



Insight into the contribution of individual functional groups to the flash point of organic compounds



Didier Mathieu*, Thibaud Alaime

CEA, DAM, Le Ripault, F-37260 Monts, France

HIGHLIGHTS

- A remarkably simple approach to flash point prediction is put forward.
- It combines a power law expression with geometrical fragments.
- It proves as reliable as models based on advanced QSPR techniques.
- In addition, it provides insight into the contributions of functional groups.
- Training and test sets include respectively 287 and 1170 compounds.

ARTICLE INFO

Article history:

Received 5 August 2013
Received in revised form
19 December 2013
Accepted 21 December 2013
Available online 3 January 2014

Keywords:

Flash point
Modeling
Prediction methods
Additivity methods

ABSTRACT

Flash point temperatures of organic compounds are predicted on the basis of a power law involving 21 additive contributions associated with non-hydrogen atoms and ring structures. The model is parametrized against a previous data set of 287 simple organic molecules. An average absolute error of 8.6 K and a maximal error of about 50 K are obtained when applying this model to an external test set of 488 compounds within its applicability domain. The overall performances of the method are remarkable given its simplicity and the small number of parameters involved. In addition, the present work provides valuable insight into the influence of individual functional groups to flash point temperatures.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

In recent years, many studies have been devoted to the development of predictive methods to estimate the flash point temperature (T_F) of organic compounds, as summarized in recent reviews [1–3]. Such methods are needed for the design of new compounds to be used in processes or devices involving energy storage or transfer. For instance, current efforts to develop new organic electrolytes for lithium-based rechargeable batteries can be made more efficient if T_F may be estimated beforehand, in addition to performance-related properties [4,5]. The availability of a quantitative model for T_F is especially critical for such applications. Indeed, while relatively large compounds might be preferred to ensure a low volatility and therefore high T_F values, the requirement of a low viscosity puts some stringent limitations on the size of the solvent molecules in the electrolyte. Quantitative models for the various properties are

therefore needed to identify acceptable trade-offs between safety and performance characteristics.

Reliable predictions of T_F may be obtained from relationships involving other experimental properties, including boiling point (T_b) [6–10], heat of vaporization ($\Delta_{vap}H$) [11] and heat of combustion [12]. The applicability of such methods being restricted by the need for experimental data, this work focuses on T_F prediction from molecular formula only. Many approaches are being developed to this aim, some of them quite successful to predict the flash point of alkanes [13,14] or alkenes [15]. However, more general models are less reliable, with average absolute errors (AAE) typically in the range 10–20 K [16–22], unless they focus on specific classes of molecules such as diesel fuel components for which T_F predictions yield AAE values slightly <10 K [23]. Whereas extensively parameterized methods lead to small errors for fitted T_F data [18–21], significantly larger errors cannot be ruled out as these models are used to predict T_F for new compounds. For instance, the performance of a group contribution (GC) method is characterized by AAE = 10.7 K using statistics derived from the 418 compounds in the training set [18], to be compared with the value AAE = 14.2 K

* Corresponding author. Tel.: +33 0247344185.

E-mail address: didier.mathieu@cea.fr (D. Mathieu).

reported from predictions for an alternative data set of 287 organic compounds [21].

To date, only advanced techniques involving many empirical parameters led to AAE values <10 K for predicted flash point temperatures. These include support vector regression (SVR) [24], artificial neural networks [25] or consensus models (CM) obtained by averaging results obtained using a number of such techniques [23]. However, their complexity hampers their widespread adoption by chemical engineers. Furthermore, they do not provide straightforward insight into the relationships between molecular structure and flash point. The recent positional distributive contribution method exhibits similar drawbacks [21]. Applied to T_F prediction, it yields a value as small as 3.77 K for the AAE derived from the training set, as expected from the ratio between the number of adjustable parameters (93) and the size of the data set (287). Unfortunately, no attempt was made to assess the robustness and predictive value of this method.

2. Modeling approach

To overcome these limitations, we develop an alternative approach based upon the observation that a simple square root expression involving only 4 adjustable parameters outperforms a dedicated ANN to predict T_F for alkanes [14]. A straightforward generalization of the method to organosilicon compounds demonstrates the performance of this modeling approach compared to mainstream quantitative structure–property relationships (QSPR) methodologies [26]. However, although in principle applicable to arbitrary compounds, this generalized scheme is clearly unlikely to provide the best performance for purely organic compounds owing to its focus on Si-containing molecules. Therefore, the approach is presently applied to simple organic molecules. Following earlier studies [14,26], T_F is obtained according to a power law expression:

$$T_F = f^\alpha \quad (1)$$

where $\alpha < 1$ is an empirical exponent and f is a size-extensive quantity obtained as a sum of additive contributions associated with molecular fragments and structural features i :

$$f = \sum_i n_i f_i = \sum_i n_i t_i^{1/\alpha} \quad (2)$$

These equations were obtained on the basis of simple considerations regarding the dependence of T_F on the molecular size: (1) for small compounds, T_F must increase with molecular size as larger molecules require higher temperatures to vaporize and (2) for very large molecules, T_F must be less dependent on the molecular size as the compounds are likely to decompose before entering the vapor phase. As a consequence, T_F must be a concave increasing function of molecular size. Rather than reporting values for the additive contributions f_i derived from a multilinear regression against $f = T_F^{1/\alpha}$, we will focus on the newly introduced parameters $t_i = f_i^\alpha$ as they have the dimension of temperatures and are thus more amenable to physical interpretation. As for the exponent α , the value of 1/2 derived from n -alkane data [14] will be revised.

The molecular fragments i are defined as follows: every non-hydrogen atom is assigned a contribution denoted $Xn_c - n_H$, where X , n_c and n_H stand respectively for its atomic symbol, coordination number and number of attached hydrogen atoms. For instance, carbon atoms in methyl, methylene, ketone and cyano groups define additive fragments denoted C4-3, C4-2, C3-0 and C2-0, respectively. Such fragments are somewhat simpler than conventional functional groups $-\text{CH}_3$, $-\text{CH}_2-$, $>\text{C}(=\text{O})$ or $-\text{C}\equiv\text{N}$ because bond orders are ignored. Furthermore, a fragment such as C2-0 reflects the contribution of carbon atoms not only in cyano groups but also in

Table 1

The 21 parameters f_i of the present model with corresponding standard deviations (s.d.) and occurrence numbers N_{occ} .

i	t_i (K)	s.d. (K)	N_{occ}
<i>Atomic groups:</i>			
C3-0	136.3	0.3	111
C3-1	125.6	0.2	93
C3-2	108.5	0.3	19
C4-0	126.6	0.4	24
C4-1	122.0	0.3	87
C4-2	132.3	0.0	223
C4-3	109.5	0.2	272
N2-0	145.3	0.6	5
N3-0	168.1	0.7	3
N3-1	182.2	0.5	7
N3-2	205.6	0.3	14
O1-0	175.2	0.3	72
O2-0	112.1	0.2	47
O2-1	241.3	0.2	63
S2-0	202.8	0.6	4
S2-1	190.1	0.4	8
Cl1-0	178.7	0.4	5
<i>Contribution for the cyano group:</i>			
$-\text{C}\equiv\text{N}$	242.3	0.6	4
<i>Ring corrections:</i>			
R_5	-113.0	0.8	2
R_6	-81.9	0.5	9
R_{6a}	156.7	1.0	50

alkynes. Similarly, O1-0 describes the contribution of oxygen atoms in various groups including ketones, carboxylic acids, N-oxides, etc. Therefore, present fragments allow for a drastic decrease of the number of adjustable parameters with respect to more conventional group contribution methods. Since they were initially introduced on the basis of straightforward geometric considerations [27], they are hereafter referred to as geometrical fragments (GF).

In practice, a pure GF approach is not possible on the basis of the present data set, as a standard group contribution $-\text{C}\equiv\text{N}$ has to be introduced for cyano groups. This is because the reference data set does not exhibit any alkyne compounds. As a result, sp carbon atoms occur only in nitriles and the individual values of the C2-0 and N1-0 contributions cannot be determined independently.

Finally, three ring corrections are used: R_5 and R_6 for five and six-membered non-aromatic rings, and R_{6a} for 6-membered aromatic rings. Before being applied to T_F prediction, this approach proved highly successful for simpler properties, namely crystal density [27] and sublimation enthalpy [28]. An obvious limitation of this scheme is its inability to distinguish isomeric compounds.

The selection of the data set is critical to any method involving empirical parameters. For comparison purposes, it is clearly desirable to re-use a previous set from literature rather than introducing still another one. It is tempting to select the most extensive data sets reported so far, which contain over 1600 compounds [22]. However, methods based on molecular fragments, including the present one, have difficulties in taking advantage of this particular or a similar database, because some fragments are under-represented and their values are therefore prone to be numerically ill-defined. This is especially clear considering the number of occurrences of the fragments used to predict flash points of organosilicon compounds, listed in Table 1 of Ref. [26]. This issue is even more critical with methods based on more elaborated fragment definitions, such as GC methods [25].

In this context, the data set of 287 organic compounds introduced by Jia et al. [21] is of special interest as it is built from extensive sampling of homologous series of molecules spanning the most common functional groups, including ketones, alcohols, ethers, esters, amines, nitriles, thiols, carboxylic acids, and chloro compounds. In contrast, peroxides, nitro compounds, amides, or

Download English Version:

<https://daneshyari.com/en/article/576917>

Download Persian Version:

<https://daneshyari.com/article/576917>

[Daneshyari.com](https://daneshyari.com)