A METHOD FOR ESTIMATING FLASH POINTS OF ORGANIC COMPOUNDS FROM MOLECULAR STRUCTURES

TAKAHIRO SUZUKI, KAZUHISA OHTAGUCHI AND KOZO KOIDE
Department of Chemical Engineering, Tokyo Institute of Technology, Tokyo 152

Key Words: Flash Point, Principal Component Analysis, Molecular Connectivity Index

Introduction

The flash point provides a simple, convenient index of the flammability and combustibility of substances. Regulations for the safe handling, transportation, and storage of combustible substances are dependent on the classification according to their flash points. The flash point is therefore of great importance in the chemical industry. Experimental flash point data are desirable, but due to the rapid advance of technology in discovery or synthesis of new compounds there is a significant gap between demand for such data and their availability. Moreover, for some toxic, explosive, or radioactive compounds the experimental determination of the flash point is extremely difficult. Hence a reliable method of estimating the flash point is desired.

In our previous report[5], the principal-component analysis (*PCA*) of a set of data including flash points and ten other flammability-related physicochemical properties of 50 organic compounds was performed to explore the number of significant structural factors affecting the flash point. Just two structural factors were found to be sufficient to reproduce the flash-point data.

This study is designed to extend the above analysis to a broader data set for 100 compounds. The object is to interpret the second factor more clearly and to develop a practically applicable method for estimating the flash point.

* Received September 25, 1990. Correspondence concerning this article should be addressed to T. Suzuki.

1. Methods

Principal-component analysis (*PCA*)[6] is adopted here, as in our previous work[5] for revealing the intrinsic factors affecting the flash point. The data format for *PCA* is the same as in the earlier paper.[5] The data set includes 100 compounds (for which experimental closed-cup flash points are available), ranging in size from methanol to ethyl(*o*-(ethoxycarbonyl)benzoyloxyacetate) and in polarity from alkanes to multifunctional compounds. Selected properties are the flash point and 10 other flammability-related properties, normal boiling point, critical temperature, molecular weight, molar refractivity, liquid-state molar volume, enthalpy of vaporization at normal boiling point, parachor, molar magnetic susceptibility, and a and b constants of van der Waal's equation of state. This compound/property table contains $100 \times 11 = 1100$ values. Most of the values have been taken from the literature.[1,2] Unfortunately, about 25% of the data consist of missing values, all of which have been estimated and included in the data matrix. This data matrix is submitted to *PCA* using standard statistical procedures, which enables us to treat the structure-flash point problem as a multivariable one.

2. Results and Discussion

The application of *PCA* to the whole data set (including the estimated values) reveals that first two components are significant according to the cross-validation. These components are new variables created from linear combinations of the 11 starting variables. Each component is orthogonal (uncorrelated) with all the other principal components. The first principal component contains the largest part of the variance of the data set, with subsequent principal components containing correspondingly smaller amounts of variance. In this case, the first two components together describe 94.0% of the total variance in the original data set, where the first component accounts for 80.7% and the second for 13.3%. Due to the relatively small contribution of the third (accounting for only 2.4% of the variance) and subsequent components, we disregard them in the following discussion. Thus it can be determined that the intrinsic dimensionality of the initial 11 parameters (T_f, T_b, $\cdots$, b) is two. The results of the analysis are essentially unchanged for this more-extended data set. Then the flash point can be described as linear combinations of these two components.[5]

One interpretation of the principal components is that they are abstract representations of the molecular structural features. We tentatively interpreted two components (z_1 and z_2) as related to bulk (or size) and polarity-related property, respectively, by inspecting the correlation between the components scores of z_1 and z_2 and some physico-chemical parameters.[5] The model using two components is in the form shown in Eq. (1):

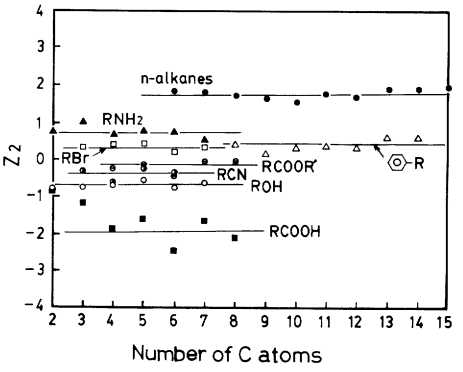

Fig. 1. z_2 for a number of homologous series of compounds

$$T_f \equiv \text{bulk} + \text{polarity-related factor} \quad (1)$$

The first term was modeled by using topological molecular descriptors called the first-order molecular connectivity index, ${}^1X$. The index is defined by Eq. (2) and can be easily calculated by considering the hydrogen-suppressed molecular skeleton[4]:

$${}^1X = \Sigma(\delta_i \delta_j)^{-1/2} \quad (2)$$

where the sum is over all bonds (the edges of the graph) in the molecule and δ_i is the number of atoms connected to it. Atoms i and j are formally bonded.

A plot of the first principal component score from this analysis against ${}^1X$ gives essentially the same results as our earlier report for 50 compounds.[5] The regression equation observed for these data was excellent, as shown in Eq. (3).

$$z_1 = 1.926\,{}^1X - 7.009 \quad (3)$$
$$(n = 100,\ r = 0.967,\ s = 0.764)$$

Figure 1 shows the second component (z_2) scores obtained for a number of homologous compounds in the data set. It suggests that the second factor is independent of molecular size and is related to the specific polar characteristics of the functional groups. A number of parameters have been used as a measure of polarity. Previously we investigated a correlation between the second factor and Fujita's inorganicity value[3] and a statistically significant correlation ($r = 0.898$) was observed for 30 compounds, but the correlation did not hold for all 50 compounds ($r = 0.540$).[5] Hence, in our current attempt we assume that the second factor can be correlated directly in terms of a structural additivity scheme.

From the above considerations, Eq. (4) for the flash point estimation is proposed:

$$T_f = c_1^1 X + \Sigma n_i G_i + c_0 \quad (4)$$

where T_f is in degrees Celsius, G_i is the atom and group contribution which occurs n_i times, and c_1 and c_0 are constants. To determine c_1, c_0 and G_i, the experimental closed-cup flash point values of 400 compounds were used, including 33 aliphatic hydrocarbons, 26 aromatic hydrocarbons, 30 alcohols, 10 phenols, 23 ethers, 10 aldehydes, 17 ketones, 11 acids, 38 esters, 8 nitriles, 31 amines and anilines, 4 N-containing ring compounds, 6 amides and anilides, 8 nitro compounds, 4 isocyanates, 10 thiols, 7 sulfides and thiophenes, 2 sulfones, 4 isothiocyanates, 35 halogenated compounds, and 83 multifunctional compounds. The following best-fit equation was found using linear regression analysis.

$$T_f = 25.57^1X + \Sigma n_i G_i - 86.0 \quad (5)$$
$$(n=400,\ r=0.967,\ s=13.52)$$

where G_i values determined are shown in **Table 1**.

Estimation results for the 400 compounds used as the original data set are summarized in **Table 2**. Average absolute and bias errors are shown for individual classes of compounds. The average absolute error of all available data was 10.3°C. For phenols the absolute error was about twice that for all compounds. This suggests the presence of some ad hoc intrinsic or unusual factors. One of the predominant features of this method is its wide applicability. Even for halogenated and multifunctional compounds, estimation results comparable with those for monofunctional compounds have been obtained.

Conclusion

A new method for estimating the flash point has been proposed. It requires only the structural formula of a molecule of interest and can be applied to diverse compounds. This method is useful not only for assessing the flammability of compounds that are as yet unknown or not readily available but also for preliminary selection of an optimal experimental condition for measuring the flash point.

Supplementary material available

All data used in *PCA* and a list of the experimental

Table 1. Atomic and group increments for Eq. (5)

Atom or Group	G_i	Atom or Group	G_i
C (aromatic)	3.1	–NH	20.4
–OH(alcohols)	55.8	>N–(aliphatic)	6.0
–OH(phenols)	31.4	N (aromatic)	15.9
–O–	2.8	F	−7.4
–CHO	21.4	Cl	21.7
>C=O	25.3	Br	46.4
>COOH	85.7	I	63.0
–COO–	13.3	–CONH–	112.7
–CN	52.8	–CON<	41.3
–SH	27.3	$-SO_2-$	102.3
–S–	30.8	–NCS	57.9
$-NO_2$	41.3	–NCO	21.1
$-NH_2$	32.5		

Table 2. Statistical evaluation of estimation results

Chemical Class	Number of Compounds	Flash Point Range (°C)	Estimated T_f (°C) Av. Error[a)]	Bias[b)]
Aliphatic hydrocarbons	33	−54.0–168.0	12.2	−1.1
Aromatic hydrocarbons	26	−11.0–140.6	6.1	2.0
Alcohols	30	11.0–110.0	6.8	1.0
Phenols	10	79.0–153.0	21.5	−11.8
Ethers	23	−45.0–135.0	9.1	1.4
Aldehydes	10	−27.0–63.0	7.3	1.4
Ketones	17	−20.0–77.0	8.1	−5.1
Acids	11	39.0–121.0	7.4	1.0
Esters	38	−20.0–148.0	5.8	0.3
Nitriles	8	0.0–71.0	10.5	10.5
Amines and Anilines	31	−29.0–153.0	14.5	6.5
N-containing ring compounds	4	20.0–99.0	7.0	0.0
Amides and Anilides	6	107.2–173.0	7.7	0.0
Nitro compounds	8	24.0–179.0	8.7	−8.7
Isocyanates	4	−18.0–55.0	4.6	2.8
Thiols	10	−20.0–87.0	7.7	0.0
Sulfides and Thiophenes	7	−36.0–76.0	12.9	1.3
Sulfones	2	126.0–143.0	4.3	0.0
Isothiocyanates	4	32.0–87.0	8.7	0.0
Halogenated hydrocarbons	35	−43.0–143.9	9.9	1.3
Multifunctional compounds	83	−15.0–196.0	14.2	−1.2
Total	400	−54.0–196.0	10.3	0.2

a) Av. Error $= \Sigma \mid T_f(\text{estd}) - T_f(\text{obsd}) \mid$ /number of compounds
b) Bias $= \Sigma (T_f(\text{estd}) - T_f(\text{obsd}))$/number of compounds

and estimated flash points for 400 compounds can be obtained from the authors on request.

Nomenclature

n	=	number of datam points in set	[—]
r	=	multiple correlation coefficient	[—]
s	=	standard deviation	[°C or —]
T_f	=	flash point	[°C]
z_i	=	principal component	[—]
δ	=	number assigned to each atom	[—]
${}^1X$	=	first-order molecular connectivity index	[—]

Literature Cited

1) Dean, J. A.: "Lange's Handbook of Chemistry," 13th ed, Mcgraw-Hill, New York (1985).
2) Dean, J. A.: "Handbook of Organic Chemistry," McGraw-Hill, New York (1987).
3) Fujita, A: *Pharm. Bull.* (*Tokyo*), **2**, 163 (1954).
4) Kier, L. B. and L. W. Hall: "Molecular Connectivity in Structure-Activity Analysis," Research Studies Press, Letchworth (1986).
5) Suzuki, T., K. Ohtaguchi and K. Koide: *Kagaku Kogaku Ronbunshu*, **16**, 1224 (1990).
6) Tou, J. T. and R. C. Gonzalez: "Pattern Recognition Principles," Addison-Wesley, Reading, MA (1974).