH ₂)(H) ₃	1.4758	C—(C) ₂ (Br)(H)	66.9734
C—(CH)(H) ₃	0.2688	C—(C) ₃ (Br)	90.1982
C—(C)(H) ₃	-1.3343	C—(C)(Cl)(H) ₂	48.0809
C—(C) ₂ (H) ₂	21.2521	C—(C) ₂ (Cl)(H)	59.6140
C—(C) ₃ (H)	38.4077	C—(C) ₃ (Cl)	70.3454
C—(C) ₄	53.6393	C—(C)(Cl) ₂ (H)	61.2838
Cd—(H)(O)	156.3760	C—(S)(H) ₃	53.5025
Cd—(H) ₂	-10.2225	C—(C)(S)(H) ₂	64.2428
Cd—(C)(H)	-4.7170	C—(C) ₂ (S)(H)	76.2968
C—(Cd)(C)(H) ₂	44.1587	C—(C) ₃ (S)	83.8271
C—(Cd)(H) ₃	25.7919	C—(N)(H) ₃	19.3695
Cd—(C) ₂	-19.4800	C—(C)(N)(H) ₂	33.4842
C—(Cd)(C) ₂ (H)	56.8007	C—(C) ₂ (N)(H)	43.1275
Cd—(Cd)(H)	110.7816	C—(C) ₃ (N)	47.1994
C—(Cd)(O)(H) ₂	74.2396	C—(C)(CN)(H) ₂	112.8793
C—(O)(H) ₃	22.4291	C—(C) ₂ (CN)(H)	134.4457
C—(CO)(H) ₃	32.7983	C—(C) ₃ (CN)	121.8172
C—(C)(CO)(H) ₂	58.8031	N—(C)(H) ₂	155.7015
C—(C) ₂ (CO)(H)	68.9866	N—(C) ₂ (H)	96.0855
C—(C) ₃ (CO)	85.9694	N—(C) ₃	-33.1968
C—(C)(O)(H) ₂	39.1793	N—(Cb)(H) ₂	216.3523
C—(C) ₂ (O)(H)	49.1420	N—(Cb)(C)(H)	232.2677
C—(C) ₃ (O)	48.1774	N—(Cb)(C) ₂	91.1189
CO—(CH ₃)(O)	143.0435	NI—(Cb) ₂	241.0696
CO—(CH ₂)(O)	229.4055	S—(C)(H)	29.0358
CO—(CH)(O)	250.8732	S—(C) ₂	-133.9381
CO—(O)(H)	202.0675	<i>ortho</i> -correction ^①	3.7143
CO—(C)(H)	124.5559	<i>meta</i> -correction ^①	-2.0200
CO—(C) ₂	60.2010	cyclopropane correction	-23.8546
CO—(Cd)(O)	413.1023	cyclobutane correction	-25.1213
Cb—(H)	149.7429	cyclopentane correction	-26.7249
Cb—(C)	-100.7275	cyclohexane correction	-26.7179
C—(Cb)(H) ₃	52.4348	<i>Co</i> ^②	-4.5041
C—(Cb)(C)(H) ₂	80.1334	<i>Cm</i> ^②	10.7843
C—(Cb)(C) ₂ (H)	91.3915	<i>Cp</i> ^②	13.0660
C—(Cb)(C) ₃	117.2959	—(CH)<position correction	-0.3567
Cb—(O)	1252.7497	>(C)<position correction	2.2817
O—(Cb)(H)	-482.0553	double bond position correction	-0.1466
O—(C)(H)	340.1722	hydroxyl position correction	-4.1845
O—(C) ₂	-102.7000	<i>trans</i> - or <i>cis</i> -structure correction ^③	0.7583
O—(CO)(CH ₃)	-0.7253	carbonyl position correction	-5.9976
O—(CO)(CH ₂)	8.0117	phenol position correction	0.7988
O—(CO)(CH)	38.3230	exp(1/N)	17.3615
O—(CO)(O)	298.2085	Cb-(N) benzenamine correction	76.1923
T_{b0}	4775.5907	exp(1/ M_w)	-4491.3432

Note: The formalism of Benson groups is described in Ref. [1]. The first symbol represents the element that forms the centre of the group. The symbols between parentheses represent the elements to which it is linked. Usual symbols are used to represent the elements in their normal valence state. Elements in other valence states are distinguished using additional characters; furthermore, different symbols represent multiple bonded carbons, depending on the element at the other end of the multiple bond: Cd, carbon forming a double bond with another carbon; Cb, carbon involved in a benzene or a pyridine ring; CO, C O group; CN, C $\equiv$ N group; NI, nitrogen of the imide (C N) function. Also used for the nitrogen of pyridine derivatives. The pyridine ring is considered as formed of five Cb and one NI. For instance, pyridine will be decomposed into: one NI-(Cb)₂, five Cb-(H). Unlike Benson, who included contributions for gauche interactions to ideal gas phase properties, no contribution for the gauche interaction to normal boiling points and vaporization enthalpies were considered in this work.

① *Ortho*- and *meta*-corrections consider interactions between alkyl chains through a benzene ring.

② Corrections for pyridines: *Co*, *Cm*, and *Cp* pyridine corrections take into account alkyl ligands in position *ortho*, *meta*, and *para* with respect to the N element, respectively.

③ *trans*- or *cis*-correction: *cis*-structure correction is 1 and *trans*-structure is -1.

Table 3 Comparison of T_b predicted with our method and with the methods of Joback and Constantinou

Chemical family	No. of compounds	Joback		Constantinou		This study	
		$\bar{\epsilon}$ /K	$\bar{\delta}$ /%	$\bar{\epsilon}$ /K	$\bar{\delta}$ /%	$\bar{\epsilon}$ /K	$\bar{\delta}$ /%
alkanes	112	8.9	2.5	5.3	1.6	4.2	1.1
cycloalkanes	28	8.9	2.5	6.7	1.8	3.2	0.8
alkenes	35	9.5	2.7	4.8	1.4	1.1	0.3
aromatics	25	7.2	1.6	5.3	1.2	3.6	0.8
ketones and aldehydes	30	8.8	2.1	12.2	2.9	6.0	1.5
alcohols	33	13.8	3.4	16.2	3.9	5.1	1.2
acids	12	11.5	2.3	2.2	0.5	10.6	2.3
phenols and ether oxides	22	17.2	4.1	10.3	2.7	4.4	1.0
esters	26	18.7	4.0	7.3	2.0	5.9	1.6
bromo and chloroalkanes	19	18.0	4.9	6.0	2.0	3.3	1.0
amines and anilines	27	24.8	6.6	62.8	14.1	3.0	0.9
nitriles	9	18.1	4.5	34.7	9.0	9.4	2.2
pyridines	10	9.3	2.2	15.1	3.5	1.6	0.4
alkane thiols	17	16.2	4.3	46.4	13.2	3.5	0.9
thio ethers	12	8.6	2.5	15.8	4.4	1.8	0.5
overall	417	12.3	3.2	12.1	3.1	4.2	1.0

deviation of about 4.2 K, less than three times of the well known first order method of Joback and Reid and more than three times of the more precise second order method of Constantinou and Gani.

This method also can be applied for estimation of critical properties of organics effectively [18–20].

5 CONCLUSIONS

It is proposed that a quantitative relationship for normal boiling point based on position group contribution method and that position distribution function has been developed, which could distinguish most isomers including *cis*- or *trans*-structure of organic compounds for their thermodynamics properties. The estimated results for 417 different types of organics show that the position group contribution method is more precise than an existing method that uses first or second order groups.

REFERENCES

- Sanghvi, R., Yalkowsky, S.H., "Estimation of normal boiling point of organic compounds", *Ind. Eng. Chem. Res.*, **45**, 2856–2861 (2006).
- Tadeusz, H., "Prediction of thermodynamic properties of the systems formed by *n*-alkanes, aliphatic monoethers, and *i*-chloroalkanes using a cell-hole group contribution model", *J. Phy. Chem. B*, **108**, 2383–2397 (2004).
- Rebelo, L.P., Canongia, J.N., Esperanca, J.M., Filipe, E., "On the critical temperature normal boiling point and vapor pressure of ionic liquids", *J. Phy. Chem. B*, **109**, 6040–6043 (2005).
- Valderrama, J.D., Robles, P.A., "Critical properties, normal boiling temperature and acentric factors of fifty ionic liquids", *Ind. Eng. Chem. Res.*, **46**, 1338–1344 (2007).
- Dearden, J.C., "Quantitative structure-property relationships for prediction of boiling point, vapor pressure and melting point", *Environ. Toxicol. Chem.*, **22**, 1696 (2003).
- Joback, K.G., Reid, R., "Estimation of pure-component properties from group-contributions", *Chem. Eng. Commun.*, **57**, 233–243 (1987).
- Davotta, S., Pendyala, V.R., "Modified Joback group contribution method for normal boiling point of aliphatic halogenated compounds", *Ind. Eng. Chem. Res.*, **31**, 2042–2046 (1992).
- Constantinou, L., Pricken, S.E., Mavrouniotis, M.L., "Estimation of thermodynamic and physical properties of acyclic hydrocarbons using the ABC approach and conjugation operators", *Ind. Eng. Chem. Res.*, **32**, 1734–1742 (1993).
- Constantinou, L., Gani, R., "A new group contribution method for estimation of properties of pure compounds", *AIChE J.*, **40**, 1697–1710 (1994).
- Marrero-Morejon, J., Pardillo-Fontdevila, E., "Estimation of pure compound properties using group-interaction contributions", *AIChE J.*, **45**, 615–621 (1999).
- Krzyzaniak, J.F., Myrdal, P., Simamora, P., Yalkowsky, S.H., "Boiling point and melting point prediction for aliphatic nonhydrogen bonding compounds", *Ind. Eng. Chem. Res.*, **34**, 2530–2535 (1995).
- Yalkowsky, S.H., Danmenfelser, P., Myrdal, P., Simamora, P., "Unified physical property estimation relationships", *Chemosphere*, **28**, 1657 (1994).
- Zhao, L., Yalkowsky, S.H., "A combined group contribution and molecular geometry approach for predicting melting point of aliphatic compounds", *Ind. Eng. Chem. Res.*, **38**, 3581 (1999).
- Poling, E.B., Prausnitz, J.M., O'Connell, P.J., "The Properties of Gases and Liquids, 5th ed., McGraw-Hill, Inc., New York (2001).
- Frenkel, M., Gadalla, N.M., Hall, K.R., Hong, X., Marsh, K.N., Wilhoit, R.C., TRC Thermodynamic Tables: Hydrocarbon; Non-Hydrocarbon, Thermodynamic Research Center, The Texas A&M University System (1997).
- Ma, P.S., Handbook of Property Data of Organic Compound, Chemical Industry Press, Beijing (2006).
- Benson, S.W., Thermochemical Kinetics, 2nd ed., John Wiley and Sons, New York (1976).
- Wang, Q., Ma, P.S., "Position group contribution method for the prediction of critical temperatures of organic compounds", *J. Chem. Eng. Data*, **53**, 1103–1109 (2008).
- Wang, Q., Ma, P.S., "Position group contribution method for the prediction of critical pressure of organic compounds", *J. Chem. Eng. Data*, **53**, 1877–1885 (2008).
- Jia, Q.Z., Wang, Q., Ma, P.S., "Position group contribution method for the prediction of critical volume of organic compounds", *J. Chem. Eng. Data*, **53**, 2606–2612 (2008).