# Mixture descriptors toward the development of Quantitative Structure-Property Relationship models for the flash points of organic mixtures

Théophile Gaudin, Patricia Rotureau, Guillaume Fayet

# Mixture descriptors towards the development of QSPR models for the flash points of organic mixtures

Théophile Gaudin, Patricia Rotureau, Guillaume Fayet*

INERIS, Parc Technologique Alata, BP2, 60550 Verneuil-en-Halatte, France

*Corresponding author: guillaume.fayet@ineris.fr; tel: +33(0)344618126; fax: +33(0)344556565

**Abstract**

Quantitative structure property relationships (QSPR) are increasingly used for the prediction of physico-chemical properties of pure compounds but only few were developed to predict the properties of mixtures.

In this work, a series of existing and new formula were proposed to derive mixture descriptors to develop QSPR models for mixtures. These mixture descriptors were used to model the flash point of a series of 435 organic mixture compositions. Multilinear models were obtained using twelve different mathematic formulas taking into account the linear or non-linear dependences of the flash point with the concentration of each compound.

The best model, issued from the newly proposed $(x_1 d_1 + x_2 d_2)^2$ formula, was a four parameter model presenting good prediction capabilities (with a mean absolute error in prediction of 10.3°C) compared to existing predictive methods for both mixtures and pure compounds.

**Keywords:** QSPR models; mixtures; flash point; mixture descriptors

**Introduction**

Quantitative Structure Property Relationships (QSPR) are predictive models allowing the prediction of macroscopic properties by correlation with descriptors of the molecular structure of chemicals[1]. These molecular descriptors are of various categories[1, 2]: constitutional, topological, geometric or quantum chemical. Such methods have been largely used for biological activities in the field of toxicology[3], ecotoxicology[4] or pharmaceutics[5, 6] and are increasingly used for physico-chemical properties[7-9].

Various models have been developed for hazardous physico-chemical properties[10-19] such as flammability[10, 11], thermal stability[12-14] and explosibility[15-19]. Up to now, the QSPR approach was mostly dedicated to pure compounds and only few recent works were dedicated to mixtures[20]. Ajmani et al. proposed various models to predict the density[21, 22] and the infinite-dilution activity coefficient[23] of binary mixtures. In these studies, the molecular descriptors for each pure compound were combined, e.g. by mole weighted averaging[24], to derive mixture descriptors. These mixture descriptors were then correlated to the property of the studied mixtures. Several studies were also dedicated to azeotropic mixtures[25-29]. In particular, Oprisiu et al.[29] developed several QSPR models to predict the boiling point[25, 28] of azeotropic binary mixtures based on fragment descriptors.

The flash point (FP) is the temperature at which the vapor above a flammable liquid ignites under the effect of a spark.[30] This property characterizes flammability hazards of liquids and is a key safety issue in the risk assessment of industrial processes and in various regulatory frameworks dedicated to chemicals (for use, storage and transport)[31, 32]. Flash point of pure compounds was studied in several works in view of developing predictive methods taking advantage of the large availability of data[33, 34]. Among them, many were based on the knowledge of other properties like the boiling point[24, 35, 36]. The highest performances were obtained by Carroll et al. [37] with a model based on boiling points and functional group counts with a mean absolute error (MAE) of 2.5°C for a data set of 1000 pure compounds. Other QSPR models were developed without using any experimental parameters[11, 38, 39]. For instance, Katritzky[11], Saldana[38] and Rowley[39] proposed QSPR models only based on theoretical descriptors with prediction errors estimated at 16.1°C, 10.3°C and 9.8°C, respectively, for organic compounds. If interesting performances were obtained with these global models (for large families of compounds), models focusing on particular families of compounds were also developed in recent works. For instance, an accurate model was obtained for the flash points of amines with a coefficient of determination $R^2$ of 0.91 in prediction that corresponds to a MAE of 14.9°C[40]. Moreover, Khajeh et al.[41] developed a specific model for esters with an error in prediction of 15.1°C.

Concerning the prediction of the flash point of mixtures, mixing rules are commonly used by estimating the flash point of the mixture from the ones of each constituent as function of their respective concentration in the mixture. Various mixing rules exist for the flash point. Wickey[42] and Affens[43] proposed methods more likely dedicated to ideal mixtures, i.e. presenting no important interactions between their components. Then, over decades, studies taking into account intermolecular interactions between constituents through the use of their activity coefficients[44-47] were conducted. Liaw developed a computational approach based on the mixing rule proposed by Wu[47] and published several experimental studies[48-50] to show that this approach could yield accurate predictions. This last approach succeeded to predict the behavior of particular mixture features like those presenting a flash point below the flash point of their components (e.g. octane/methanol[48] or

octane/i-propanol[49]) or, on the contrary presenting maximum flash-point behavior (e.g. cyclohexanol/phenol[50]).

In recent work, Saldana et al.[51] proposed the first QSPR model for the prediction of flash point of organic liquid mixtures. In this final model, the mole weighted average values of the molecular descriptors of each pure compound of the mixture were used in a genetic algorithm to develop a multilinear regression model (GA-MLR) with four descriptors and an error in prediction of 10.1°C.

In a previous work[52], we also demonstrated the possibility to combine QSPR models with mixing rules notably by developing a full predictive method based on the mixing rule of Liaw combined with the model of Rowley (only based on group contributions) with an error in prediction of only 4.4°C. Saldana[51] obtained same accuracy (MAE = 3.4°C) based on a similar approach using three SVM-based QSPR models for the predictions of the flash point, heat of vaporization and boiling point of pure compounds that were introduced in a modified Liaw mixing rule.

The present study aimed to develop new QSPR models for the flash point of binary mixtures and to evaluate the potential of various mathematic formulas of mixture descriptors for the development of such models for the prediction of this property. Indeed, the definition of mixture descriptors was identified as a critical parameter in existing QSPR models for mixtures, in particular for such property that can follow a non linear trend with concentration of each component. Moreover, in this study, an extended database was used, by gathering additional data from literature to the ones already collected and used by Saldana et al.[51] from 287 to 435 data, to extend the diversity of the compounds involved in the mixtures.

## Materials and methods

### Experimental data

The flash point is the lowest temperature at which a liquid releases enough vapor to form a flammable mixture with air above the liquid surface. This property in notably used in the assessment risks associated to the storage and use of flammable liquids. In practice, this property is measured as the lower temperature at which a flash is obtained by application of a flame above the liquid surface. Various apparatus are used, according to standards (e.g. ISO 1516, DIN 51755-1, ASTM D56, etc.)[32] recommended in the European regulation related to the registration, evaluation, authorisation and restriction of chemical substances (REACH)[53] and in the CLP regulation[32]. For instance, the ASTM D56 standard[54] (used for most of the experimental data) corresponds to flash points in a tag closed tester, for liquids presenting a viscosity lower to 5.5 $mm^2/s$ and for flash points lower than 93°C. The reliability of the measures within this standard is estimated to be 4.3°C below 60°C and reaches 5.8°C above 60°C.

Concerning liquid mixtures, various phenomena affect the flash point thus involving that it does not follow a linear trend with the concentration of each component[30]. At first, the difference in volatility of each compound influences the concentration of vapors above the liquid surface, which therefore follows a non linear trend. Moreover, in non-ideal mixtures, intermolecular interactions favor or disfavor the transfer of molecules from the liquid to the vapor phase leading to positive or negative deviation of the profile of flash point with the concentration of each compound.

In this study, a large dataset of 435 experimental data on binary mixtures was gathered issued from 13 references in literature[47-50, 55-63]. These data are available in Supporting Information (table S1). This dataset represents, to our knowledge, the largest collection of flash points of mixtures. It comprises 43 different binary mixtures of 34 pure compounds in various concentrations. Different families of compounds are represented, like hydrocarbons, alcohols, ketones, esters or acids. Flash point values range from -27.5°C to 110.4°C and molar fractions from 0 to 1.

All flash points were obtained according to standards recommended by the REACH regulation. Nevertheless, as explained before, each standard protocol presents its own limits and uncertainties. For this reason, it has been chosen in this work to develop models based on data obtained according to a single standard to ensure at best the homogeneity of protocols. The remaining data were used to build the validation set used to evaluate the predictive power of the developed models.

So, the training set of the models was constituted by the 284 flash points obtained by Liaw et al.[48-50, 55, 56], using the ASTM D56 standard. As summarized in table 1, this training set contains 22 mixtures of 21 pure compounds. These data were, for the most of them, already used by Liaw et al. to assess the performances of the mixing rule of Wu[47] and by Saldana et al.[51].

Then, to estimate the predictive power of the obtained models, the validation set of the 151 flash points obtained using other standards (recommended by REACH) were used. 8 out of the 21 pure compounds involved in these mixtures were already present in the training set and the 13 remaining ones are representatives of chemical families present in the training set. So, they can all be considered as being within the applicability domains of the developed models. It should be noted that all the mixtures involved in the validation set were constituted with at least one pure compound not represented into the training set. To the end, as presented in figure 1, the distribution of flash points is homogeneous in both sets compared to that of the entire dataset.

**Molecular descriptors**

The structures of the 34 pure compounds of the gathered dataset were calculated using the density functional theory (DFT) in Gaussian09 package[64]. Geometric structures were optimized using the B3LYP functional[65] and 6-31+G(d,p) basis set, already commonly used for organic compounds[66-68] . Vibrational frequencies were computed at same level of theory to ensure that all stable species presented no imaginary frequency.

More than 300 molecular descriptors were computed to characterize these molecular structures. These descriptors can be considered into four classes. Constitutional descriptors are related to the identification and count of specific features of molecules, e.g. numbers of atoms, bonds or chemical groups. Topological descriptors arise from the two-dimension (2D) structure of molecules. Calculated from the connectivity table of molecules, they characterize their size and shape. Geometric descriptors stem for the three-dimension (3D) structure, like distances, angles in the molecules or molecular volumes. Finally, quantum chemical descriptors allow access to binding, energetic, electronic and thermodynamic information.

Most of these descriptors were calculated using Codessa software[69] from the previously DFT calculated structures. More details on the calculated descriptors are available in Ref. [70]. Additional descriptors were also considered, such as the presence and counts of functional groups identified in

the dataset. This was also the case of quantum chemical descriptors such as conceptual DFT descriptors[71, 72], as already done in previous works [73, 74].

**Mixture descriptors**

Upon the type of property and mixtures under study, different strategies can be used to access QSPR models applicable to mixtures[20]. In the present case, mixture descriptors *D* were developed by combining the molecular descriptors $d_i$ of each component taking into account their respective molar fractions $x_i$ as shown in eq. 1.

$$D = f(d_i, x_i) \quad (1)$$

As already observed[48], the flash point of mixtures does not follow simple dilution effect. Indeed, it is also affected by other effects like the affinity between compounds and by their respective vapor pressure[33]. So, not only linear but also non-linear formula could be of great interest. As a consequence, 12 mathematic formulas were tested in this study for the development of QSPR models to predict the flash point of binary mixtures: five were extracted from literature whereas the other seven ones were newly proposed.

The 12 proposed formulas were divided into three classes. The six first ones were dedicated to the direct correlation with the property of the mixture (direct combinations) whereas the three ones of the second class aimed to describe the deviation of the property from the linear contribution of each component with their respective concentrations (deviation combinations). The last class gathered three other combinations that did not fit with the constraints of the first two classes (other combinations).

*Direct combinations*

At first, the proposed direct combinations must be not dependent on the order of the two components of the binary mixture. Indeed, the flash point of a mixture of 30 %mol ethanol in n-octane (70%) is the same as the one of a mixture of 70 %mol n-octane in ethanol (30%). So, all proposed formulas were symmetric regarding molecular descriptors and molar fractions (see eq. 2).

$$D = f(d_1, x_1, d_2, x_2) = f(d_2, x_2, d_1, x_1) \quad (2)$$

Moreover, to be applicable to pure compounds, these formulas may be dependent on the molecular descriptor $d_1$ (or $d_2$) of only one component when $x_1$ (or $x_2$) becomes equal to 1, i.e.:

$$if \ \ x_1 = 0, \quad D = f(d_2, x_2) \quad (3)$$

The first proposed formula is the molar contribution (**fmol_sum**), corresponding to the linear combination of the molecular descriptors $d_i$ of both compounds weighted by their respective molar fraction $x_i$ in the mixture.

$$D = x_1 d_1 + x_2 d_2 \quad (4)$$

This formula has already been used in previous QSPR models for mixtures[21, 26, 75] and in particular by Saldana et al.[51] for the flash point of mixtures.

Four other formulas were designed in this study following these criteria. The weighted difference (**fmol_diff**) consisted in the absolute difference between the descriptors of each component weighted by their respective molar fraction in the mixture.

$$D = \left| x_1 d_1 - x_2 d_2 \right| \tag{5}$$

The square molar fraction (**sqr_fmol**) and root square molar fraction (**root_fmol**) formulas correspond to a linear contribution of the descriptors of each component with respect to their square and root square molar fraction, respectively.

$$D = {x_1}^2 d_1 + {x_2}^2 d_2 \tag{6}$$

$$D = \sqrt{x_1} d_1 + \sqrt{x_2} d_2 \tag{7}$$

The square molar contribution (**sqr_fmol_sum**) is the square of the molar contribution.

$$D = \left( x_1 d_1 + x_2 d_2 \right)^2 \tag{8}$$

At last, the norm of the molar contribution (**norm_cont**, in Eq. 9), already proposed by Saldana et al. [51], was also tested, even if it was not selected for their final model.

$$D = \sqrt{\left( x_1 d_1 \right)^2 + \left( x_2 d_2 \right)^2} \tag{9}$$

*Deviation combinations*

As for direct combinations, all tested deviation combinations were symmetric regarding molecular descriptors and molar fractions (see eq. 2). However, to be applicable to pure compounds, these last formulas may not be dependent on any molecular descriptor, $d_1$ or $d_2$. Indeed, for pure component, the deviation to linearity becomes equal to 0, and formulas must follow eq. 10.

$$If \;\; x_1 = 0 \; or \; x_2 = 0, \quad D = 0 \tag{10}$$

Three combinations were proposed in this category, based on the difference between the molecular descriptors $\Delta d_i$ and a term depending on the absolute difference between the molar fractions of both components of the mixture $\Delta x_i$. None of these formulas were already tested.

The most simple, denoted **mol_dev** (in eq. 11), considered the fact that the deviation issued by the mixture of two compounds is related to the difference between the descriptor value for each pure compound and the difference in molar fraction. Moreover, the more the mixture is balanced between the two components, the more the property of the mixture deviates from a linear behavior.

$$D = \left( 1 - \Delta x \right) \Delta d \tag{11}$$

The two other formulas, denoted **sqr_mol_dev** and **mol_dev_sqr**, were obtained by squaring $\Delta x_i$ or the $(1 - \Delta x_i)$ term.

$$D = \left( 1 - \Delta x^2 \right) \Delta d \tag{12}$$

$$D = (1 - \Delta x)^2 \Delta d \quad (13)$$

*Other combinations*

In addition, other combinations have been considered as they were already used for the development of previous QSPR models for mixtures[22, 76]. For instance, the centroid approach (**cent**) has notably been used by Jover et al.[76] for the prediction of the dissociation constant (pKa) of phenols in various solvents. In this approach, the mixture descriptor is calculated as the average value of the molecular descriptor of the *n* constituents of the mixture.

$$D = \frac{\sum d_i}{n} \quad (14)$$

Ajmani[22] proposed another formula to predict the density of liquids by using the square of the difference (**sqr_diff**) between the two constituents of the mixture.

$$D = (d_1 - d_2)^2 \quad (15)$$

At last, it has been proposed to also consider the absolute difference (**abs_diff**) between the molecular descriptors of each constituent of the mixture.

$$D = |d_1 - d_2| \quad (16)$$

It should be noted that these formula do not take into account the respective molar fraction of each constituent in the mixture. So, they are invariant within the whole range of molar fraction from 0 to 1 and could be strictly not applicable for pure compounds since the molecular descriptor values for two compounds are needed.

**Development of models**

In this study, multilinear regression models have been developed following eq. 17.

$$Y = \sum a_i D_i + a_0 \quad (17)$$

where *Y* is the calculated property, $D_i$ are the mixture descriptors and $a_i$ the regression constants.

In these equations, an excess of descriptors could lead to an over-parameterization of the final model. So, a selection of the descriptors included into the model has to be performed. In the present study, the heuristic method[70], implemented in Codessa software[69], was used due to its successful use in previous works for the development of QSPR models of pure compounds, and in particular for the flash point of organic compounds[11].

In a first step, descriptors presenting missing values or small magnitude of variation were eliminated. Then, one-parameter correlations were computed and descriptors presenting the lowest correlation with the property were canceled. Moreover, if two descriptors were highly inter-correlated, the one presenting the lowest correlation with the property was eliminated. Then, descriptors were listed by decreasing order of correlation with the property. The two-parameter equations were developed and the most correlated ones presenting no inter-correlation between the included descriptors were

selected. After that, additional descriptors were added for each selected pairs with the same procedure as long as the models increase in correlation. The output of the algorithm was a list of the best correlations obtained.

To the end, the final models were selected among the list proposed by the heuristic algorithm, to represent the best compromise between the correlation of the models and the number of descriptors.

The reliability of the model was evaluated by using internal and external validations[77]. The correlation of the final models was evaluated based on the correlation coefficient ($R^2$) and the mean absolute error (MAE). The pertinence of each descriptor in the model was checked based on a Student's t-test at a 95% level of confidence. Then, the robustness of models was estimated by leave-one-out cross-validation ($Q^2$). Finally, the predictive power of models was estimated by external validation based on the correlation ($R^2_{EXT}$) and mean absolute error ($MAE_{EXT}$) for the predictions on the validation set. Since all the mixtures involved in the validation set were constituted with at least one pure compound that was not represented into the training set, the performed external validation represented a "compounds-out" external validation as defined by Muratov[78] and Oprisiu[25].

**Results and discussions**

New QSPR models for the flash point of organic binary mixtures were developed to evaluate the potential of the twelve proposed formulas (among which seven were new) for the development of such models for the prediction of this property. Their details are available in Supporting Information (tables S2-S13) and their performances are gathered in table S14.

At first, it should be noted that the deviation formulas aimed to describe the deviation of the property from the linear contribution of each component with their respective concentrations. So, the three corresponding formulas (**mol_dev**, **sqr_mol_dev** and **mol_dev_sqr**) were not tested directly on the flash point but on the difference between the actual experimental flash point and the linear contribution of the experimental flash points of the pure compounds weighted by their respective molar fraction in the mixture (deviation to linearity). But, all three deviation formulas (Tables S8-S10) failed in accessing reliable models with only 0.34 in $R^2$ for **sqr_mol_dev** and 0.32 in $R^2_{EXT}$ for the **mol_dev** and the **sqr_mol_dev** formulas. Moreover, the mean absolute errors of these models were high, about 7°C, which is high regarding the values of the deviations to linearity of the tested mixtures, as shown in figure 2. Indeed, 222 out of the 435 data presented deviations to linearity lower than 5°C and 331 lower than 10°C.

Concerning the nine others formulas, directly used to develop models for the prediction of FP, better results were obtained. In terms of goodness-of-fit, the mean absolute errors obtained for the training set were lower than 12.1°C. This value is in the range of values obtained for existing QSPR models to predict the flash points of pure compounds[8] (exception given for those needing experimental data, e.g. boiling point). Nevertheless, some models demonstrated better predictive capabilities than other, as shown on figure 3 which represents their performances in terms of MAE and $MAE_{EXT}$ values. Indeed, even if they presented low MAE values, the formulas based on differences, **fmol_diff**, **sqr_diff** and **abs_diff** presented the highest errors in prediction with $MAE_{EXT}$ = 17.6°C, 26.2°C and 28.1°c, respectively. The best results were obtained with the centroid approach (**cent**) with $MAE_{EXT}$ = 9.7°C and with the square linear combination weighted by the molar fraction (**sqr_fmol_sum**) with

$MAE_{EXT}$ = 10.3°C. Two other formulas, based on weighted sums, were also interesting from a predictive point of view, the molar contribution **fmol_sum** and the root square weighted combination **root_fmol**, with models reaching $MAE_{EXT}$=11.8°C and 11.9°C, respectively. In conclusion, four out of the twelve models were selected from a statistical point of view: **cent**, **sqr_fmol_sum**, **fmol_sum** and **root_fmol**.

To select the best formula among these four combinations, the capacity of the models issued from each combination to reproduce qualitatively the shapes of the mixture profiles of the dataset was examined. From this point of view, the centroid approach may be less pertinent. Indeed, the flash point, as various properties[79, 80], varies with the composition of the mixture whereas the centroid approach does not take in account molar fractions. For instance, the centroid formula gives the same value of FP for the whole range of concentration for the 4-methyl-2-pentanone/1-butanol mixture (as shown in figure 4), even if this profile is among the simplest one with an ideal profile. The **sqr_fmol_sum**, **fmol_sum** and **root_fmol** approaches more likely reproduced the dilution effect by taking into account the molar fraction of each component of the mixture.

In the dataset, different families of compounds are represented and different types of mixtures can be identified. Table 2 summarized the formulas allowing access to the MAE lower than 10°C for different types of mixtures. The formulas revealed to be the most efficient were globally the same. In particular, the three selected combinations, i.e. the **cent**, **sqr_fmol_sum**, **fmol_sum** and **root_fmol** combinations, presented good capabilities for various types of mixtures.

Nevertheless, among these three mathematic formulas, the molar contribution (**fmol_sum**) appeared to be the less adaptable to reproduce strongly non linear profiles for non-ideal mixtures like in the case of the 1-propanol/octane mixture (in figure 5). This type of mixture represents a critical safety issue since the mixture revealed more flammable than the pure compounds individually. So, it is very interesting for the predictive model to be able to identify and characterize such behaviors.

In both figures 4 and 5, it should be noted that a large part of the deviations observed along the profile could be due to large errors for the prediction for pure compounds. In particular, none of the tested formula led to models lower than 10°C in MAE for hydrocarbon/hydrocabon mixtures due to this failure (see table 2). In these cases, as these mixtures present nearly ideal profiles, errors can be likely attributed to the ability to predict the flash point for the pure compounds, for instance of the predictions of pure n-decane and n-dodecane as shown in table 3. This was clearly illustrated in figure 6 for the n-octane/n-decane mixture for which the flash point of pure n-decane was predicted with more than 25°C in error leading to important errors for the largest concentrations in n-decane.

Such failure was due to the small diversity of compounds in the training set. Indeed, if the overall number of data was important (284 data), the diversity in terms of chemical structures remained low with only 22 mixtures of 21 pure compounds. Figure 7 confirmed this statement since the errors calculated for the pure compounds (between 12.4°C to 22.1°C) were higher than for the mixtures (in figure 3). Nevertheless, these performances were close to the ones obtained for existing QSPR models for the flash point of pure compounds involving solely computed molecular descriptors. Higher accuracy was found for models using the boiling point[8] like the models of Carroll[37] and of Gharagheizi[81], developed for highly larger diversity of compounds (1471 and 1000 pure compounds, respectively) that reached MAE of 2.8°C and 8.5°C, respectively.

The last point to discuss is the type of descriptors used in the developed models. In the 12 developed models, 23 molecular descriptors were used, as shown in Supporting Information (tables S2-S13). In particular, the Randic index (order 3) ${}^{3}\chi$ is used in 7 models. Moreover, it is included into the four models presenting the best predictive capabilities (**cent**, **sqr_fmol_sum**, **fmol_sum** and **root_fmol**). Randic indices ${}^{n}\chi$ are topological indices that were introduced in 1975[82], to characterize molecular branching and that have been notably correlated to the boiling point of alcanes[82]. Since the evaporation of the liquid is a critical parameter within the flash point property, the introduction of a Randic index within the models is meaningful.

**Details of the best model**

Based on the previous analysis, the model issued from the square molar contribution formula (**sqr_fmol_sum**) was selected as the best model among the 12 calculated ones. Indeed, it not only presented reliable predictions on the validation set from a statistical point of view ($MAE_{EXT}$=10.3°C), but it also appeared as applicable to pure compounds and for ideal and non-ideal mixtures as a function of molar fractions, as illustrated in figures 4 and 5.

$$\begin{aligned} FP(^{\circ}C) = \; & 50.3 + 16.3\left(x_1 {}^{3}\chi_1 + x_2 {}^{3}\chi_2\right)^2 + 5.5\cdot 10^{-3}\left(x_1 HDCA2_1 + x_2 HDCA2_2\right)^2 \\ & - 2.4\cdot 10^{-6}\left(x_1\Delta\alpha_1 + x_2\Delta\alpha_2\right)^2 - 88.0\left(x_1 V_{\min,H,1} + x_2 V_{\min,H,2}\right)^2 \end{aligned} \quad (18)$$

where $x$ is the molar fraction, ${}^{3}\chi$ is the Randic index (order 3), *HDCA2* is the HDCA H-donors charged surface area, $\Delta\alpha$ is the anisotropic polarizability and $V_{min,H}$ is the minimum valency of a H atom.

The pertinence of the Randic index has been already discussed in previous section. It also should be noted that the H-donor charged surface area (HDCA2) can be associated to the capacity of a molecule to interact with surrounding molecules through hydrogen bonds that influences relative liquid vapor equilibrium to each compound of the mixture.

This model presented good reliability on both the training and validation sets with MAE=6.3°C and 10.3°C, respectively and it was evidenced as robust by cross validation with $Q^2$=0.79, i.e. close to $R^2$=0.80. The predictions issued from model eq. 18 were already accurate compared with existing QSPR models for pure compounds that commonly present about 7-8°C in prediction error[10], especially when only based on calculated parameters. If more diverse data were available, it could become even more accurate.

As can be seen in figure 8, good correlations were obtained for most part of the dataset. Nevertheless, three specific mixtures (isoamylic alcohol/isoamylacetate, o-xylene/cyclohexanol and o-xylene/heptane), circled in figure 8 (one in the training set, two in the validation set), seemed to decrease the calculated performances of the model. This can be explained by considering the constitution of the data set. Indeed, the first one, the isoamylic alcohol/isoamylacetate mixture, is the only alcohol/ester mixture of the training set. Moreover, only two other esters were used in the training set with structures relatively different from the isoamylacetate. Indeed, methylacetate and methylacrylate did not present a carbon chain, as in isoamylacetate. So, this mixture could be considered in the limit of the applicability domain of the model. Considering the two other mixtures presenting large errors, they both include the o-xylene and it should be noted that only one aromatic compound (toluene) was represented within the whole training set used for the fitting of the model. So, this mixture could also be considered out of the applicability domain of the model. Finally, when

excluding the mixtures of the validation set with aromatic compounds, the predictive power of the model reached a $MAE_{EXT}$ of only 8.3°C (with respect to 10.3°C when considering these three mixtures). This demonstrated the importance of the constitution of the dataset for the development of reliable models not only in terms of quantity of data but also in terms of diversity of compounds and mixtures.

Besides, this model presents similar predictive capabilities than the only existing QSPR model developed by Saldana et al.[51] that presented an error in prediction $MAE_{EXT}$ of 10.1°C evaluated for a more reduced set of compounds. Indeed, the validation set was composed of 30% of the entire dataset that was almost identical to the training set of the present study. Moreover, the $MAE_{EXT}$ of the new model (Eq. 18) decreased to 8.3°C when excluding the mixtures with aromatic compounds that were not represented in the model of Saldana.

Finally, this QSPR model for mixtures was compared to the full predictive combined approach proposed in our previous work[52]. This last approach lied on a combination of the Liaw mixing rule[49] with the QSPR model of Rowley[39] for the calculation of the flash point of the pure compounds and fixed averaged values for their Antoine's coefficients. Table 4 summarizes the performances of the two approaches for the 151 data of the validation set that were not already used in previous works. The performances of the QSPR model revealed lower than the combined approach ($MAE_{EXT}$ = 10.3°C vs. 4.7°C). The availability of a more diverse dataset in terms of chemical structures would help as the prediction of flash points on pure compounds revealed in this study of great importance on the performances of the various tested formula. Nevertheless, this limitation also highlights the need for further methodological developments in the field of mixture QSPR modeling, beyond mixture descriptors formulation, for instance by considering algorithms and strategies for the development of mixture QSPR models taking into account the specificities of mixtures.

**Conclusions**

In this study, a series of mixture descriptors have been tested with the aim to achieve reliable models to predict the flash point of mixtures. A data set of 435 experimental data points for 43 binary mixtures was gathered which represents, to the best of our knowledge, the largest exiting database of the flash points for mixtures.

Considering that in the case of flash point, non linear effects are observed with concentration, new non-linear formulas were tested. In total, twelve combinations were proposed to calculate mixture descriptors for the molecular descriptors of pure compounds. For each of them, QSPR models were developed and validated on an external data set. In the case of the flash point, non linear effects are observed with concentration. So, the best results were obtained with a new non-linear formula, the square molar contribution $(x_1d_1 + x_2d_2)^2$, due to the fact that non linear effects were observed on the FP with concentration. Considering the limited chemical diversity of the available data (with only 34 different pure compounds), the model developed with this formula reached yet promising predictive capabilities, with only 10.3°C in MAE for the validation set, but it can also be fitted for either ideal and non-ideal mixtures. Moreover, excluding mixtures with aromatic compounds (that can be considered as outliers), this model reached a MAE of 8.3°C.

These first results encourage further developments with larger and more diverse databases to access high predictive models by using such non-linear mixture descriptors.

**Acknowledgement**

This work was performed using HPC resources from GENCI-CCRT (Grant 2013-t2013086639).

**Supporting Information**

Tables containing experimental data, final models obtained with each mixture descriptor's formula and a summary of the performances of these models. This information is available free of charge via the Internet at http://pubs.acs.org/.

**References**

(1) Karelson, M.; Lobanov, V. S.; Katritzky, A. R. Quantum-Chemical Descriptors in QSAR/QSPR Studies. *Chem. Rev.* **1996,** *96*, 1027-1044.
(2) Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*. Wiley: Weinheim, 2000.
(3) Benigni, R.; Bossa, C.; Netzeva, T.; Worth, A. *Collection and Evaluation of (Q)SAR Models for Mutagenicity and Carcinogenicity*; European Commission, Joint Research Centre: 2007.
(4) Bradbury, S. P. Quantitative structure-activity relationships and ecological risk assessment: an overview of predictive aquatic toxicology research. *Toxicol. Lett.* **1995,** *79*, 229-237.
(5) Grover, M.; Singh, B.; Bakshi, M.; Singh, S. Quantitative structure-property relationships in pharmaceutical research - Part 2. *Pharm. Sci. Technol. To.* **2000,** *3*, 50-57.
(6) Grover, M.; Singh, B.; Bakshi, M.; Singh, S. Quantitative structure-property relationships in pharmaceutical research - Part 1. *Pharm. Sci. Technol. To.* **2000,** *3*, 28-35.
(7) Katritzky, A. R.; Kuanar, M.; Slavov, S.; Hall, C. D.; Karelson, M.; Kahn, I.; Dobchev, D. A. Quantitative Correlation of Physical and Chemical Properties with Chemical Structure: Utility for Prediction. *Chem. Rev.* **2010,** *110*, 5714-5789.
(8) Dearden, J. C.; Rotureau, P.; Fayet, G. QSPR prediction of physico-chemical properties for REACH. *SAR QSAR Environ. Res.* **2013,** *24*, 545-584.
(9) Le, T.; Epa, V. C.; Burden, F. R.; Winkler, D. A. Quantitative Structure-Property Relationship Modeling of Diverse Materials Properties. *Chem. Rev.* **2012,** *112*, 2889-2919.
(10) Gharagheizi, F. An accurate model for prediction of autoignition temperature of pure compounds. *J. Hazard. Mater.* **2011,** *189*, 211-221.
(11) Katritzky, A. R.; Petrukhin, R.; Jain, R.; Karelson, M. QSPR Analysis of Flash Points. *J. Chem. Inf. Comput. Sci.* **2001,** *41*, 1521-1530.
(12) Fayet, G.; Rotureau, P.; Joubert, L.; Adamo, C. Development of a QSPR model for predicting thermal stabilities of nitroaromatic compounds taking into account their decomposition mechanisms. *J. Mol. Model.* **2011,** *17*, 2443-2453.
(13) Lu, Y.; Ng, D.; Mannan, M. S. Prediction of the Reactivity Hazards for Organic Peroxides Using the QSPR Approach. *Ind. Eng. Chem. Res.* **2011,** *50*, 1515-1522.
(14) Prana, V.; Rotureau, P.; Fayet, G.; André, D.; Hub, S.; Vicot, P.; Rao, L.; Adamo, C. Prediction of the termal demcomposition of organic peroxides by validated QSPR models. *J. Hazard. Mater.* **2014,** *276*, 216-224.
(15) Rice, B. M.; Hare, J. J. A Quantum Mechanical Investigation of the Relation between Impact Sensitivity and the Charge Distribution in Energetic Molecules. *J. Phys. Chem. A* **2002,** *106*, 1770-1783.
(16) Keshavarz, M. H. Theoretical prediction of electric spark sensitivity of nitroaromatic energetic compounds based on molecular structure. *J. Hazard. Mater.* **2008,** *153*, 201-206.
(17) Bénazet, S.; Jacob, G.; Pèpe, G. GenMolTM Supramolecular Descriptors Predicting Reliable Sensitivity of Energetic Compounds. *Propel. Explos. Pyrotech.* **2009,** *34*, 120-135.
(18) Fayet, G.; Rotureau, P.; Joubert, L.; Adamo, C. Predicting explosibility properties of chemicals from Quantitative Structure-Property Relationships. *Process Saf. Prog.* **2010,** *29*, 359-371.

(19) Fayet, G.; Rotureau, P.; Prana, V.; Adamo, C. Global and local quantitative structure–property relationship models to predict the impact sensitivity of nitro compounds. *Process Saf. Prog.* **2012,** *31*, 291-303.
(20) Muratov, E. N.; Varlamova, E. V.; Artemenko, A. G.; Polishchuk, P. G.; Kuz'min, V. E. Existing and Developing Approaches for QSAR Analysis of Mixtures. *Mol. Inform.* **2012,** *31*, 202-221.
(21) Ajmani, S.; Rogers, S. C.; Barley, M. H.; Livingstone, D. J. Application of QSPR to Mixtures. *J. Chem. Inf. Model.* **2006,** *46*, 2043-2055.
(22) Ajmani, S.; Rogers, S. C.; Barley, M. H.; Burgess, A. N.; Livingstone, D. J. Characterization of Mixtures. Part 2: QSPR Models for Prediction of Excess Molar Volume and Liquid Density Using Neural Networks. *Mol. Inf.* **2010,** *29*, 645-653.
(23) Ajmani, S.; Rogers, S. C.; Barley, M. H.; Burgess, A. N.; Livingstone, D. J. Characterization of Mixtures Part 1: Prediction of Infinite-Dilution Activity Coefficients Using Neural Network-Based QSPR Models. *QSAR Comb. Sci* **2008,** *27*, 1346-1361.
(24) Lyman, W. J.; Reehl, W. F.; Rosenblatt, D. H., *Handbook of Chemical Property Estimation Methods*. American Chemical Society: Washington DC, 1990.
(25) Oprisiu, I.; Varlamova, E.; Muratov, E.; Artemenko, A.; Marcou, G.; Polishchuk, P.; Kuz'min, V.; Varnek, A. QSPR Approach to Predict Nonadditive Properties of Mixtures. Application to Bubble Point Temperatures of Binary Mixtures of Liquids. *Mol. Inf.* **2012,** *31*, 491-502.
(26) Katritzky, A. R.; Stoyanova-Slavova, I. B.; Tämm, K.; Tamm, T.; Karelson, M. Application of the QSPR Approach to the Boiling Points of Azeotropes. *J. Phys. Chem. A* **2011,** *115*, 3475-3479.
(27) Oliferenko, A. A.; Oliferenko, P. V.; Torrecilla, J. S.; Katritzky, A. R. Boiling Points of Ternary Azeotropic Mixtures Modeled with the Use of the Universal Solvation Equation and Neural Networks. *Ind. Eng. Chem. Res.* **2012,** *51*, 9123-9128.
(28) Solov'ev, V. P.; Oprisiu, I.; Marcou, G.; Varnek, A. Quantitative Structure-Property Relationship (QSPR) Modeling of Normal Boiling Point Temperature and Composition of Binary Azeotropes. *Ind. Eng. Chem. Res.* **2011,** *50*, 14162-14167.
(29) Oprisiu, I.; Novotarskyi, S.; Tetko, I. Modeling of non-additive mixture properties using the Online CHEmical database and Modeling environment (OCHEM). *J. Chemometr.* **2013,** *5*, 4.
(30) Babrauskas, V. *Ignition Handbook*. Fire Sciences Publishers: 2003.
(31) Globally harmonized system of classification and labelling of chemicals (GHS). In Third revised edition, ST/SG/AC.A10/30/Rev.3 ed.; United Nations, Ed. 2009.
(32) Regulation (EC) N°1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labelling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) N° 1907/2006. In.
(33) Vidal, M.; Rogers, W. J.; Holste, J. C.; Mannan, M. S. A review of estimation methods for flash points and flammability limits. *Process Saf. Prog.* **2004,** *23*, 47-55.
(34) Liu, X.; Liu, Z. Research Progress on Flash Point Prediction. *J. Chem. Eng. Data* **2010,** *55*, 2943-2950.
(35) Hshieh, F.-Y. Correlation of closed-cup flash points with normal boiling points for silicone and general organic compounds. *Fire Mater.* **1997,** *21*, 277-282.
(36) Catoire, L.; Naudet, V. A Unique Equation to Estimate Flash Points of Selected Pure Liquids Application to the Correction of Probably Erroneous Flash Point Values. *J. Phys. Chem. Ref. Data* **2004,** *33*, 1083.
(37) Carroll, F. A.; Lin, C.-Y.; Quina, F. H. Simple Method to Evaluate and to Predict Flash Points of OrganicCompounds. *Ind. Eng. Chem. Res.* **2011,** *50*, 4796-4800.
(38) Saldana, D. A.; Starck, L.; Mougin, P.; Rousseau, B.; Pidol, L.; Jeuland, N.; Creton, B. Flash Point and Cetane Number Predictions for Fuel Compounds Using Quantitative Structure Property Relationship (QSPR) Methods. *Energy Fuels* **2011,** *25*, 3900-3908.
(39) Rowley, J. R.; Rowley, R. L.; Wilding, W. V. Estimation of the Flash Point of Pure Organic Chemicals from Structural Contributions. *Process Saf. Prog.* **2010,** *29*, 353-358.
(40) Fayet, G.; Rotureau, P.; Prana, V.; Adamo, C. Prediction of Physico-Chemical Properties for REACH Based on QSPR Models. *Chem. Eng. Trans.* **2013,** *31*.

(41) Khajeh, A.; Modarress, H. QSPR prediction of flash point of esters by means of GFA and ANFIS. *J. Hazard. Mater.* **2010,** *179*, 715-720.
(42) Wickey, R. O.; Chittenden, D. H. Flash Point of Blends Correlated. *Pretroleum Refiner* **1963,** *42*, 157-158.
(43) Affens, W. A.; McLaren, G. W. Flammability properties of hydrocarbon solutions in air. *J. Chem. Eng. Data* **1972,** *17*, 482-488.
(44) Walsham, J. G., 5. Prediction of Flash Points for Solvent Mixtures. In *In Solvents Theory and Practice*, Advances in Chemistry, American Chemical Society ed.; Tess, R.: Washington, DC, 1973; pp 56-69.
(45) Thorne, P. F. *Flash Points of Mixtures of Flammable and Non-Flammable Liquids*; 1022; Fire Research Station: Fire Research Station BOREHAMWOOD Hertfordshire WD6 2BL, 1975, 1975.
(46) Gmehling, J.; Rassmussen, P. Flash Point of Flammable Liquid Mixtures Using UNIFAC. *Ind. Eng. Chem. Fundam* **1982,** *21*, 186-188.
(47) Wu, D. T.; Finkelman, R. A mathematical model for the prediction of closed cup flash points. *Org. Coat. Plast. Chem.* **1978,** *38*, 61-67.
(48) Liaw, H.-J.; Lu, W.-H.; Gerbaud, V.; Chen, C.-C. Flash-point prediction for binary partially miscible mixtures of flammable solvents. *J. Hazard. Mater.* **2008,** *153*, 1165-1175.
(49) Liaw, H.-J.; Gerbaud, V.; Li, Y.-H. Prediction of miscible mixtures flash-point from UNIFAC group contribution methods. *Fluid Phase Equilib.* **2011,** *300*, 70-82.
(50) Liaw, H.-J.; Lin, S.-C. Binary mixtures exhibiting maximum flash-point behavior. *J. Hazard. Mater.* **2007,** *140*, 155-164.
(51) Saldana, D. A.; Starck, L.; Mougin, P.; Rousseau, B.; Creton, B. Prediction of Flash Points for Fuel Mixtures Using Machine Learning and a Novel Equation. *Energy Fuels* **2013,** *27*, 3811-3820.
(52) Gaudin, T.; Rotureau, P.; Fayet, G. Combining mixing rules with QSPR models for pure chemicals to predict the flash points of binary organic liquid mixtures. *Fire Safety Journal* **2014,** *74*, 61-70.
(53) Regulation (EC) N° 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH). In.
(54) ASTM D56 - Standard Test Method for Flash Point by Tag Closed Cup Tester. In 2010.
(55) Liaw, H.-J.; Gerbaud, V.; Wu, H.-T. Flash Point Measurements and Modeling for Ternary Partially Miscible Aqueous-Organic Mixtures. *J. Chem. Eng. Data* **2010,** *55*, 3451-3461.
(56) Liaw, H.-J.; Gerbaud, V.; Chao-Yang, C. Flash Point for Ternary Partially Miscible Mixtures of Flammable Solvents. *J. Chem. Eng. Data* **2010,** *55*, 134-146.
(57) Mitu, M.; Brandes, E.; Dieter, P. Lower Explosion Limits, Lower Explosion Points and Flash Points of (1-octanol + n-amylacetate) Mixtures. *Revista de Chimie* **2006,** *57*, 770-772.
(58) Noorollahy, M.; Moghadam, A. Z.; Ghasrodashti, A. A. Calculation of mixture equilibrium binary interaction parameters using closed cup flash point measurements. *Chem. Eng. Res. Des.* **2010,** *88*, 81-86.
(59) Ha, D.-M.; Lee, S. Prediction and measurement of the lower flash points for flammable binary solutions by using a setaflash closed cup tester. *Korean J. Chem. Eng.* **2011,** *28*, 1161-1165.
(60) Moghaddam, A. Z.; Rafiei, A.; Khalili, T. Assessing prediction models on calculating the flash point of organic acid, ketone and alcohol mixtures. *Fluid Phase Equilib.* **2012,** *316*, 117-121.
(61) Donaldson, A. B. Flash Point For Binary Mixtures. *Eng. J. Qatar Univ.* **1994,** *7*, 27-36.
(62) Hristova, M. Measurement and prediction of binary mixture flash point. *Cent. Eur. J. Chem.* **2013,** *11*, 57-62.
(63) Hristova, M.; Damgaliev, D. Flash point of organic binary mixtures containing alcohols: experiment and prediction. *Cent. Eur. J. Chem.* **2013,** *11*, 388-393.
(64) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi,

M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian09*, Gaussian Inc.: Wallington CT, 2009.
(65) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993,** *98*, 5648-5652.
(66) Tirado-Rives, J.; Jorgensen, W. L. Performance of B3LYP Density Functional Methods for a Large Set of Organic Molecules. *Journal of Chemical Theory and Computation* **2008,** *4*, 297-306.
(67) Crounse, J. D.; Nielsen, L. B.; Jörgensen, S.; Kjaergaard, H. G.; Wennberg, P. O. Autoxidation of Organic Compounds in the Atmosphere. *The Journal of Physical Chemistry Letters* **2013,** *4*, 3513-3520.
(68) Di Tommaso, S.; Rotureau, P.; Benaissa, W.; Gruez, P.; Adamo, C. Theoretical and Experimental Study on the Inhibition of Diethyl Ether Oxidation. *Energy Fuels* **2014,** *28*, 2821-2829.
(69) *CodessaPro*, University of Florida: 2002.
(70) Karelson, M. *Molecular Descriptors in QSAR/QSPR*. Wiley: New York, 2000.
(71) Geerlings, P.; De Proft, F.; Langenaeker, W. Conceptual Density Functional Theory. *Chem. Rev.* **2003,** *103*, 1793-1874.
(72) Chermette, H. Chemical reactivity indexes in density functional theory. *J. Comput. Chem.* **1999,** *20*, 129-154.
(73) Fayet, G.; Joubert, L.; Rotureau, P.; Adamo, C. On the use of descriptors arising from the conceptual density functional theory for the prediction of chemicals explosibility. *Chem. Phys. Lett.* **2009,** *467*, 407-411.
(74) Fayet, G.; Rotureau, P.; Joubert, L.; Adamo, C. On the prediction of thermal stability of nitroaromatic compounds using quantum chemical calculations. *J. Hazard. Mater.* **2009,** *171*, 845-850.
(75) Kay, W. B. Density of Hydrocarbon. *Ind. Eng. Chem.* **1936,** *28*, 1014-1019.
(76) Jover, J.; Bosque, R.; Sales, J. Neural Network Based QSPR Study for Predicting pKa of Phenols in Different Solvents. *J. Comb. Sci.* **2007**, 385-397.
(77) Tropsha, A. Best Practices for QSAR Model Development, Validation, and Exploitation. *Mol. Inform.* **2010,** *29*, 476-488.
(78) Muratov, E. N.; Varlamova, E. V.; Artamenko, A. G.; Polishchuk, P. G.; Kuz'min, V. E. Existing and Developing Approaches for QSAR Analysis of Mixtures. *Mol. Inf.* **2012,** *31*, 202-221.
(79) Oprisiu, I. Modélisation QSPR de mélanges binaires non-additifs. Application au comportement azéotropique. phD Thesis, University of Strasbourg, 2012.
(80) Ajmani, S.; Rogers, S. C.; Barley, M. H.; Livingstone, D. J. Application of QSPR to Mixtures. *J. Chem. Inf. Model.* **2006,** *46*, 2043-2055.
(81) Gharagheizi, F.; Eslamimanesh, A.; Mohammadi, A. H.; Richon, D. Empirical Method for Representing the Flash-Point Temperature of Pure Compounds. *Ind. Eng. Chem. Res.* **2011,** *50*, 5877-5880.
(82) Randic, M. The connectivity index 25 years after. *J. Mol. Graphics Modell.* **2001,** *20*, 19-35.

**Table 1 –** Composition of the training and validation sets.

| Set | Number of data | Number of mixtures | Number of pure compounds |
|---|---|---|---|
| Entire dataset | 435 | 43 | 34 |
| Training set | 284 | 22 | 21 |
| Validation set | 151 | 21 | 21 |

**Table 2 –** Models allowing reaching $MAE_{EXT}$ < 10°C for the prediction of FP for various types of mixtures.

| Type of mixtures | $MAE_{EXT}$ < 10°C |
|---|---|
| Alcohol / Alcohol | **sqr_fmol_sum**, **root_fmol**, **cent** |
| Alcohol / Carbonyl | **fmol_sum**, **root_fmol**, **sqr_fmol_sum**, **cent** |
| Alcohol / Ester | **root_fmol**, **sqr_fmol_sum**, **abs_diff** |
| Alcohol / Hydrocarbon | **fmol_sum**, **norm_cont**, **sqr_fmol** |
| Hydrocarbon / Ester | **fmol_sum**, **norm_cont**, **cent** |
| Hydrocarbon / Hydrocarbon | - |

**Table 3 –** Experimental vs. calculated FP (in °C) for pure n-decane and n-dodecane with the various formulas.

| | exp | fmol_sum | fmol_diff | sqr_fmol | root_fmol | sqr_fmol_sum | norm_cont |
|---|---|---|---|---|---|---|---|
| **n-decane** | 51.8 | 16.3 | 16.0 | 18.3 | 24.2 | 35.1 | 17.4 |
| **n-dodecane** | 84.0 | 42.2 | 27.7 | 33.2 | 18.6 | 52.8 | 25.1 |

**Table 4 –** Performances of the present QSPR model and the combined approach developed in previous work[52] for the 151 data of the validation set.

| | **Mixture QSPR (Eq. 18)** | **Combined approach [52]** |
|---|---|---|
| **$R^2_{EXT}$** | 0.71 | 0.96 |
| **$MAE_{EXT}$ (°C)** | 10.3 | 4.7 |

**Figure 1 –** Distributions of the datasets in flash point.

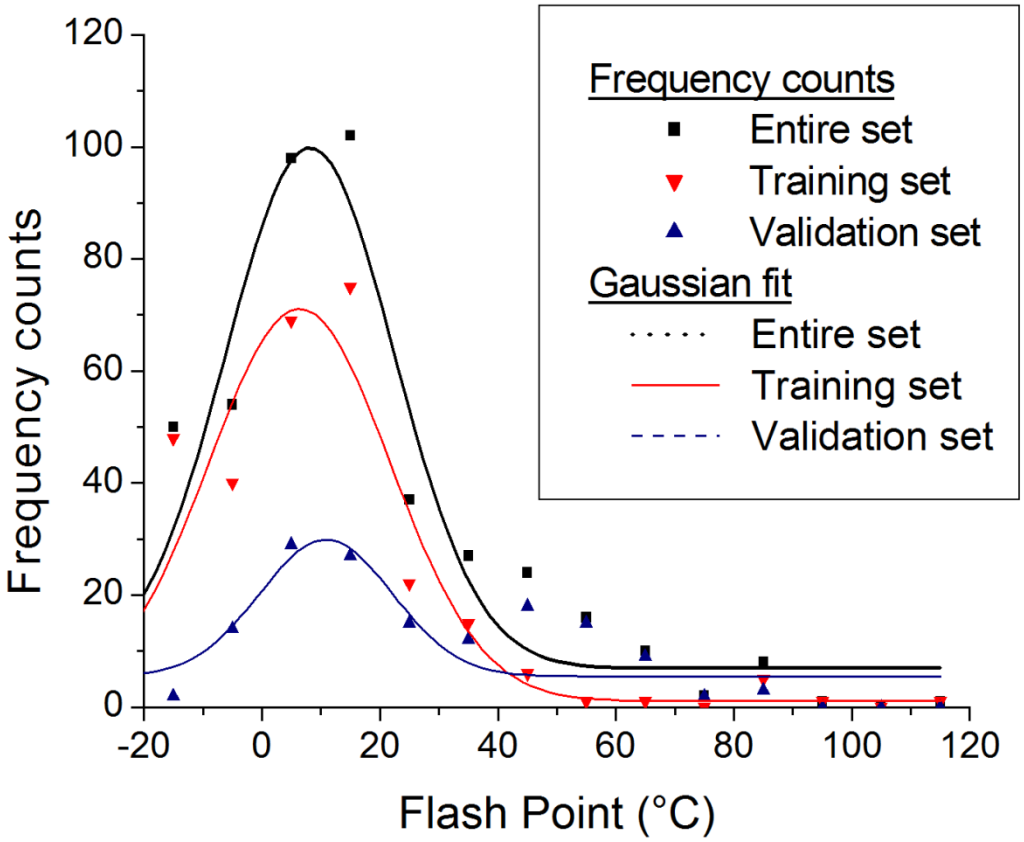

**Figure 2 -** Distribution of the deviations of experimental flash points of mixtures to linear behavior with respect to concentration.

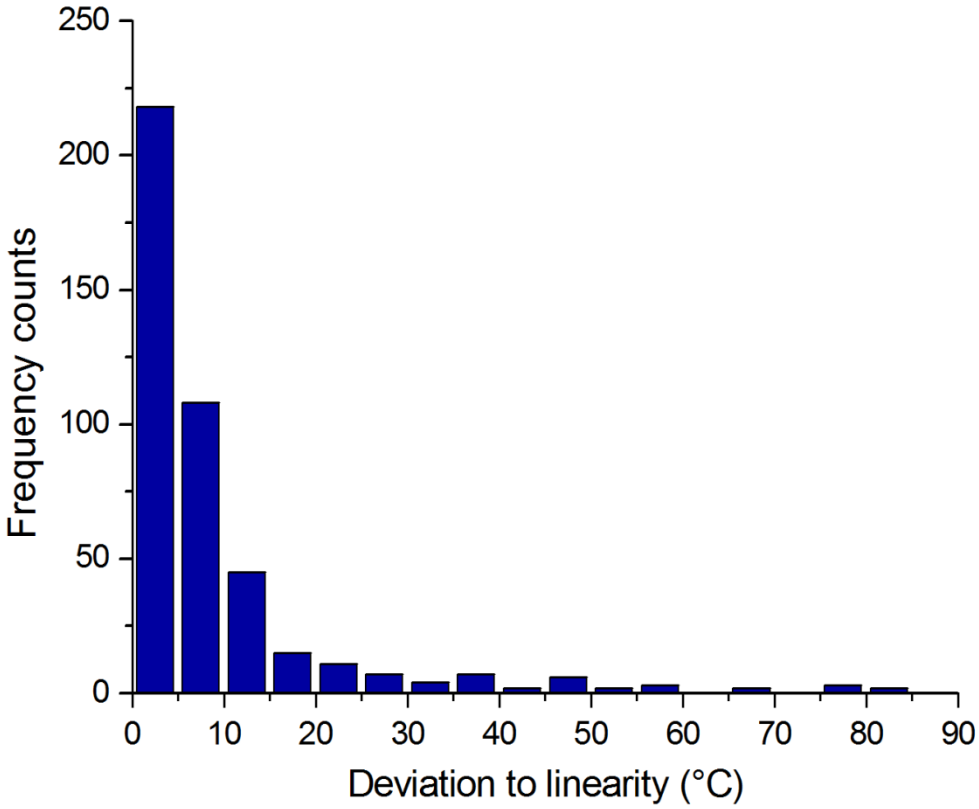

**Figure 3 –** Performances of the QSPR models for the flash point of mixtures upon the type of mixture descriptors.

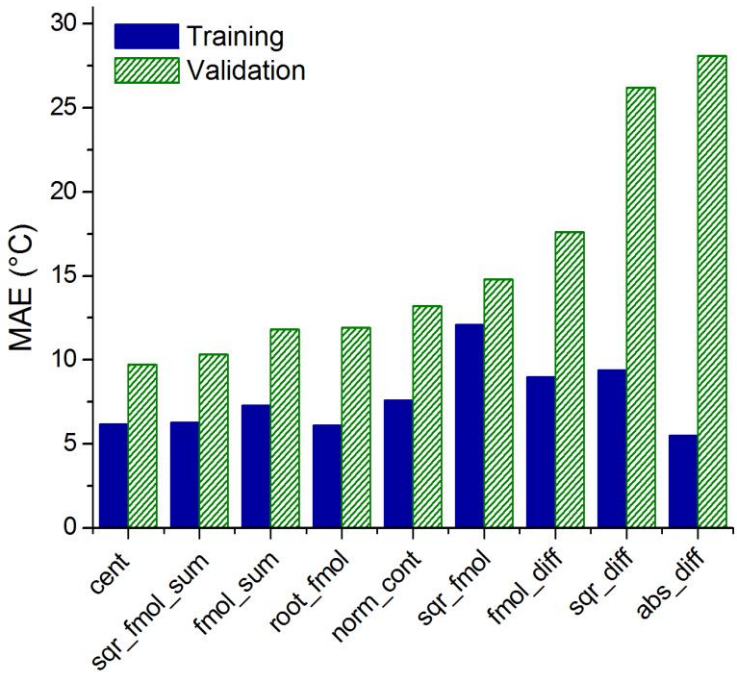

**Figure 4 –** Experimental[57] and calculated flash point profiles of 4-methyl-2-pentanone/1-butanol mixtures.

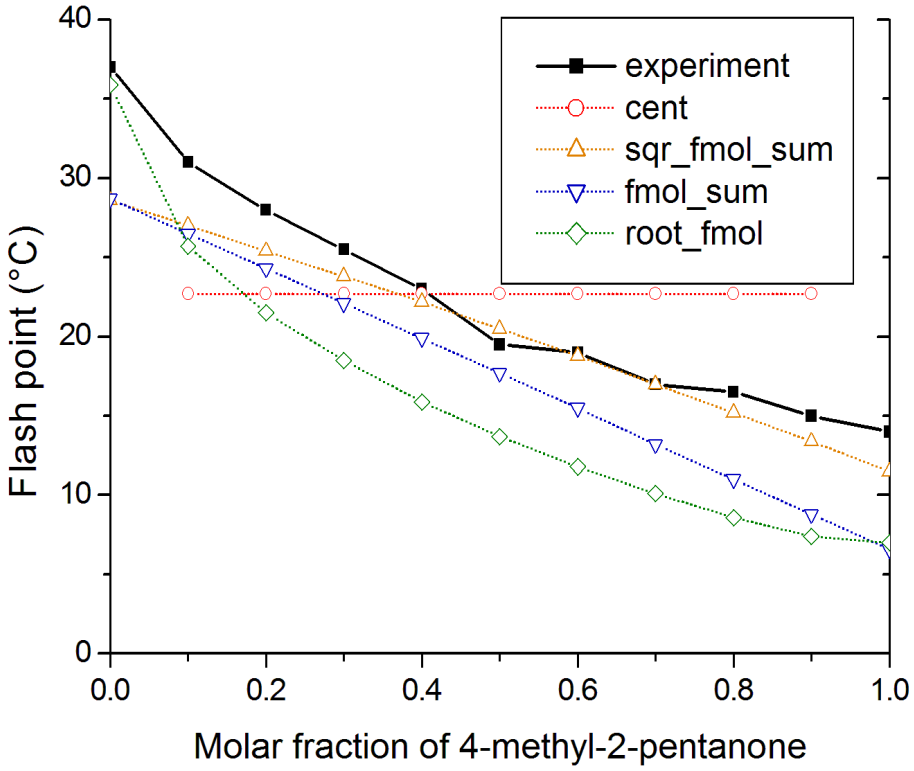

**Figure 5 –** Experimental[58] and calculated flash point profiles of 1-propanol/octane mixtures.

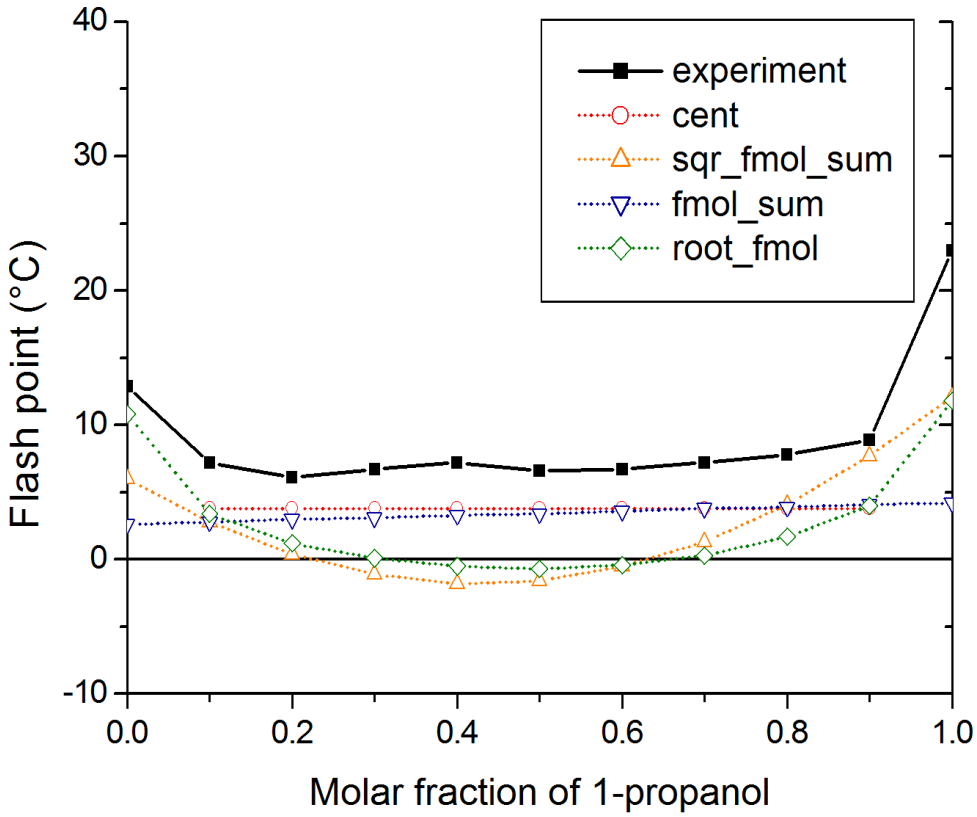

**Figure 6 –** Experimental[58] and calculated flash point profiles of n-octane/n-decane mixtures

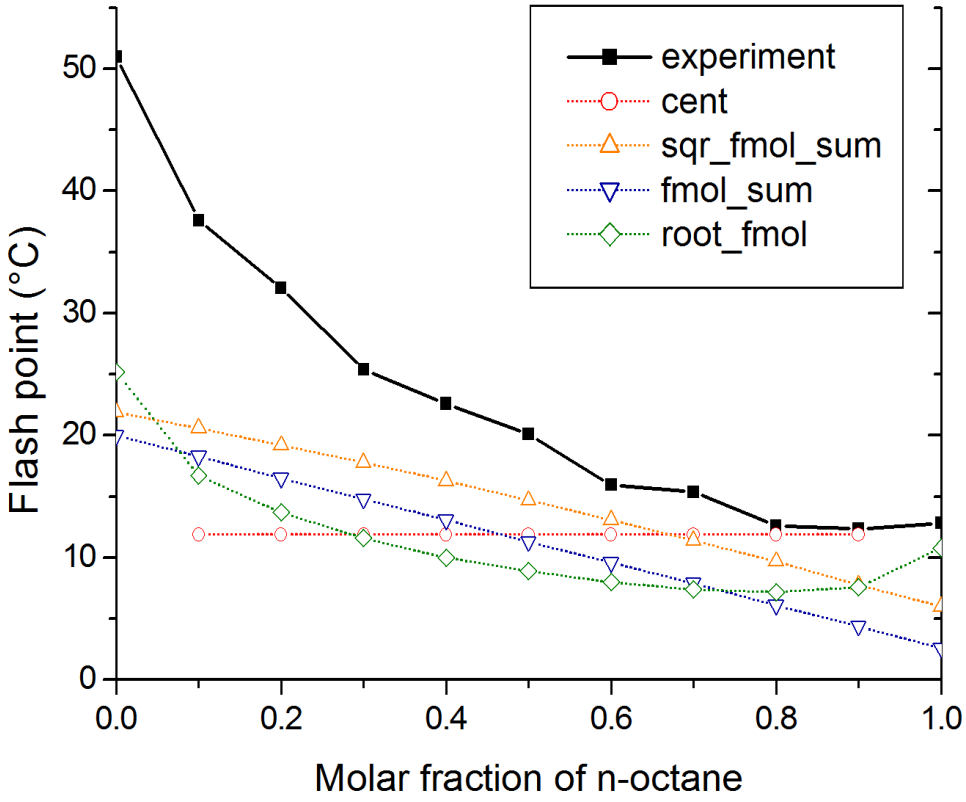

**Figure 7 –** Performances of the new QSPR models issued from the best three combinations for the flash point of pure compounds.

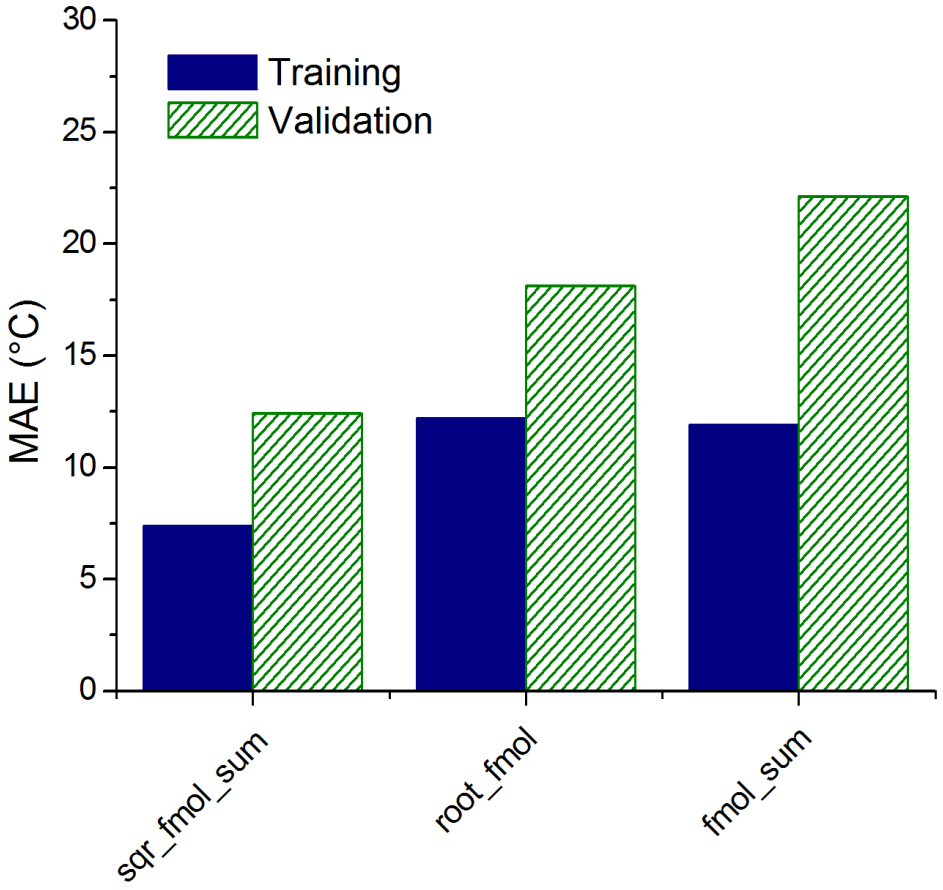

**Figure 8 –** Experimental vs. calculated flash points for 43 mixtures (435 data points) using the best model obtained with the **sqr_fmol_sum** formula (eq. 18)