

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/231376432>

Prediction of Flash Point of Organosilicon Compounds Using Quantitative Structure Property Relationship Approach

ARTICLE *in* INDUSTRIAL & ENGINEERING CHEMISTRY RESEARCH · NOVEMBER 2010

Impact Factor: 2.59 · DOI: 10.1021/ie101381b

CITATIONS

11

READS

20

3 AUTHORS, INCLUDING:



Chan-Cheng Chen

National Kaohsiung First University of Scien...

28 PUBLICATIONS 297 CITATIONS

SEE PROFILE



Horng-Jang Liaw

National Kaohsiung First University of Scien...

40 PUBLICATIONS 563 CITATIONS

SEE PROFILE

Prediction of Flash Points of Organosilicon Compounds by Structure Group Contribution Approach

Gow-Bin Wang,[†] Chan-Cheng Chen,^{*,‡} Horng-Jang Liaw,[§] and Yi-Jen Tsai[§]

[†]Department of Chemical and Materials Engineering, Chang Gung University, 259 Wen-Hwa First Road, Kwei-Shan, Tao-Yuan County 33302, Taiwan, ROC

[‡]Department of Safety, Health and Environmental Engineering, National Kaohsiung First University of Science and Technology, 2 Jhuoyue Rd., Nanzih, Kaohsiung City 811, Taiwan, ROC

[§]Department of Occupational Safety and Health, China Medical University, 91 Hsueh-Shih Road, Taichung 40402, Taiwan, ROC

 Supporting Information

ABSTRACT: Flash point (FP) is the primary property to evaluate fire hazards of a flammable liquid. In most countries regulations for safe handling, transporting, and storage of liquid chemicals mainly depend on the FPs of liquid chemicals. Due to the advancement of technology in discovery or synthesis of new compounds, FP data are desirable for related industries, but there is often a significant gap between the demand for such data and their availability. Thus, a reliable method to predict the FPs of flammable compounds seems very important in this regard. In the present work a predictive model of FP for organosilicon compounds is proposed via the structure group contribution (SGC) approach. This model is built up by using a training set of 184 organosilicon compounds with the fitting ability (R^2) of 0.9330, the average error of 8.91 K, and the average error in percentage of 2.84%. The predictive capability of the proposed model has been demonstrated on a testing set of 46 organosilicon compounds with the predictive capability (Q^2) of 0.8868, the average error of 11.15 K, and the average error in percentage of 3.66%. Because the known error for measuring FP by experiment is reported to be about 6–10 K, the proposed method offers a reasonable estimate of the FP for organosilicon compounds. Moreover, the proposed SGC model requires only the molecular structure of a compound to estimate its FP, so it also offers an effective way to approximate the FP of a novel chemical for which its quantity is still not readily available for measuring its FP by experiments.

1. INTRODUCTION

Flash point (FP) is the temperature at which a liquid chemical emits sufficient flammable vapor to bring the concentration of the flammable vapor in the head space above the liquid to the lower flammability limit. Thus, FP is usually adopted to classify liquid chemicals for the purpose of assessing their fire and explosion hazards. In fact, in most countries regulations for the safe handling, transporting, and storage of liquid chemicals mainly depend on this classification, so the FP of liquid chemicals is information of great importance for the sake of safety practices. For example, a liquid chemical of which the FP is below 60.5 °C is defined to be a flammable liquid by the U.S. Department of Transportation (DOT 49 CFR 173.120), and special protective procedures should be implemented while storing or transporting it. Although the data of FP are desirable, due to the advancement of technology in the discovery or synthesis of new compounds, there is often a significant gap between the demand for such data and their availability. Moreover, for some toxic or radioactive chemicals, determining their FPs by experiments might be extremely difficult in practice. In this regard a reliable method to estimate the FP is still indispensable.

Many correlations have been proposed in the literature to predict FPs of pure chemicals. The work by Liu and Liu¹ could be referred to for a recent review. Roughly speaking, these correlations could be divided into three categories: the first category contains those correlations which require other physical

properties such as boiling point, density, enthalpy of vaporization, and so on.^{2–6} Correlations in this category have some drawbacks. First, because these correlations require other physical properties, their accuracies depend on the accuracies of the required properties or methods used to predict these properties. Furthermore, if any of these properties is not available for a specific compound, these correlations do not work in this case. Therefore, these correlations do not apply in predicting the FP of a novel substance for which most of its properties are often unavailable.

The second category is known as the quantitative structure–property relationship (QSPR) method. In this category, many molecular-based parameters, which are often called “molecular descriptors” in the literature, are used to estimate the FP of a compound. In this category the required molecular descriptors are calculated from the molecular structure of a molecule. Because this approach does not require any physical property, it can be applied to predict the FP of a novel substance. Recently many researchers have adopted this approach.^{7–13} The disadvantage of the QSPR approach is that calculation of the required molecular descriptors is usually very tedious to users. Unless a

Received: May 25, 2011

Accepted: September 26, 2011

Revised: September 13, 2011

Published: September 26, 2011

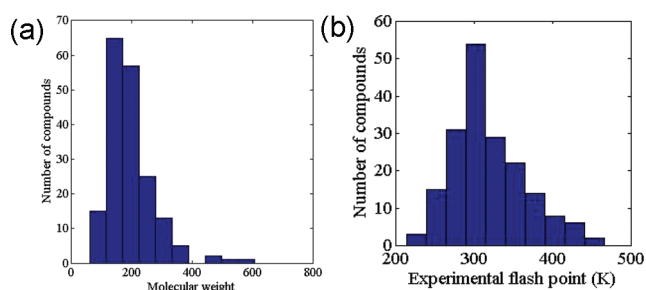


Figure 1. Distributions of molecular weight and flash point of the training set: (a) distribution of molecular weight; (b) distribution of the flash point.

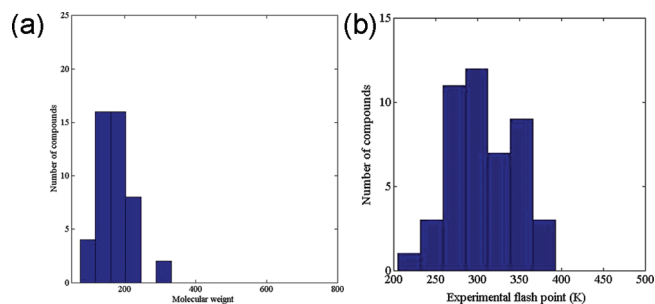


Figure 2. Distributions of molecular weight and flash point of the testing set: (a) distribution of molecular weight; (b) distribution of the flash point.

specialized software for calculating molecular descriptors is available by the user, it seems almost impossible to calculate the required molecular descriptors by users themselves.

The third category contains the well-known structure group contribution (SGC) correlations.^{14–16} In this category, a group contribution table of different functional groups which are deemed to contribute to the FP of a compound is first constructed, and then a correlation between the FP and the counts of these contributed groups is established. This correlation is then used to predict the FP of a new compound according to its counts of the contributed groups. Albahri had proposed a SGC model to predict FP for a data set of 287 pure hydrocarbons with 35 contributed groups, which gave the fitting ability (R^2) of 0.99 and maximum fitting error in percentage of 13.5%.¹⁴ Pan et al. had developed an artificial neural network (ANN) model to predict the FP of alkane with nine groups.¹⁵ There are 92 alkanes in their work, which are divided into a training set of 62 alkanes, a validation set of 15 alkanes, and a testing set of 15 alkanes. Gharagheizi et al. also proposed an ANN model to predict the FP for a data set of 1378 pure components, including heteroatoms such as nitrogen, sulfide, and so on.¹⁶ They divided their data into a training set of 1241 compounds and a testing set of 137 compounds. No validation set was considered in their work. Finally, 79 contributed groups are suggested, and the R^2 and Q^2 are reported to be of 0.9767 and 0.9661, respectively.

Although many studies of predicting FP of organic compounds have been explored in the literature, there are, however, few studies regarding predicting FPs of organosilicon compounds.^{2,12} Hsieh had correlated FPs of organosilicon compounds with their normal boiling points, and the following

correlation had been proposed in his work:²

$$T_f = -51.2385 + 0.4994T_b + 0.00047T_b^2 \quad (1)$$

where T_f and T_b are FP and normal boiling point with the unit of °C, respectively. He reached this conclusion from a data set of 207 organosilicon compounds, and the R^2 value of this correlation was reported to be 0.967. As eq 1 shows, this correlation requires the normal boiling point of a compound to predict its FP; thus the accuracy of the predicted FP highly depends on the accuracy of the measured normal boiling point. Furthermore, while a normal boiling point is unavailable for a specific compound, this correlation does not apply in this situation. Recently, Chen et al. also proposed predictive models of FP for a data set of 230 organosilicon compounds.¹² Two linear models, which are six-descriptor and 13-descriptor, respectively, are suggested in their works. Although these two models give reasonable performances, the QSPR approach seems not friendly to most users because calculation of the required molecular descriptors usually needs specialized software.

Although the applicability range of SGC models are limited by the studied data set (the new chemicals with functional groups not included in the explored models will be out of the model applicability range) and they usually provide a weak ability in distinguishing the isomeric compounds, they do not depend on the existing properties and are friendly to most practical engineers. Moreover, the semiconductor related industries have made great progress in recent years, so a predictive model of FP for organosilicon compounds via the SGC approach is still indispensable.

2. DATA SET, GROUP DEFINITIONS, AND METHODOLOGY

The FPs for 230 organosilicon compounds were collected from Hsieh's work² and The Hazardous Chemical Database.¹⁷ The collected organosilicon compounds showed wide variability in the distributions of both FP and molecular weight. The FP ranges from 204 to 466 K, and molecular weight ranges from 60.2 to 607.4. These compounds are then randomly distributed into a training set of 184 compounds and a testing set of 46 compounds. The distributions of molecular weight and FP for the training set in Figure 1, and Figure 2 shows these two distributions for the case of the testing set. It could be seen from these two figures that the distribution of molecular weight for the training set and for the testing set are quite similar; however, the compounds with FP higher than 400 K are all distributed into the training set. As the number of such compounds is very little, we think this will not distort the conclusion.

The contributed groups for predicting the FP of organosilicon compounds are listed in Table 1. To more clearly elucidate these groups, we first briefly discuss about the relation between the FP of a liquid chemical and its molecular structure. As FP is the temperature at which a liquid chemical emits sufficient flammable vapor to bring the concentration of the flammable vapor to the lower flammability limit, it is obvious that the FP of a liquid chemical highly depends on its ability to emit vapor, which is characterized by its normal boiling point. At the molecular level, the normal boiling point of a compound depends on intermolecular forces. For ionic molecules, ion–ion force is the main intermolecular force. While considering the nonpolar molecules, van der Waals force is the main attribution to the intermolecular force. The factors determining the van der Waals force of a

Table 1. Group Definitions for Predicting FP of Organosilicon Compounds

group no.	group	remark	group no.	group	remark
1	-CH ₃		21	-NH ₂	
2	>CH ₂		22	>NH	
3	>CH-		23	>NH	ring increments
4	>C<		24	>N-	
5	=CH ₂		25	-N=	ring increments
6	=CH-		26	-CN	
7	=C<		27	-S-	
8	≡CH		28	-SiH ₃	
9	≡C-		29	>SiH ₂	
10	-CH ₂ -	ring increments	30	>SiH-	
11	>CH-	ring increments	31	>Si<	
12	=CH-	ring increments	32	>SiH-	ring increments
13	=C<	ring increments	33	>Si<	ring increments
14	-F	attached to C	34	>N-	ring increments
15	-Cl	attached to C	35	-Cl	attached to Si
16	-Br	attached to C	36	-I	attached to Si
17	-OH	alcohol	37	-N=C=S	
18	-O-		38	-N=C=O	
19	-O-	ring increments	39	-N=C=N-	
20	>C=O				

Table 2. Pauling Electronegativities of Some Elements

element	electro-negativity	element	electro-negativity	element	electro-negativity
C	2.5	N	3.0	Cl	3.0
H	2.2	S	2.5	Br	2.8
O	3.5	F	4.0	Si	1.8

molecule include the following: molecular weight, molecular size, and the relative polarizability of electrons of the atoms involved. However, most organic molecules are neither ionic nor nonpolar but have permanent dipoles resulting from a nonuniform distribution of the bonding electrons. The dipole of a bond is mainly determined by the difference in electronegativity between atoms joined by this bond.¹⁸

Now we discuss the contributed groups defined in Table 1. Table 1 is basically modified from the contributed groups defined in Joback and Reid's work,¹⁹ in which contributed groups were designed to predict the normal boiling point of organic compounds. As mentioned earlier, the FP of a compound is found to be highly related to its normal boiling point, so contributed groups defined in their work are adopted to predict FP here. Groups 1–27 are those directly adopted from their work. Groups 28–39 are newly added contributed groups in the present work. These newly added groups are discussed as follows.

Table 2, which is condensed from Table 1.2 in ref 18, shows the Pauling's electronegativities of some atoms. As it is shown by Table 2, the electronegativity of carbon is higher than that of hydrogen but the electronegativity of silicon is less than that of hydrogen. Thus, bonding electrons are closer to the carbon atom in a carbon–hydrogen bond, but they are closer to the hydrogen atom in a silicon–hydrogen bond. Groups 28–33 are proposed to distinguish the difference in electronegativity between carbon

Table 3. Group Contributions in Different Models

serial	group	remark	MLR	degree 2	degree 3	degree 4
1	-CH ₃		-5.8949	-4.7543	-4.0297	-4.0298
2	>CH ₂		11.6851	15.6399	13.7287	13.7287
3	>CH-		24.9147	29.7614	26.2355	26.2356
4	>C<		34.8034	41.8024	36.535	36.535
5	=CH ₂		-2.5988	-0.9360	-0.6978	-0.6978
6	=CH-		8.7628	11.8316	10.4606	10.4606
7	=C<		26.2462	28.2627	24.5365	24.5364
8	≡CH		17.0872	25.8011	23.2633	23.2633
9	≡C-		9.7398	10.14026	9.0672	9.0672
10	-CH ₂ -	ring increments	9.4772	12.0741	10.6532	10.6532
11	>CH-	ring increments	17.0767	26.3594	23.4316	23.4317
12	=CH-	ring increments	8.5096	11.6540	10.1581	10.1581
13	=C<	ring increments	23.1199	28.2310	24.549	24.549
14	-F	attached to C	-15.5806	-17.3884	-15.0968	-15.0968
15	-Cl	attached to C	14.715	21.5543	19.2371	19.2371
16	-Br	attached to C	18.9825	29.2475	25.8415	25.8415
17	-OH	alcohol	40.3115	49.4208	43.6808	43.6808
18	-O-		6.8130	8.7550	7.4483	7.4482
19	-O-	ring increments	21.1070	17.4888	14.8537	14.8537
20	>C=O		29.8577	36.9982	32.5291	32.5291
21	-NH ₂		21.7836	26.7186	22.411	22.4109
22	>NH		13.6770	17.0723	15.2071	15.207
23	>NH	ring increments	23.7222	20.6411	17.7693	17.7693
24	>N-		19.2356	23.0458	19.8475	19.8475
25	-N=	ring increments	30.9789	58.1676	73.1871	73.1046
26	-CN		36.1882	46.3254	40.4311	40.4311
27	-S-		30.7360	38.2233	34.0047	34.0047
28	-SiH ₃		0.1893	5.1962	3.6484	3.6483
29	>SiH ₂		-2.8718	7.0559	5.0994	5.0996
30	>SiH-		20.2789	22.9298	20.1597	20.1597
31	>Si<		31.2739	35.2149	30.5686	30.5686
32	>SiH-	ring increments	0.5165	8.7554	8.1994	8.1994
33	>Si<	ring increments	12.5462	21.0877	18.5585	18.5585
34	>N-	ring increments	-22.5224	-19.7121	-16.284	-16.2837
35	-Cl	attached to Si	5.6747	9.4766	8.3391	8.3391
36	-I	attached to Si	33.3250	46.5773	41.2847	41.2847
37	-N=C=S		53.3250	69.7229	61.7407	61.7407
38	-N=C=O		12.660	18.3904	16.3327	16.3327
39	-N=C=N-		46.736	57.1338	50.5372	50.5372

Table 4. Estimates of Model Structure Parameters in Different Models

coefficient	MLR	degree 2	degree 3	degree 4
<i>a</i>	2.4124×10^2	2.2454×10^2	2.2842×10^2	2.2842×10^2
<i>b</i>		-8.5904×10^{-4}	1.8434×10^{-4}	1.8437×10^{-4}
<i>c</i>			-3.2489×10^{-6}	-3.2487×10^{-6}
<i>d</i>				7.5885×10^{-13}

atom and silicon atom. Group 34 is a newly added group for nitrogen atom in a ring structure which is not defined in Joback and Reid's work. It could be also seen from Table 2 that the electronegativity difference between halogen atom and carbon

Table 5. 95% Confidence Intervals for Model Structure Parameters in Different Models

models	a_-	a_+	b_-	b_+	c_-	c_+	d_-	d_+
MLR	2.1399×10^2	2.6848×10^2						
degree 2	1.9090×10^2	2.5817×10^2	-1.1294×10^{-3}	-5.8867×10^{-4}				
degree 3	1.9747×10^2	2.5937×10^2	-3.0546×10^{-3}	3.4232×10^{-3}	-1.4166×10^{-5}	7.6686×10^{-6}		
degree 4	1.9490×10^2	2.6194×10^2	-1.0907×10^{-2}	1.1276×10^{-2}	-6.8982×10^{-5}	6.2484×10^{-5}	-1.0583×10^{-7}	1.0583×10^{-7}

Table 6. Performances Comparison for Models with Different Complexity

models	training Set	testing Set		
	R^2	Q^2	max err (%)	av err (%)
MLR	0.9278	0.6957	31.32	4.60
degree 2	0.9330	0.7740	25.58	4.13
degree 3	0.9334	-3.0217	84.19	19.46
degree 4	0.9334	0.0081	35.25	9.90

atom is much larger than the one between halogen atom and silicon atom, so groups 35 and 36 are introduced to differentiate this effect. Groups 37–39 are considered to address the effect of highly symmetric structure in a molecule.

With regard to the model structure, the simplest form should be the multiple linear regression (MLR) model; that is,

$$FP = f_0 + \sum_{i=1}^n v_i f_i \quad (2)$$

where n is the number of groups defined in Table 1, v_i is the number of group i in a molecule, f_i ($i = 1, \dots, n$) is the group contribution for the i th defined group, and f_0 is the intersect of the fitting line. For more accurate estimation, other nonlinear models could be considered. For example, ANN model is one common choice for this purpose. However, ANN model usually requires too many parameters in a model, and this makes the model too complicated to be used. For example, a three-layer ANN model with 39 neurons in the input layer, six neurons in a hidden layer, and one neuron in an output layer will need $(39 \times 6 + 6) + (6 \times 1 + 1) = 247$ parameters. Because the number of available data in the present work is only 230, the ANN model seems impractical in this situation. A nonlinear model of polynomial form had also been suggested to predict FPs of organic compounds in the literature¹⁴ and discussion on the feasibility of such a polynomial model had been also reported.²⁰ Because the nonlinear model of the polynomial form, i.e., eq 3, had been proven to be effective in predicting FP, it is adopted in the present work.

$$FP = a + \left(\sum_i v_i f_i\right) + b\left(\sum_i v_i f_i\right)^2 + c\left(\sum_i v_i f_i\right)^3 + d\left(\sum_i v_i f_i\right)^4 \quad (3)$$

In eq 3, v_i is the number of group i in a molecule, f_i is the group contribution for the i th contributed group, and parameters a , b , c , and d are the model structure parameters. Parameters in eq 3 are usually solved by the least-squares error criterion, which is a nonlinear regression problem. Many algorithms could be considered for this problem, but solutions may differ slightly by distinct algorithms. However, this difference usually has no

practical significance. In the present work the linearized algorithm is adopted because it can simultaneously estimate the required parameters and their corresponding confidence interval. Discussion on this linearized algorithm could be found in the literature.²⁰

To evaluate the feasibility of a model, we must assess its performances both in goodness of fit ability and in goodness of predictive capability. In this study, the quantitative measure of goodness of fit ability is demonstrated by the explained variation in the training set (R^2), whereas the goodness of predictive capability is given by the predicted variation in the testing set (Q^2). These two indices are defined as follows.

$$R^2 = 1 - \frac{\sum_{i=1}^K (y_i - \hat{y}_i)^2}{\sum_{i=1}^K (y_i - \bar{y})^2} \quad \text{for the training set}$$

$$Q^2 = 1 - \frac{\sum_{i=1}^M (y_i - \hat{y}_i)^2}{\sum_{i=1}^M (y_i - \bar{y})^2} \quad \text{for the testing set}$$

where y_i is the i th sample measurement, \hat{y}_i is the predicted value of the i th sample, \bar{y} is the average of all sample measurements in the specific data set, and K and M are the numbers of compounds in the training set and testing set, respectively. Usually, R^2 and Q^2 behave differently with increasing model complexity (i.e., number of parameters in a model). The R^2 is inflationary and approaches unity as the model complexity increases. Hence, it is not sufficient to only have a high R^2 for a practical model. The Q^2 , on the other hand, is not inflationary and will not automatically approach unity with increasing model complexity. Commencing with a very simple model, Q^2 will increase as model complexity increases. However, at a certain degree of complexity, Q^2 will reach a plateau and subsequently reduce. Usually, the point at which Q^2 reaches a plateau is the trade-off point between the fit ability and predictive capability. An alternative to decide the model complexity is to conduct hypothesis testing for a specific parameter in the model to see whether or not it is significant in the corresponding model. In the present work, both Q^2 statistics and hypothesis testing will be considered to decide the feasible model complexity.

3. RESULTS AND DISCUSSION

Polynomial models of different degrees, including MLR and polynomial models of degrees 2–4, were explored in this work. For a model of given degree, the group contributions (f_i) of different contributed groups and other model structure parameters (i.e., a , b , c , and d in eq 3) were solved to minimize the sum of squared errors for the training set. Parameters in the MLR

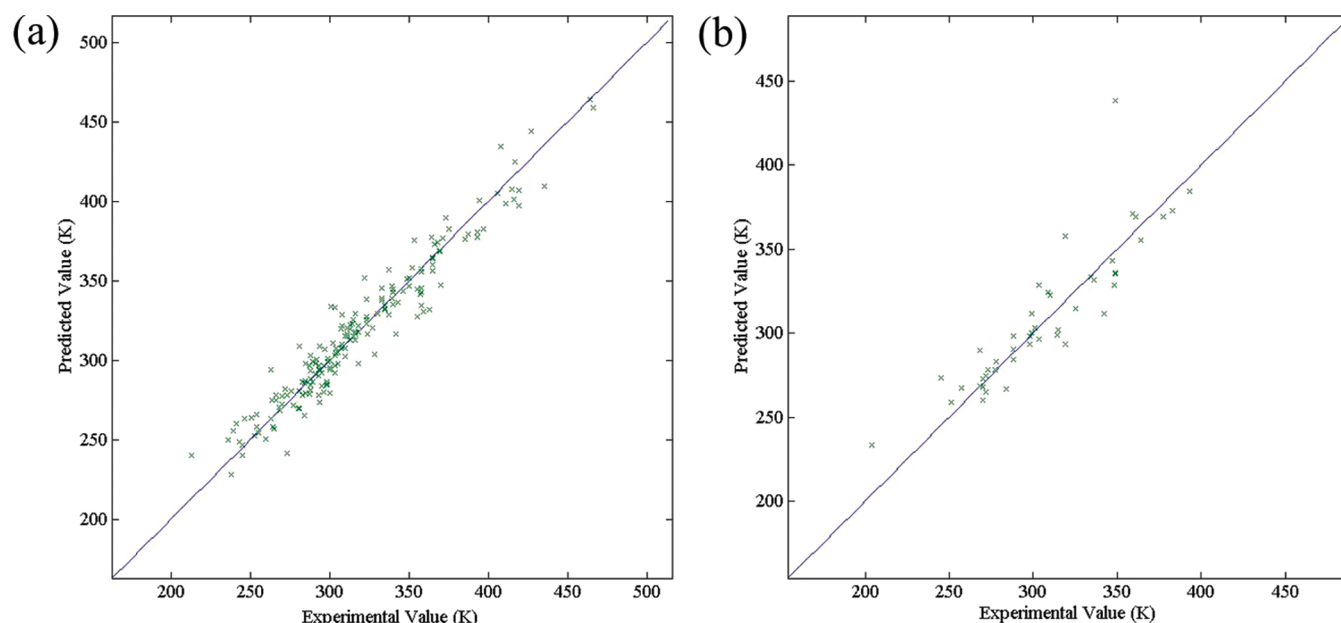


Figure 3. Performances of the polynomial model of degree 2: (a) training data; (b) testing data.

model were directly calculated by the classical least-squares method. For nonlinear models, the aforementioned linearized algorithm was used to solve all parameters. Table 3 summarizes numerical values of group contributions in different models for all contributed groups defined in Table 1. The model structure parameters are listed in Table 4.

Table 5 lists the 95% confidence intervals for model structure parameters a , b , c , and d in all investigated models. For the polynomial model of degree 3, the 95% confidence intervals for parameter b and for parameter c both contain the value of zero, which means that neither the hypothesis $b = 0$ nor the hypothesis $c = 0$ could be rejected with 95% confidence. Thus, the polynomial model of degree 3 is an overfitting model. With similar arguments, it is also found that the polynomial model of degree 4 is an overfitting model, too. For the polynomial model of degree 2, although the estimate of b is very small (i.e., -8.5904×10^{-4}), its 95% confidence interval does not include a value of zero, so the hypothesis $b = 0$ is rejected with 95% confidence. Thus, it is concluded from these hypothesis tests that the adequate model is the one of degree 2.

Table 6 shows R^2 , Q^2 , the maximum predictive error in percentage, and the average predictive error in percentage for all explored models. It could be seen from Table 6 that although the R^2 value increases as the model degree increases, the Q^2 value, however, will reach the maximum while the model is of degree 2. Thus, the Q^2 -statistics approach also indicates that the polynomial models of degree 3 and of degree 4 are overfitting models. So, the polynomial model of degree 2 is adopted in the present work. As it is shown by Table 6, this model gives the fitting ability of $R^2 = 0.9330$ and predictive capability of $Q^2 = 0.7740$, respectively. A rule of thumb for developing a practical model is as follows: the difference between R^2 and Q^2 must not be too large and preferably not exceeding 0.2–0.3. Moreover, a $Q^2 > 0.5$ is regarded as good and a $Q^2 > 0.9$ as excellent.²¹ Thus, this model offers a reasonable estimate of FP for organosilicon compounds. The predicted FPs by this polynomial model of degree 2, experimental FPs, and group counts pertaining to all defined

contributed groups are supplied in the Supporting Information for all collected organosilicon compounds.

The predicted FPs were plotted against their experimental values in Figure 3. Figure 3a shows the case of the training set, and Figure 3b shows the case of the testing set. There is, as shown in Figure 3b, an obvious outlier in the testing set. This outlier is found to be of triphenylsilane. Its FP is reported to be 349.15 K (76 °C) in The Chemical Database, but the proposed method predicts its FP to be of 438.44 K (165.3 °C). Recently, Chen et al. had pointed out the FP of triphenylsilane reported in The Chemical Database is a mistake. They had measured its FP according to the ASTM D93 standard test method, and the newly measured FP had been reported to be of 143.2 °C.¹² Obviously, the proposed SGC method has also successfully pointed out this mistake. After correcting this outlier, the Q^2 value will now increase from 0.7740 to 0.8868.

A general comparison between the present work and those reported in the literature is as follow. Hsieh had correlated the FP of an organosilicon compound with its normal boiling point, and the correlation of eq 1 had been proposed in his work.¹ He reached this equation from a data set of 207 organosilicon compounds, and the R^2 value of his model was announced to be 0.967. Although the 207 organosilicon compounds in his work are all included in the present 230-compound data set, the R^2 value will decrease to 0.8667, while eq 1 is applied to the whole data set (after correcting the outlier of triphenylsilane). However, because the newly added organosilicon compounds are not included in building up eq 1, such a comparison might be a little unfair to his work. To make a more adequate comparison, we refit the relation between FP and normal boiling point with our training set, and the following correlation is obtained:

$$T_f = -45.7042 + 0.485939T_b + 0.000411T_b^2 \quad (4)$$

where units of the corresponding terms in eq 4 are the same as those in eq 1. This equation is then applied to the testing set to evaluate its predictive capability. The results for both the fitting ability and predictive capability are summarized in Table 7.

Table 7. Performance Comparison between the Proposed SGC Method and Other Works

model*	data set	MSE	R^2/Q^2	max err (K)	av err (K)	max err (%)	av err (%)
Hshieh (eq 1)	whole	280.38	0.8667	107.67	10.64	25.86	3.29
Hshieh (eq 4)	training	232.68	0.8922	88.39	10.52	24.21	3.26
Hshieh (eq 4)	testing	436.81	0.7650	105.55	11.96	25.35	3.81
Chen et al. (A)	training	154.27	0.9293	43.86	9.44	13.79	3.01
Chen et al. (A)	testing	134.26	0.9245	27.91	9.08	11.27	2.97
Chen et al. (B)	training	222.50	0.8981	46.84	11.21	14.72	3.58
Chen et al. (B)	testing	260.83	0.8533	45.26	12.17	15.71	4.05
this work	training	144.68	0.9330	32.83	8.91	12.78	2.84
this work	testing	206.94	0.8868	38.26	11.15	14.20	3.66

* Chen et al. (A), 13-descriptor model; Chen et al. (B), six-descriptor model.

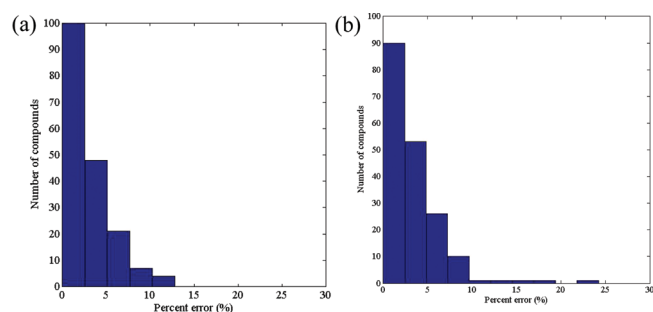


Figure 4. Distribution of percentage errors for the training set: (a) this work; (b) Hshieh's model.

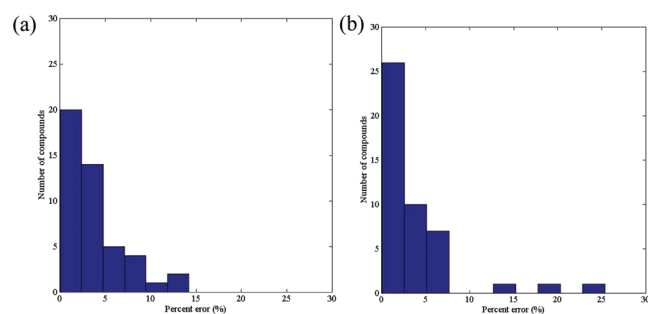


Figure 5. Distribution of percentage errors for the testing set: (a) this work; (b) Hshieh's model.

It could be seen from this table that although the R^2 value, after refitting, will increase from 0.8667 to 0.8922, there is still a gap between 0.8922 and the announced value of 0.967. Q^2 of this correlation is found to be of 0.7650. Other performance indices, which include mean square error (MSE), maximum error, average error, maximum error in percentage, and average error in percentage, are also listed in Table 7. It could be seen from Table 7 that the proposed method performs better than Hshieh's method in all these indices. The histograms of fitting errors in percentage for Hshieh's model and for the proposed model are compared in Figure 4; the cases of predictive errors are compared in Figure 5. It could be seen from these two figures that for the proposed model the fitting errors and predictive errors are less than 15% for all investigated compounds, but for Hshieh's model both the fitting error and predictive error may be higher than 20% for some organosilicon compounds. Chen et al.¹² had also proposed a 13-descriptor model and a

six-descriptor model for the organosilicon compounds via the QSPR approach. The performances of these two models are also summarized in Table 7. It could be observed from Table 7 that the proposed SGC model is superior to their six-descriptor model and comparable to their 13-descriptor model for all of the explored performance indices. However, as we have mentioned, the QSPR model usually needs specialized software to calculate the required molecular descriptors; thus, the proposed SGC model seems more friendly to practical engineers.

4. CONCLUSIONS

In the present work a predictive model of FP for organosilicon compounds is proposed on the basis of the SGC approach. The proposed equation to predict the FPs for organosilicon compounds is

$$FP = 2.2454 \times 10^2 + \left(\sum_i v_{fi} \right) - 8.5904 \times 10^{-4} \left(\sum_i v_{fi} \right)^2 \quad (5)$$

The proposed SGC model includes 39 contributed groups and is a polynomial model of degree 2. This model was deduced from a training set of 184 organosilicon compounds. The fitting ability (R^2) of the proposed model is found to be of 0.9330, with an average error of 8.91 K and an average error in percentage of 2.84%. The predictive capability of this model is demonstrated on a testing set of 46 compounds which were not included in the original training set. The predictive capability (Q^2) of the proposed model is of 0.8868, with an average error of 11.15 K and an average error in percentage of 3.66%.

As compared with Hshieh's model, the proposed model exhibits much better performances. While comparing with Chen et al.'s work, the proposed model is superior to their six-descriptor model and is comparable to their 13-descriptor model. However, their models need specialized software to calculate the required descriptors, thus, the proposed SGC model seems more friendly to practical engineers. The proposed SGC model has successfully indicated the FP of triphenylsilane reported in The Chemical Database is a mistake. Moreover, the average error in measuring FP by experiment is reported to be 6–10 K in ASTM D93 method, so the proposed SGC method, which gives an average fitting error of 8.91 K and an average predictive error of 11.15 K, obviously offers a reasonable estimate of the FP of an organosilicon compound while its experimental value is unavailable.

■ ASSOCIATED CONTENT

S Supporting Information. Tables listing names of 230 organosilicon compounds used in this study, their experimental FP, the predicted FPs given by proposed methods, and their group counts for contributed groups. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*Tel.: 886-7-6011000 ext. 2311. Fax: 886-7-6011061. E-mail: chch_chen@ccms.nkfust.edu.tw.

■ ACKNOWLEDGMENT

We thank the National Science Council of the ROC for supporting this study financially under Grant No. NSC 98-2221-E-039-002-MY3.

■ REFERENCES

- (1) Liu, X.; Liu, Z. Research progress on flash point prediction. *J. Chem. Eng. Data* **2010**, *55*, 2943–2950.
- (2) Hsieh, F. Y. Correlation of closed-cup flash points with normal boiling points for silicone and general organic compounds. *Fire Mater.* **1997**, *21*, 277–282.
- (3) Metcalfe, E.; Metcalfe, A. E. M. On the correlation of flash points. *Fire Mater.* **1992**, *16*, 153–154.
- (4) Patil, G. S. Estimation of flash point. *Fire Mater.* **1988**, *12*, 127–131.
- (5) Satyanarayana, K.; Kakati, M. C. Correlation of flash points. *Fire Mater.* **1991**, *15*, 97–100.
- (6) Satyanarayana, K.; Rao, P. G. Improved equation to estimate flash points of organic compounds. *J. Hazard. Mater.* **1992**, *32*, 81–85.
- (7) Katritzky, A. R.; Petrukhin, R.; Jain, R.; Karelson, M. QSPR analysis of flash points. *J. Chem. Inf. Comput. Sci.* **2001**, *411*, 521–1530.
- (8) Katritzky, A. R.; Stoyanova-Slavova, I. B.; Dobchev, D. A.; Karelson, M. QSPR modeling of flash points: An update. *J. Mol. Graphics Modell.* **2007**, *26*, 529–536.
- (9) Gharagheizia, F.; Alamdari, R. F. Prediction of flash point temperature of pure components using a quantitative structure–property relationship model. *QSAR Comb. Sci.* **2008**, *27* (6), 679–683.
- (10) Patel, S. J.; Ng, D.; Mannan, M. S. QSPR flash point prediction of solvents using topological indices for application in computer aided molecular design. *Ind. Eng. Chem. Res.* **2009**, *48*, 7378–7387.
- (11) Pan, Y.; Jiang, J.; Ding, X.; Wang, R.; Jiang, J. Prediction of flammability characteristics of pure hydrocarbons from molecular structures. *AIChE J.* **2010**, *56*, 690–701.
- (12) Chen, C. C.; Liaw, H. J.; Tsai, Y. J. Prediction of flash point of organosilicon compounds using quantitative structure property relationship approach. *Ind. Eng. Chem. Res.* **2010**, *49*, 12702–12708.
- (13) Suzuki, T.; Ohtaguchi, K.; Koide, K. A method for estimating flash point of organic compounds from molecular structure. *J. Chem. Eng. Jpn.* **1991**, *24* (2), 258–261.
- (14) Albahri, T. A. Flammability characteristics of pure hydrocarbons. *Chem. Eng. Sci.* **2003**, *58*, 3629–3641.
- (15) Pan, Y.; Jiang, J.; Wang, Z. Quantitative structure-property relationship studies for predicting flash points of alkanes using group bond contribution method with back-propagation neural network. *J. Hazard. Mater.* **2007**, *147*, 424–430.
- (16) Gharagheizi, F.; Alamdari, R. F.; Angaji, M. T. A new network-group contribution method for estimation of flash point temperature of pure components. *Energy Fuels* **2008**, *22*, 1628–1635.
- (17) The Chemical Database. The Department of Chemistry at the University Akron; <http://ull.chemistry.uakron.edu/erd/index.html>.
- (18) Solomons, T. W. G. *Fundamentals of Organic Chemistry*; John Wiley and Sons: New York, 1982.
- (19) Joback, K. G.; Reid, R. C. Estimation of pure-component properties from group-contributions. *Chem. Eng. Commun.* **1987**, *57*, 233–243.
- (20) Chen, C. C.; Liaw, H. J.; Kuo, Y. Y. Prediction of autoignition temperatures of organic compounds by the structural group contribution approach. *J. Hazard. Mater.* **2009**, *162*, 746–762.
- (21) Umetrics, A. B. Multi- and Megavariate Data Analysis; Umetrics AB: Umea, Sweden, 2001.