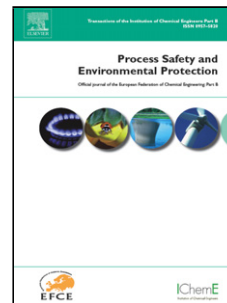


Accepted Manuscript

Title: A simple and reliable method for prediction of the flash point temperature of alcohols on the basis of elemental composition and structural parameters

Author: Mohammad Hossein Keshavarz Mohammad Jafari
Mohammad Kamalvand Akram Karami Zahra Keshavarz
Ahmad Zamani Saeedeh Rajaei



PII: S0957-5820(16)00023-9
DOI: <http://dx.doi.org/doi:10.1016/j.psep.2016.01.018>
Reference: PSEP 689

To appear in: *Process Safety and Environment Protection*

Received date: 6-10-2015
Revised date: 21-1-2016
Accepted date: 25-1-2016

Please cite this article as: Keshavarz, M.H., Jafari, M., Kamalvand, M., Karami, A., Keshavarz, Z., Zamani, A., Rajaei, S., A simple and reliable method for prediction of the flash point temperature of alcohols on the basis of elemental composition and structural parameters, *Process Safety and Environment Protection* (2016), <http://dx.doi.org/10.1016/j.psep.2016.01.018>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Highlights

- A new approach was introduced for predicting the flash point of alcohols.
- A large dataset of alcohols containing 929 data points was used.
- The new model is based on elemental composition, hydrogen bonding, and some fragments.
- Different validation methods prove high reliability of the new method.

A simple and reliable method for prediction of the flash point temperature of alcohols on the basis of elemental composition and structural parameters

Mohammad Hossein Keshavarz^{*1}, Mohammad Jafari^{*1}, Mohammad Kamalvand², Akram Karami¹, Zahra Keshavarz³, Ahmad Zamani¹ and Saeedeh Rajaei⁴

¹ Department of Chemistry, Malek-ashtar University of Technology, Shahin-shahr P.O. Box 83145-115, Islamic Republic of Iran

² Department of Chemistry, Faculty of Science, Yazd University, Yazd P.O. Box 89195/741, Islamic Republic of Iran

³ Department of Civil Engineering, Isfahan University of Technology (IUT) P.O. Box 84155-198, Isfahan, Islamic Republic of Iran

⁴ Chemistry Department, Payame Noor University, 19395-4697 Tehran, I. R. Iran Islamic Republic of Iran

***To whom correspondence should be addressed:** (1) Prof. M. H. Keshavarz, mheshavarz@mut-es.ac.ir; keshavarz7@gmail.com; Tel: (0098) 031 45225071; Fax: (0098) 031 45225068; or (2) M. Jafari, jafariemamzadeh@gmail.com

Abstract

Flash point (*FP*) is an important parameter in hazard classification, safe handling, transportation and storage of flammable liquids. On the basis of 929 experimental *FP* values of different alcohols, which were collected from different sources, a new correlation was constructed for the estimation of *FP* of alcohols. Different alcohols and phenols were studied that contain various molecular structures, *i.e.* acyclic and cyclic alcohols as well as phenols and alcohols with composite aliphatic-aromatic structures. The new correlation is based on the elemental composition and some of the structural parameters, such as intermolecular hydrogen bonding, which can be extracted from the chemical structure of any type of alcohols. The root mean square deviation of the new model for the external validation dataset containing 164 compounds is 16.4 K. Among all 929 data points, absolute percent error of the new model is greater than 10% only in 24 alcohols and less than 5% in 714 cases. Since the measured *FP* values were reported in literature sometimes differ in tens of degrees, predictions of the new simple model has been compared with one of the best available predictive methods that give much more reliable results with maximum errors less than about 70 K.

Keywords: Flash point; Alcohol; Correlation; Molecular structure; Safety

Accepted Manuscript

1. Introduction

Flash point (*FP*) is the lowest temperature at which a liquid can form an ignitable mixture in air near the surface of the liquid where its measurement requires an ignition source. Since *FP* of a volatile material is the lowest temperature at which vapors of a fluid will ignite, liquids with lower *FP* can be ignited easier. Moreover, the vapor may cease to burn when the ignition source is removed at *FP* which is different from the fire point, i.e. the temperature at which the vapor continues to burn after being ignited. Both *FP* and the fire point do not depend the temperature of the ignition source. Since a certain concentration of vapor for each flammable liquid in the air is necessary to sustain combustion, the *FP* of a flammable liquid is the lowest temperature at which there will be enough flammable vapor to ignite when an ignition source of sufficient strength is applied.

The *FP* of a pure substance or mixture is a principal indicator of flammability, which is an important parameter in hazard classification of flammable liquids. Therefore, it is essential information in practice and for safe handling, transportation and storage of many substances (Albahri, 2015; Katritzky et al., 1994; Mannan, 2012). Moreover, the knowledge of *FP* values helps firefighters extinguishing fires, because a high *FP* liquid fire pool can be extinguished by cooling with water mist, whereas low *FP* substances usually need to be blanketed by dry chemicals or foams (King, 1990; Mannan, 2012). For example, cooking oil has a high *FP*, and its temperature during burning is high (603-733K). Its gasification heat is also high and additional thermal energy is required for fuel evaporation. The cooking oil fire can be extinguished by water mist mainly through cooling the fuel surface, as the rate of supply of fuel vapor or burning rate is reduced sufficiently enough not to support the flame (Liu et al., 2004). Water, due to its large heat capacity, absorbs high quantities of heat and pulls down the temperature of fire, so it is a good candidate for extinguishing the fires of high flash point liquids. Foam, on the other hand, can be used for extinguishing the fires of high or low flash point because it can act as a blanket not as a heat absorber (Mannan, 2012). The *FP* is also used to determine the vapor explosion potential in industrial processes (Jones and Pujadó, 2006; Kong et al., 2003).

There are two methods for measuring the *FP* (King, 1990): (1) the closed cup method and (2) the open cup method, which are used for liquids with low and high *FP*, respectively. The open

cup tester approximates conditions that are met in open vessels and that would be encountered in spills (Lance et al., 1979). One disadvantage of open tester is that low boiling components of mixture may be lost to the surrounding atmosphere prior to the application of the flame, which give higher values of the *FP* (Lance et al., 1979). The closed cup methods tend to give somewhat lower results than open cup methods, because the physical barrier prevents the volatile particles to escape and causes to approximate an equilibrium between vapor and the air in the enclosed space (King, 1990; Lance et al., 1979). The closed cup tester provides an insight into the flammability of materials within enclosed spaces such as sealed containers (Lance et al., 1979).

The procedures for experimental determination of *FP* are described in various sources, which differ in their scope and experimental conditions: a) the ASTM D56 (D56, 1993) (Tag Closed Cup method) is used for materials with viscosities lower than 5.5 mm²/s and the *FP* of below 93 °C; b) the ASTM D93 (D93, 2002; White and Montemayor, 2013) (Pensky Martens Closed Cup method) is used for materials with the *FP* between 40 °C and 360 °C, such as distillate fuels, lubricating oils, fuel oils, mixtures of petroleum liquids with solids, and biodiesel fuels; and c) the ASTM D92 (D92, 2005; White and Montemayor, 2013) (Cleveland Open Cup method) is used for liquids with high viscosity and the *FP* between 79 °C and 400 °C, such as petroleum products except fuel oils. Based on the type of material and the available apparatus, one can select each of these standard methods. Details of these methods for determination of *FP* of unknown substances and the equality of the test methods are reviewed elsewhere (Janes and Chaineaux, 2013; Rowley et al., 2010a).

The measured *FP* values are affected by different factors: (1) sample size as well as its viscosity and homogeneity (in mixtures); (2) the type, position, and dimension of the ignition source; (3) temperature rise rate; (4) stirring of liquid; (5) mixing of vapor phase above pool; (6) drafts; (7) ambient pressure; (8) operator bias; and (9) fuel container condition (open or closed) (Ding et al., 2014; Ding et al., 2013; Ishida and Iwama, 1986; Kong et al., 2003; Lance et al., 1979). Therefore, the measured *FP* values reported in literature sometimes differ tens of degrees (Rowley et al., 2010a). It is important to have suitable predictive methods for those organic compounds where their experimental values of *FP* were not reported in literature. Moreover, the measurement of the *FP* of toxic, explosive, or radioactive materials is very difficult and even impossible in some cases (Katritzky et al., 2001; Suzuki et al., 1991). All these considerations reveal that reliable prediction of *FP* values for different classes of organic compounds is

desirable. The value of the *FP* of a substance is related directly to the boiling point and inversely to the vapor pressure at a given temperature. As a general rule, flammable liquids with high vapor pressures at "normal" temperatures commonly exhibit low boiling points and low *FP* (Lance et al., 1979).

Various theoretical methods have been developed for the calculation/prediction of the *FP* of pure compounds or mixtures of flammable liquids (Catoire et al., 2006; Liaw and Tsai, 2014; Liu and Liu, 2010; Phoon et al., 2014; Saldana et al., 2013; Valenzuela et al., 2011; Vidal et al., 2004). Catoire and Naudet (Catoire and Naudet, 2004) presented a simple correlation for predicting the *FP* values of pure organic molecules on the basis of their normal boiling point, standard enthalpy of vaporization, and the number of carbon atoms. Gharagheizi et al. (Gharagheizi et al., 2012) presented a linear model for the estimation of the *FP* of pure compounds on the basis of normal boiling points and chemical structure-based parameters. For those methods that need experimental data of physical and thermodynamic properties of organic compounds for prediction of the *FP* values, e.g. the methods of Catoire and Naudet (Catoire and Naudet, 2004) as well as Gharagheizi et al. (Gharagheizi et al., 2012), they may give large deviations for those compounds where the required physical and thermodynamic properties have not been reported in literature. These methods can provide good predictions of *FP* for those classes of organic compounds where these models have been developed, if experimental data of desired variables such as normal boiling point were available. Since the used experimental data for development of models are much more than variables, statistical analysis data have confirmed high reliability of these methods for those new compounds with similar molecular structures that have not been used in development of methods.

Group contribution methods (Benson et al., 1969; Poling et al., 2001) have been used extensively for the prediction of different properties, e.g. solid phase heats of formation (Argoub et al., 2014), gas phase heats of formation (Kamalvand et al., 2015), heats of fusion and fusion temperature (Jain et al., 2004; Yalkowsky, 2014), and heats of sublimation (Gharagheizi et al., 2013). Some group additivity methods have also been developed for the prediction of the *FP*. Albahri (Albahri, 2003) introduced a group contribution method for the prediction of the *FP* of hydrocarbons. This method cannot be used for some unsaturated hydrocarbons because several group additivity values are not defined. Rowley et al. (Rowley et al., 2010b) introduced a more complete group contribution model for a wide range of organic compounds including 62 group

contribution values. The reliability of these models decreases as compounds with strong interactions between constitutive fragments are considered (Mathieu and Alaime, 2014), *i.e.* in polyfunctional molecules with complex molecular structures.

Quantitative structure property relationship (QSPR) methods, which are based on complex molecular descriptors and computer codes, have also been used for the prediction the *FP* of some classes of compounds. Tetteh et al. (Tetteh et al., 1999) used 25 functional groups and their first-order molecular connectivity indices (Estrada, 2002) as inputs of a radial basis function neural network model to predict *FPs* and boiling points of 400 organic compounds. Later, Gharagheizi et al. (Gharagheizi et al., 2008) constructed a neural network model on the basis of 79 functional groups from 1378 pure compounds. Katritzky et al. (Katritzky et al., 2007) used geometrical, topological, quantum mechanical and electronic descriptors for modeling the *FP* of 758 organic compounds. Applicability of the methods on the basis of neural networks is limited because they often require unusual molecular descriptors which need to be obtained through specialized software such as Dragon (Mauri et al., 2006) and CODESSA (Katritzky et al., 2002) for a given molecule (Keshavarz et al., 2013). Several new models have been recently developed on the basis of simple structural parameters for predicting *FPs* of saturated and unsaturated hydrocarbons (Keshavarz, 2012; Keshavarz and Ghanbarzadeh, 2011) and amines (Keshavarz et al., 2013).

The purpose of this work is to introduce a simple and reliable model for the prediction of *FPs* of different aliphatic and aromatic alcohols with the general formula of $C_aH_bN_cO_dF_eCl_fBr_gI_hSi_i$. The model is based on the molecular structure of the desired alcohol, without using any computer code and experimental data.

2. Materials and method

Experimental *FP* values for 548 alcohols with the general formula of $C_aH_bN_cO_dF_eCl_fBr_gI_hSi_i$ were collected from different sources (Hsieh and Hsieh, 2005; Jia et al., 2012; Lide and Hynes, 2010; Rowley et al., 2010c; Stephenson, 1987; Yaws, 2003). All of these experimental values are reported in Table S.1, which is available in supporting information. Since each new model should be tested on a sufficiently large number of chemicals that were not used in the model building, the main dataset should be divided into two subsets: a training set and an external validation set. However, in order to create an independent external validation dataset,

selection of datapoints from the pool of available datapoints should be done in a random manner (Consonni et al., 2010). In this paper, the main dataset was divided (randomly) into two sets, a training set including 384 molecules, *i.e.* 70% of molecules, and an external validation set including 164 molecules, *i.e.* 30% of molecules. Random splitting of the main dataset as well as subsequent data analysis were carried out using the Matlab® package, v. 7.14 (MathWorks Inc). The multiple linear regression (MLR) method (Billo, 2001; Palm III, 2011) was used to build the correlations. Optimization was carried out on the basis of the training data set and the predictive ability of each model was investigated by several internal and external validation procedures.

3. Results and discussion

3.1. Core model on the basis of elemental composition and hydrogen bonding donors

A careful examination of the *FPs* of many alcohol derivatives revealed that the elemental composition as well as the specific structural parameters can be used to construct the new model. It was found that the *FPs* of alcohols primarily depend on the number of carbon, oxygen, and some halogen atoms. Hydrogen and oxygen atoms also play a secondary role in determining the *FP* of alcohols through intermolecular hydrogen bonding. On the basis of the training data set, the core correlation of *FPs* (FP_{core}) can be obtained as follows:

$$FP_{core}/K = 235.7 + 11.53a - 2.058b + 18.41d + 16.83(f + g + h) + 32.41HBD \quad (1)$$

where a , b , d , f , g , and h are the number of carbon, hydrogen, oxygen, chlorine, bromine, and iodine atoms, respectively; HBD is the number of H atoms attached to O and N atoms, *i.e.* $-OH$, $-NH_2$ and $>NH$ functional groups. HBD stands for "Hydrogen Bonding Donors" in intermolecular hydrogen bondings (Todeschini and Consonni, 2000). The contribution of the number of nitrogen, fluorine, and sulfur atoms is zero because they cannot improve the coefficient of determination (R^2) (Palm III, 2011) value. Magnitude of R^2 determines that whether regression accounts for the variation or not. Its extreme value is 1.0, which means that the regression accounts for all of the variations and that the correlation is deterministic. On the other hand, $R^2=0.0$ means that the regression accounts for non of the variations (Keshavarz et al., 2015). As seen in Eq. (1), all coefficients are positive except the coefficient of b , which indicate that increasing the values of a , d , f , g , h and HBD in a desired alcohol or phenol can rise its *FP*.

Since the value of coefficient b is smaller than the other coefficients especially HBD , its contribution in lowering FP is minor. Predictions of Eq. (1) were given in Table S.1 for all of the 548 alcohols.

3.2. Improvement of the model by adding correcting factor

In order to enhance the reliability of the model and the construction of a precise and yet simple model, a Correction Factor (CF) was considered. The effects of specific intramolecular attractions, branches, heavier atoms, and long chain alkyl groups on raising or lowering the FP s were included in the CF . The multiple linear regression method (Palm III, 2011) and the aforementioned training dataset were used to construct the second model. The optimized correlation for the FP can be given as follows:

$$FP(/K) = 233.7 + 11.67a - 2.028b + 18.08d + 17.13(f+g+h) + 33.74HBD - 18.56CF \quad (2)$$

The values of CF can be defined as follows:

- (a) *The attachment of $-OH$, $-NH_2$ or $>NH$ to tertiary carbon:* The value of CF is equal to the number of hydrogens of these functional groups, which are affected by tertiary carbon steric hindrance. For example, the reported value of the FP of 1-butanol is in the range of 302.0 to 319.3 K but for tert-butanol, the FP value is between 282.1 to 288.7 K.
- (b) *The attachment of large alkyl groups to $-OH$:* The value of CF is -1.0 for alcohols containing a long chain saturated alkyl group with more than eleven carbon atoms. For unsaturated aliphatic alcohols with more than eleven carbon atoms, the value of CF is +3.
- (c) *The presence of $C-S-C$ or $-SH$ groups in aliphatic alcohols:* It was found that in aromatic alcohols the presence of one or more heavy sulfur atoms does not affect the FP , but in aliphatic alcohols containing S atoms, the predicted FP values are underestimated. Thus, a CF should be added to include the effect of sulfur atoms in raising the FP of the molecule. For example, different experimental FP s in the range of 288.0 to 305.4 K were reported in the case of 1-propanol while if the methyl group is replaced by a mercapto- ($-SH$) group, *i.e.* in 2-mercaptoethanol, the FP will increase to

345 K. In these cases, CF is equal to -2.0.

For those compounds where above conditions are not satisfied, the value of CF is zero. Predictions of Eq. (2) are also given in Table S.1 for all of the 548 alcohols.

3.3. Coefficients of variables in models and their significance

Table 1 shows regression coefficients, standard errors, t-statistic, P-value, as well as the upper and lower bounds of a 95% confidence interval. The statistical significance of the regression coefficients in predicting the FP values can be evaluated on the basis of standard errors, *i.e.* if the standard error is small relative to each coefficient, its variable is significant. The t-statistic is the ratio of coefficients to their standard errors, so higher t-statistic values correspond to the more significant coefficients (Al-Fahemi et al., 2014; Billo, 2001). The P value shows the probability that a parameter estimated from the measured data should have the value which was determined. For P value of less than 0.05, the effect is significant and the observed effect is not due to random variations (Keshavarz et al., 2015). The t-statistic and P-values, as seen in Table 1, show that the proposed five descriptors in Eq. (1) and six descriptors in Eq. (2) have a highly significant ability to predict the FP .

As seen in Eqs. (1) and (2), all of the elements except hydrogen have positive coefficients, which mean that the FP increases with increasing the number of elements in the molecule. This is in accordance with the fact that compounds with higher molecular weights have higher FP s (Katritzky et al., 2001). The contribution of hydrogen and oxygen atoms is somewhat complicated with respect to the other elements due to their participation in the formation of hydrogen bonds. The coefficients of HBD of Eqs. (1) and (2) are positive, which is consistent with the fact that intermolecular hydrogen bonding increases the FP (Gharagheizi and Fareghi Alamdari, 2008). On the other hand, the coefficients of the number of hydrogen atoms is negative and small, which has a small effect in decreasing the FP with respect to the other elements.

The interval of experimental FP values is broad in some cases, e. g. the reported FP for 1,2-butanediol is in the range of 313.2 - 366.5 K. Hence, such experimental FP values are accompanied with considerable errors, which decreases the coefficients of determination. Table 2 shows a comparison between statistical parameters of Eqs. (1) and (2) for model building, external validation and the whole dataset compared to the method of Rowley et al. (Rowley et

al., 2010b). As a general rule, R^2 values of Eq. (2) are higher than the corresponding values for Eq. (1). The Root Mean Squared (RMS) error provides a reliable indication of the fitness of the model, which is independent of the distribution of datapoints. RMS values should be low and as similar as possible to ensure both the predictive ability (low values) and generalizability (similar values) (Gramatica, 2007). The Mean Absolute Deviation (MAD) is also a linear measure of errors that assesses the average size of errors when negative signs are ignored. For a normal (Gaussian) distribution of errors, the RMS/MAD ratio is ca. 1.25 (Keshavarz et al., 2015; Willmott et al., 2009). For Eq. (1), this ratio is 1.28 for the training set and 1.30 for the external validation set. For Eq. (2) these values are 1.26 and 1.30, respectively.

3.4. Internal Validation of models

R^2 is a useful statistical parameter but attending it solely can be misleading, so the model should be evaluated by further validation techniques, which are divided into internal and external validation methods (Leach and Gillet, 2007). The most commonly used technique for internal validation, is the cross validation (CV). In cross validation, different proportions of chemicals are iteratively held-out from the training set used for model development and predicted as unknown by the developed model in order to verify internal predictivity (Gramatica, 2007). Two modes of cross validation are leave one-out (LOO CV) and K-fold cross validation (KfCV). In LOO CV just a single data point is removed each time, and in KfCV the whole training set is divided into K groups (randomly) and in turn, each group is left out. KfCV is executed a large number of times (in the order of 100) by selecting different grouping schemes and finally a mean cross validated R^2 , which is commonly known as Q^2 , is derived.

R^2_{Model} measures goodness-of-fit whereas Q^2 values are a measure of goodness-of-prediction. In general, Q^2 is smaller than R^2 but for a well-behaved system, there should not be a significant difference between these parameters (Leach and Gillet, 2007). Robust models have Q^2 values stable upon partition size and close to R^2 (Fayet and Rotureau, 2014) and a QSPR model is called predictive if $R^2 > 0.6$ and $Q^2 > 0.5$ (Tropsha, 2010). Coefficients of determination and RMS values in different kinds of cross validation are reported in Table 3. It should be noted that the reported $Q^2_{5\text{CV}}$ and $Q^2_{10\text{CV}}$ values are the average of 1000 runs for 5-fold CV and 10-fold CV, respectively. As seen in Tables 2 and 3, R^2_{Model} and Q^2 values are all greater than 0.7, so both

models are predictive. Moreover, Q^2 values are near to the R^2_{Model} , which confirms that both models are well-behaved, robust, and more importantly, not over-fitted.

RMS values can be calculated on different datasets: a) on a training set, which is also called Standard Deviation Error in Calculation (SDEC), b) on CV sets, which is also called Standard Deviation Error in Prediction (SDEP) and c) on an external validation set (Gramatica, 2007). As seen in Tables 2 and 3, RMS values of the new models are small and very similar in training and cross validation sets, which confirms the predictive ability and generalizability of both models.

3.5. External Validation of models

Further assessment of the predictive ability of the new models can be performed by external validation (Golbraikh and Tropsha, 2002). As seen in Table 2, R^2_{External} values are slightly lower than the other coefficients of determination of each model. This is an expected result because these molecules have not been used in the model development. On the other hand, R^2_{External} values for both models are larger than the R^2_{Rowley} , which means that the new models have a better predictive ability despite their simplicity. Fig. 1 also shows a graphical comparison between the new models and the group additivity method of Rowley et al. (Rowley et al., 2010b). As seen, the predictions of the methods exhibit a lower dispersion with respect to the model of Rowley et al. (Rowley et al., 2010b). This is consistent with the fact that both models have higher R^2_{External} and lower RMS values.

Fig. 2 shows the range of absolute percent error of the new models ($\text{APE} = 100 \times |FP_{\text{exp.}} - FP_{\text{pred.}}| / FP_{\text{exp.}}$) for all 929 data points. For Eq. (1), the APE values are greater than 10 percent for 47 alcohols but are lower than 5% in 648 cases. For Eq. (2), the APE values are greater than 10% in only 24 cases and are less than 5% for 714 alcohols. Hence, the accuracy of the predicted FP values of the alcohols improves going from Eq. (1) to Eq. (2).

4. Conclusions

A simple and accurate model was developed for the prediction of the FP of various acyclic, cyclic, and aromatic alcohols with the general formula $C_aH_bN_cO_dF_eCl_fBr_gI_hSi_i$. The model is based on the elemental composition and the number of hydrogen atoms attached to O and N atoms. The effects of some of the structural parameters on the flash points of alcohols were added. These descriptors can be simply obtained from the molecular structure of alcohols. Cross

validation methods as well as an external validation technique was used to check the predictive ability of the models.

Acknowledgement

We would like to thank the research committee of Malek-ashtar University of Technology (MUT) for supporting this work.

References

- Al-Fahemi, J.H., Albis, N.A., Gad, E.A.M., 2014. QSPR Models for Octane Number Prediction. *Journal of Theoretical Chemistry* 2014, 6.
- Albahri, T.A., 2003. Flammability characteristics of pure hydrocarbons. *Chemical Engineering Science* 58, 3629-3641.
- Albahri, T.A., 2015. MNLR and ANN structural group contribution methods for predicting the flash point temperature of pure compounds in the transportation fuels range. *Process Safety and Environmental Protection* 93, 182-191.
- Argoub, K., Benkouider, A.M., Yahiaoui, A., Kessas, R., Guella, S., Bagui, F., 2014. Prediction of standard enthalpy of formation in the solid state by a third-order group contribution method. *Fluid Phase Equilibria* 380, 121-127.
- Benson, S.W., Cruickshank, F., Golden, D., Haugen, G.R., O'neal, H., Rodgers, A., Shaw, R., Walsh, R., 1969. Additivity rules for the estimation of thermochemical properties. *Chemical Reviews* 69, 279-324.
- Billo, E.J., 2001. *Excel for Chemists: A Comprehensive Guide*, 2nd ed. Wiley, New York.
- Catoire, L., Naudet, V., 2004. A unique equation to estimate flash points of selected pure liquids application to the correction of probably erroneous flash point values. *Journal of physical and chemical reference data* 33, 1083-1111.
- Catoire, L., Paulmier, S., Naudet, V., 2006. Experimental determination and estimation of closed cup flash points of mixtures of flammable solvents. *Process safety progress* 25, 33-39.
- Consonni, V., Ballabio, D., Todeschini, R., 2010. Evaluation of model predictive ability by external validation techniques. *Journal of chemometrics* 24, 194-201.
- D56, A., 1993. *Standard Test Methods for Flash Point by Tag Closed Tester*. ASTM International, West Conshohocken, PA.
- D92, A., 2005. *Standard Test Method for Flash and Fire Points by Cleveland Open Cup Tester*. ASTM International, West Conshohocken, PA.
- D93, A., 2002. *Standard Test Methods for Flash Point by Pensky-Martens Closed Cup Tester*. ASTM International, West Conshohocken, PA.
- Ding, C., Yao, W., Tang, Y., Rong, J., Zhou, D., Wang, J., 2014. Experimental study of the flash point of flammable liquids under different altitudes in Tibet plateau. *Fire and Materials* 38, 241-246.
- Ding, C., Yao, W., Zhou, D., Rong, J., Zhang, Y., Zhang, W., Wang, J., 2013. Experimental study and hazard analysis on the flash point of flammable liquids at high altitudes. *Journal of Fire Sciences* 31, 469-477.
- Estrada, E., 2002. Physicochemical interpretation of molecular connectivity indices. *The Journal of Physical Chemistry A* 106, 9085-9091.
- Fayet, G., Rotureau, P., 2014. Development of simple QSPR models for the impact sensitivity of nitramines. *Journal of Loss Prevention in the Process Industries* 30, 1-8.
- Gharagheizi, F., Fareghi Alamdari, R., 2008. Prediction of flash point temperature of pure components using a quantitative structure–property relationship model. *QSAR & Combinatorial Science* 27, 679-683.
- Gharagheizi, F., Fareghi Alamdari, R., Angaji, M.T., 2008. A new neural network– group contribution method for estimation of flash point temperature of pure components. *Energy & Fuels* 22, 1628-1635.
- Gharagheizi, F., Ilani-Kashkouli, P., Acree Jr, W.E., Mohammadi, A.H., Ramjugernath, D.,

2013. A group contribution model for determining the sublimation enthalpy of organic compounds at the standard reference temperature of 298 K. *Fluid Phase Equilibria* 354, 265-285.
- Gharagheizi, F., Keshavarz, M.H., Sattari, M., 2012. A simple accurate model for prediction of flash point temperature of pure compounds. *J Therm Anal Calorim* 110, 1005-1012.
- Golbraikh, A., Tropsha, A., 2002. Beware of q^2 ! *Journal of Molecular Graphics and Modelling* 20, 269-276.
- Gramatica, P., 2007. Principles of QSAR models validation: internal and external. *QSAR & combinatorial science* 26, 694-701.
- Hshieh, T.T., Hshieh, F.-Y., 2005. Closed-cup flash points and flammability properties of selected chemical compounds. *Journal of fire sciences* 23, 157-171.
- Ishida, H., Iwama, A., 1986. Some critical discussions on flash and fire points of liquid fuels, *Fire Safety Science—Proceedings of the first international symposium*, pp. 217-226.
- Jain, A., Yang, G., Yalkowsky, S.H., 2004. Estimation of melting points of organic compounds. *Industrial & engineering chemistry research* 43, 7618-7621.
- Janes, A., Chaîneaux, J., 2013. Experimental determination of flash points of flammable liquid aqueous solutions, 14. *International Symposium on Loss Prevention and Safety Promotion in the Process Industry. AIDIC. Milano*, pp. 943-948.
- Jia, Q., Wang, Q., Ma, P., Xia, S., Yan, F., Tang, H., 2012. Prediction of the flash point temperature of organic compounds with the positional distributive contribution method. *Journal of Chemical & Engineering Data* 57, 3357-3367.
- Jones, D.S.J., Pujadó, P.R., 2006. *Handbook of petroleum processing*. Springer.
- Kamalvand, M., Keshavarz, M.H., Jafari, M., 2015. Prediction of the strength of energetic materials using the condensed and gas phase heats of formation. *Propellants, Explosives, Pyrotechnics* 40, 551-557.
- Katritzky, A., Lobanov, V., Karelson, M., 1994. Comprehensive descriptors for structural and statistical analysis. *Revue Roumaine de Chimie* 41, 851-868.
- Katritzky, A.R., Petrukhin, R., Jain, R., Karelson, M., 2001. QSPR analysis of flash points. *Journal of chemical information and computer sciences* 41, 1521-1530.
- Katritzky, A.R., Petrukhin, R., Yang, H., Karelson, M., 2002. *CODESSA PRO User's Manual (Comprehensive Descriptors For Structural and Statistical Analysis)*. University of Florida, Florida.
- Katritzky, A.R., Stoyanova-Slavova, I.B., Dobchev, D.A., Karelson, M., 2007. QSPR modeling of flash points: An update. *Journal of Molecular Graphics and Modelling* 26, 529-536.
- Keshavarz, M.H., 2012. Estimation of the flash points of saturated and unsaturated hydrocarbons. *Indian Journal of Engineering & Materials Sciences* 19, 269-278.
- Keshavarz, M.H., Bashavard, B., Goshadro, A., Dehghan, Z., Jafari, M., 2015. Prediction of heats of sublimation of energetic compounds using their molecular structures. *J Therm Anal Calorim* 120, 1941-1951.
- Keshavarz, M.H., Ghanbarzadeh, M., 2011. Simple method for reliable predicting flash points of unsaturated hydrocarbons. *Journal of hazardous materials* 193, 335-341.
- Keshavarz, M.H., Moradi, S., Madram, A.R., Pouretedal, H.R., Esmailpour, K., Shokrolahi, A., 2013. Reliable method for prediction of the flash point of various classes of amines on the basis of some molecular moieties for safety measures in industrial processes. *Journal of Loss Prevention in the Process Industries* 26, 650-659.
- King, R., 1990. *Safety in the process industries*. Butterworth-Heinemann, London.
- Kong, D., am Ende, D.J., Brenek, S.J., Weston, N.P., 2003. Determination of flash point in air

- and pure oxygen using an equilibrium closed bomb apparatus. *Journal of hazardous materials* 102, 155-165.
- Lance, R.C., Barnard, A.J., Hooyman, J.E., 1979. Measurement of flash points: apparatus, methodology, applications. *Journal of Hazardous Materials* 3, 107-119.
- Leach, A.R., Gillet, V.J., 2007. An introduction to chemoinformatics. Springer, The Netherlands.
- Liaw, H.-J., Tsai, T.-P., 2014. Flash-point estimation for binary partially miscible mixtures of flammable solvents by UNIFAC group contribution methods. *Fluid Phase Equilibria* 375, 275-285.
- Lide, D.R., Hynes, W.M., 2010. *CRC Handbook of Chemistry and Physics*, 90th ed. CRC Press, Boca Raton, Florida.
- Liu, X., Liu, Z., 2010. Research Progress on Flash Point Prediction. *Journal of Chemical & Engineering Data* 55, 2943-2950.
- Liu, Z., Kim, A., Carpenter, D., 2004. Extinguishment of large cooking oil pool fires by the use of water mist system, Combustion Institute/Canada Section, Spring Technical Meeting, pp. 1-6.
- Mannan, S., 2012. *Lees' Loss prevention in the process industries: Hazard identification, assessment and control*. Butterworth-Heinemann, Oxford.
- Mathieu, D., Alaime, T., 2014. Insight into the contribution of individual functional groups to the flash point of organic compounds. *Journal of hazardous materials* 267, 169-174.
- Mauri, A., Consonni, V., Pavan, M., Todeschini, R., 2006. Dragon software: An easy approach to molecular descriptor calculations. *MATCH-Communications in Mathematical and in Computer Chemistry* 56, 237-248.
- Palm III, W.J., 2011. *Introduction to MATLAB for engineers*, 3rd ed. McGraw-Hill, New York.
- Phoon, L.Y., Mustaffa, A.A., Hashim, H., Mat, R., 2014. A Review of Flash Point Prediction Models for Flammable Liquid Mixtures. *Industrial & Engineering Chemistry Research* 53, 12553-12565.
- Poling, B.E., Prausnitz, J.M., John Paul, O.C., Reid, R.C., 2001. *The properties of gases and liquids*, 5th ed. McGraw-Hill, New York.
- Rowley, J., Freeman, D., Rowley, R., Oscarson, J., Giles, N., Wilding, W., 2010a. Flash point: evaluation, experimentation and estimation. *International Journal of Thermophysics* 31, 875-887.
- Rowley, J., Rowley, R., Wilding, W., 2010b. Estimation of the flash point of pure organic chemicals from structural contributions. *Process Safety Progress* 29, 353-358.
- Rowley, R., Wilding, W., Oscarson, J., Giles, N., 2010c. DIPPR® Data Compilation of Pure Chemical Properties. Design Institute for Physical Properties, AIChE, New York.
- Saldana, D.A., Starck, L., Mougin, P., Rousseau, B., Creton, B., 2013. Prediction of flash points for fuel mixtures using machine learning and a novel equation. *Energy & Fuels* 27, 3811-3820.
- Stephenson, R.M., 1987. *Flash points of organic and organometallic compounds*. Elsevier, New York.
- Suzuki, T., Ohtaguchi, K., Koide, K., 1991. A method for estimating flash points of organic compounds from molecular structures. *Journal of chemical engineering of Japan* 24, 258-261.
- Tetteh, J., Suzuki, T., Metcalfe, E., Howells, S., 1999. Quantitative structure-property relationships for the estimation of boiling point and flash point using a radial basis function neural network. *Journal of chemical information and computer sciences* 39, 491-507.
- Todeschini, R., Consonni, V., 2000. *Handbook of molecular descriptors*. John Wiley & Sons, Weinheim.
- Tropsha, A., 2010. Best practices for QSAR model development, validation, and exploitation. *Molecular Informatics* 29, 476-488.

- Valenzuela, E.M., Vázquez-Román, R., Patel, S., Mannan, M.S., 2011. Prediction models for the flash point of pure components. *Journal of Loss Prevention in the Process Industries* 24, 753-757.
- Vidal, M., Rogers, W., Holste, J., Mannan, M., 2004. A review of estimation methods for flash points and flammability limits. *Process safety progress* 23, 47-55.
- White, E.W., Montemayor, R.G., 2013. *The Practice of Flash Point Determination: A Laboratory Resource (MNL72)*. ASTM, West Conshohocken, PA., USA.
- Willmott, C.J., Matsuura, K., Robeson, S.M., 2009. Ambiguities inherent in sums-of-squares-based error statistics. *Atmospheric Environment* 43, 749-752.
- Yalkowsky, S.H., 2014. Carnelley's Rule and the Prediction of Melting Point. *Journal of Pharmaceutical Sciences*.
- Yaws, C.L., 2003. *Yaws' Handbook of Thermodynamic and Physical Properties of Chemical Compounds*. Knovel, New York.

Figure legends

Fig. 1. The predicted flash points using the developed models and the method used by Rowley et al. (Rowley et al., 2010b) versus experimental values for 164 molecules in the external validation dataset

Fig. 2. The range of absolute average errors of Eqs. (1) and (2) for the 929 data points where the predicted values of Eq. (2) are given in bold.

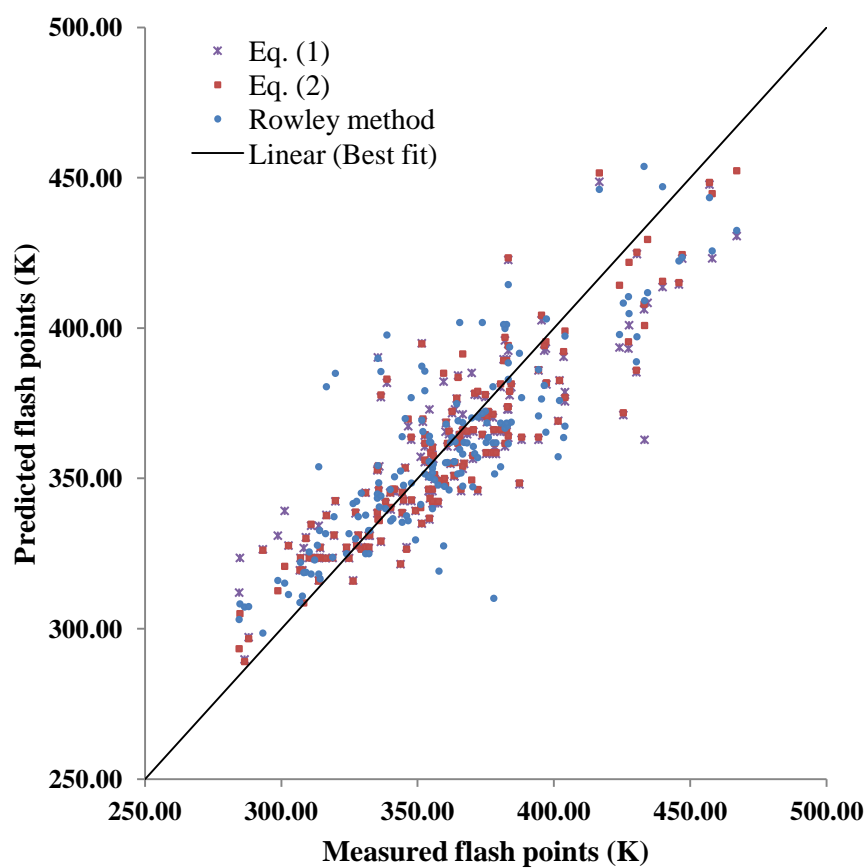


Fig. 1

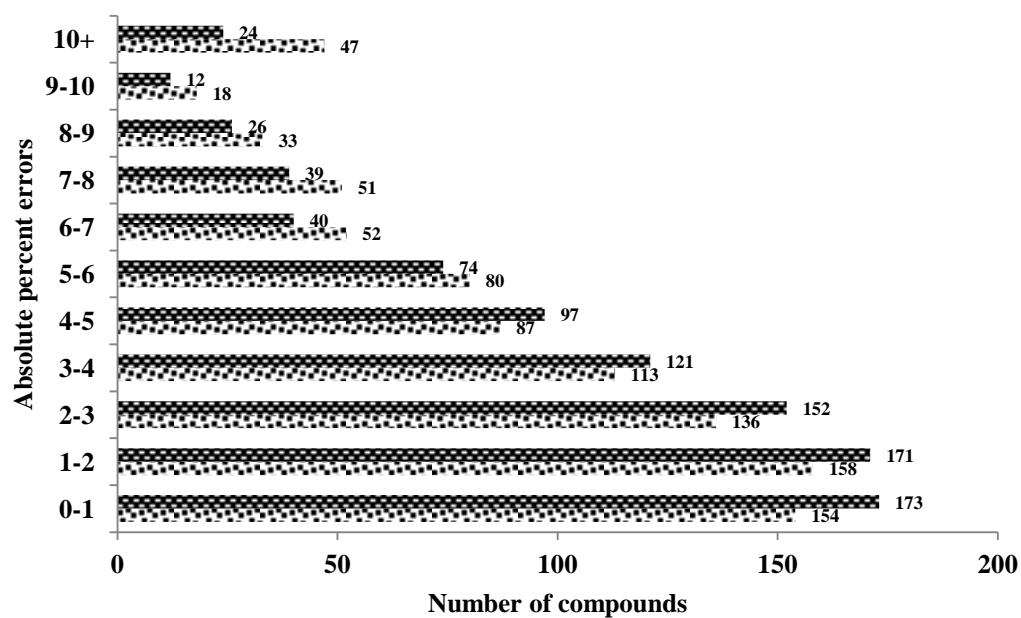


Fig. 2

Table 1.

Regression coefficients, standard errors, t-statistics, P-values, and confidence intervals for new models

Model	Variable	Coefficient	Standard error	t-statistic	P-value ¹	Lower bound ² (95%)	Upper bound ³ (95%)
Eq. (1)	Intercept	235.7	3.859	61.06	6.80E-198	228.1	243.2
	no. of C atoms (<i>a</i>)	11.53	0.4801	24.01	4.85E-78	10.58	12.47
	no. of H atoms (<i>b</i>)	-2.058	0.2436	-8.447	6.52E-16	-2.537	-1.579
	no. of O atoms (<i>d</i>)	18.41	1.330	13.84	1.54E-35	15.80	21.03
	no. of Cl, Br, and I atoms (<i>f+g+h</i>)	16.83	2.344	7.178	3.76E-12	12.22	21.44
	no. of hydrogen bonding donors (<i>HBD</i>)	32.41	1.604	20.21	4.34E-62	29.25	35.56
Eq. (2)	Intercept	233.7	3.223	72.49	1.90E-223	227.3	240.0
	no. of C atoms (<i>a</i>)	11.67	0.4007	29.13	1.50E-98	10.88	12.46
	no. of H atoms (<i>b</i>)	-2.028	0.2032	-9.976	5.88E-21	-2.427	-1.628
	no. of O atoms (<i>d</i>)	18.08	1.110	16.29	1.48E-45	15.90	20.27
	no. of Cl, Br, and I atoms (<i>f+g+h</i>)	17.13	1.956	8.760	6.70E-17	13.29	20.98
	no. of hydrogen bonding donors (<i>HBD</i>)	33.74	1.342	25.14	1.28E-82	31.10	36.38
	correcting factor (<i>CF</i>)	-18.56	1.440	-12.89	9.53E-32	-21.39	-15.73

¹ P-value=Probability of rejecting a true null hypothesis; ² Lower bound (95%)=Lower bond of a 95% confidence interval; ³ Upper bound (95%)=Upper bond of a 95% confidence interval.

Table 2.

Statistical parameters of Eqs. (1) and (2) as well as the method of Rowley et al. (Rowley et al., 2010b) for training, external validation, and total datasets

	No. of compounds	R^2			RMS			MAD ^a			Maximum of errors		
		Eq. (1)	Eq. (2)	Ref. (Rowley et al., 2010b)	Eq. (1)	Eq. (2)	Ref. (Rowley et al., 2010b)	Eq. (1)	Eq. (2)	Ref. (Rowley et al., 2010b)	Eq. (1)	Eq. (2)	Ref. (Rowley et al., 2010b)
Training	384	0.77 ^b	0.84 ^b	0.72	17.1	14.2	22.5	13.4	11.3	15.5	70.8	48.2	105.1
External validation	164	0.73 ^c	0.81 ^c	0.70	19.2	16.4	23.6	14.8	12.6	15.9	70.4	53.6	80.6
Total Dataset	548	0.78	0.84	0.77	18.2	15.7	21.0	14.0	12.1	14.4	70.8	68.5	105.1

^a Mean Absolute Deviation; ^b R^2_{Model} ; ^c R^2_{External} .

Table 3.

Coefficient of determinations and RMS values in cross validation sets

		R ²		RMS	
		Eq. (1)	Eq. (2)	Eq. (1)	Eq. (2)
Leave one-out CV	383	0.76 ^b	0.83 ^b	17.5	14.6
5-fold CV ^a	379	0.75 ^c	0.82 ^c	17.4	14.6
10-fold CV ^a	374	0.74 ^d	0.81 ^d	17.3	14.5

^a For 1000 runs; ^b Q²_{LOO}; ^c Q²_{5CV}; ^d Q²_{10CV}.