



Article

Prediction of the Auto-Ignition Temperatures of Binary Miscible Liquid Mixtures from Molecular Structures

Shijing Shen, Yong Pan *, Xianke Ji, Yuqing Ni and Juncheng Jiang

Jiangsu Key Laboratory of Hazardous Chemicals Safety and Control, College of Safety Science and Engineering, Nanjing Tech University, Nanjing 210009, China; ShijingShen_njtech@163.com (S.S.); 1090997173@njtech.edu.cn (X.J.); nyqforwork@outlook.com (Y.N.); ypnjut@126.com (J.J.)

* Correspondence: yongpan@njtech.edu.cn; Tel.: +86-25-58139873

Received: 23 March 2019; Accepted: 23 April 2019; Published: 27 April 2019



Abstract: A quantitative structure-property relationship (QSPR) study is performed to predict the auto-ignition temperatures (AITs) of binary liquid mixtures based on their molecular structures. The Simplex Representation of Molecular Structure (SiRMS) methodology was employed to describe the structure characteristics of a series of 132 binary miscible liquid mixtures. The most rigorous “compounds out” strategy was employed to divide the dataset into the training set and test set. The genetic algorithm (GA) combined with multiple linear regression (MLR) was used to select the best subset of SiRMS descriptors, which significantly contributes to the AITs of binary liquid mixtures. The result is a multilinear model with six parameters. Various strategies were employed to validate the developed model, and the results showed that the model has satisfactory robustness and predictivity. Furthermore, the applicability domain (AD) of the model was defined. The developed model could be considered as a new way to reliably predict the AITs of existing or new binary miscible liquid mixtures, belonging to its AD.

Keywords: quantitative structure-property relationship (QSPR); auto-ignition temperature (AIT); simplex representation of molecular structure (SiRMS); binary miscible liquid mixtures

1. Introduction

The auto-ignition temperature (AIT) is defined as the lowest temperature at which the substance spontaneously ignites in ambient air, without an external ignition source, such as a spark or flame. AIT is one of the most important parameters applied to classify the chemicals based on their degree of flammability [1]. The experimental AIT values are the main source of the AIT data used in production. However, the measurement of AITs is expensive and time-consuming. Especially for the mixtures, the measurement is more difficult, since the AITs of the mixtures are closely related to their compositions and ratios, which are rather difficult to test one-by-one. Therefore, it is of great significance to develop theoretical models for predicting the AITs of mixtures.

Many theoretical models for predicting the AITs of pure flammable liquids have been proposed [2–5]. However, only a few efforts have been made to predict the AITs of mixtures. Rota et al. [6] developed a kinetic model to predict the AITs of 46 gas mixtures, including ammonia, hydrogen, methane, and air at high pressure and temperature. The average relative error (ARE) and the average absolute error (AAE) of the proposed model were about 3.5% and 25 K, respectively. Peper et al. [7] proposed a simple weighting function formula to predict the AITs of polyurethane raw material mixtures. However, the results showed that the calculated values of AITs of five mixtures are 20 K higher than the measured values. Lan et al. [8] presented a zero-dimensional model to predict

the AITs of 24 binary miscible liquid mixtures. Due to the lack of a chemical kinetic mechanism, the predictive ability of the model for higher hydrocarbons is poor.

The quantitative structure-property relationship (QSPR) method, as a mathematical method, relates the properties of interest to the molecular structures of chemicals. It can be expected to capture the relationships between the molecular structures and desired properties without detailed knowledge of the mechanisms of interaction. In addition, QSPR is considered to be a time-saving and effective method for prediction of the desired properties. In recent years, several QSPR models have been successfully developed to predict the physicochemical properties of mixtures, such as toxicity, boiling point, flash point, and critical parameters, all of which showed satisfactory stability and predictivity [9–13].

The most challenging problem in QSPR studies for mixtures is the representation of structure characteristics of mixtures. There are several different descriptor types for mixtures reported in the literature: descriptors based on the partition coefficient for a mixture, integral additive descriptors, integral non-additive descriptors of mixtures, and fragment non-additive descriptors [14]. As one of the typical fragment non-additive descriptors, Simplex Representation of Molecular Structure (SiRMS) descriptors can be theoretically applied to any investigated activity or property, and could capture the interaction or joint effect of components. Recently, SiRMS descriptors have been successfully employed in QSPR studies for mixtures [15–17].

In this work, for the first time, the QSPR method is applied to study the quantitative relationships between the molecular structures and AITs of binary miscible liquid mixtures. The main purpose of this study is to develop a new method for predicting the AITs of binary miscible liquid mixtures, including: (i) development of SiRMS descriptors for mixtures; (ii) establishment of a QSPR model for the AITs of binary miscible liquid mixtures; (iii) rigorous internal and external model validations; and (iv) definition of the model applicability domain (AD).

2. Results and Discussion

2.1. Results of Prediction

According to the “Compounds out” strategy, the dataset is divided into a training set with 99 mixtures and a test set with 33 mixtures. By performing the GA-MLR procedure on the training set, starting with the calculated 434 simplex descriptors, a best subset of six descriptors was obtained. The definitions and types of these selected descriptors are shown in Table 1. The corresponding best model is presented as follows:

$$\begin{aligned} \text{AIT} &= 700.630 + 36.735X_1 - 56.130X_2 + 70.943X_3 + 52.446X_4 - 111.781X_5 - 92.718X_6 \\ \text{range} &: 496.15 \text{ K} \leq \text{AIT} \leq 798.15 \text{ K} \\ R^2 &= 0.958, Q^2_{\text{LOO}} = 0.950, s = 15.411, F = 345.869, n = 99 \end{aligned} \quad (1)$$

where n is the number of mixtures in the training set, s is the standard error of the model, and F is the Fischer F-ratio.

Table 1. Descriptors selected in the presented model for prediction of the Auto-ignition Temperature (AIT).

Symbol	Descriptor	Definition	Type	Mixing Rule	ME Value
X ₁	S n 4 CHARGE A.A-A-B			$x_1D_1 + x_2D_2$	-66.821
X ₂	S n 4 REFRACTIVITY B-B-B-B			$x_1D_1 + x_2D_2$	155.161
X ₃	S n 4 elm C-C(-C)=O			$x_1D_1 + x_2D_2$	-54.633
X ₄	S n 4 elm C-C(-O)=O			$x_1D_1 + x_2D_2$	-14.773

Table 1. Cont.

Symbol	Descriptor	Definition	Type	Mixing Rule	ME Value
X ₅	M n 4 CHARGE A-A.B-C			2x ₁ D ₁₊₂	21.835
X ₆	M n 4 REFRACTIVITY B-B.B-C			2x ₁ D ₁₊₂	59.231

Moreover, the relative significance and contribution of each descriptor on the AIT were determined by the mean effect (ME) analysis, which is calculated as follows:

$$ME_j = \frac{\beta_j \sum_{i=1}^{i=n} d_{ij}}{\sum_j^m \beta_j \sum_j^n d_{ij}} \quad (2)$$

where ME_j represents the mean effect for the descriptor *j*, β_j is the coefficient of the descriptor *j*, d_{ij} is the value of the descriptors of interest for each mixture, *m* is the number of descriptors in the model, and *n* is the number of dataset members. The symbol (positive or negative) of ME represents the trend of the impact of each descriptor on the AIT. The greater the absolute value of the coefficient is, the more important the descriptor is.

As can be concluded from Table 2, the |S|n|||4||REFRACTIVITY|B-B-B-B descriptor has the greatest influence on AIT. In addition, the relative importance and contribution of each descriptor in the model was determined and ranked as follows based on the ME values: |S|n|||4||REFRACTIVITY|B-B-B-B > |S|n|||4||CHARGE|A.A-A-B > |M|n|||4||REFRACTIVITY|B-B.B-C > |S|n|||4||elm|C-C(-C)=O > |M|n|||4||CHARGE|A-A.B-C > |S|n|||4||elm|C-C(-O)=O.

Table 2. The main statistical parameters of the obtained Multiple Linear Regression (MLR) model.

Statistical Parameters	Training Set	Test Set
R ²	0.958	0.942
Q ² _{LOO}	0.950	-
Q ² _{EXT}	-	0.942
RMSE	15.333	15.740
AAE	12.395	12.531
ARE	1.9%	1.8%
<i>n</i>	99	33

The developed model was then employed to predict the AIT values of mixtures in the test set for external validation. The predicted AIT values are presented in the Supplementary Table S1. The main statistical parameters of the model are presented in Table 2. As can be seen from Table 3, the AAE and RMSE values were as low as possible, which indicated that the presented model has acceptable predictive capability. A plot of the predicted AIT values versus the observed ones for both the training and test sets is presented in Figure 1. Thus, this showed a reasonable agreement between the predicted and observed AIT values across the whole dataset. The predicted percentage error of all the 132 mixtures was also calculated, which is shown in Figure 2. The obtained average percentage error for these mixtures was 1.8% and the maximum percentage error was 7.9%.

Table 3. Basic types of simplexes.

Basic Type	1	2	3	4	5	6	7	8	9	10	11
simplex											

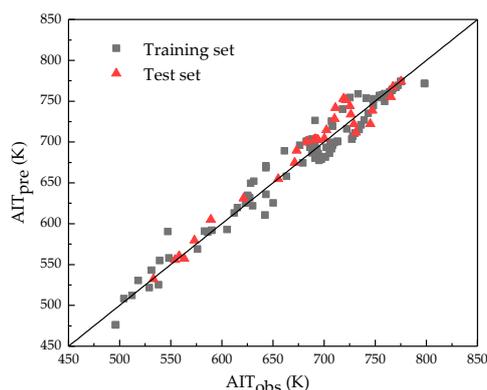


Figure 1. Correlation between the predicted and observed AIT values for both the training and test sets.

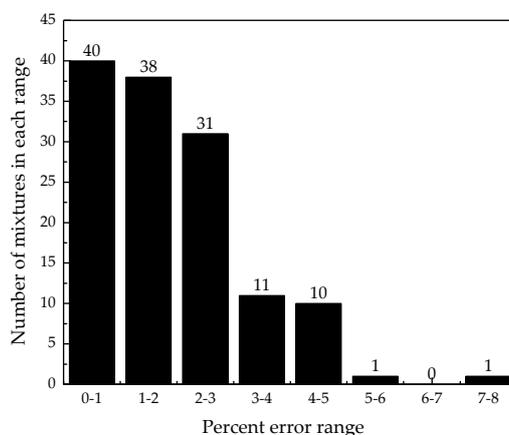


Figure 2. The percent errors obtained by the presented model and the number of mixtures in each range.

2.2. Model Stability Validation and Results Analysis

In this study, the Y-randomization test was performed on the training set 100 times. The obtained R^2 of randomization versus the frequency of occurrence of the randomized models are presented in Figure 3. The resulting maximum, minimum, and average values of the achieved highest random R^2 were 0.173, 0.004, and 0.055, respectively, while the value of SD was 0.035. The difference between the R^2 of the original MLR model and the mhr R^2 is higher than 3 SD. It can be concluded that there is no chance correlation in the proposed model.

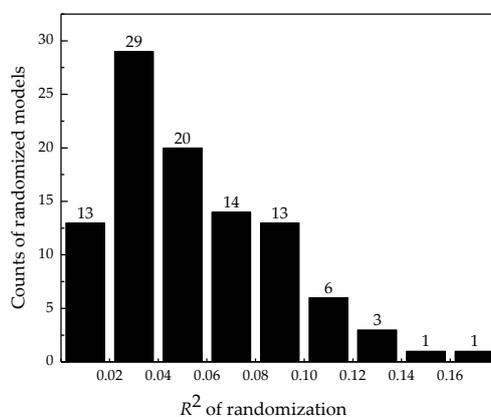


Figure 3. Histogram of R^2 of randomization versus frequency of occurrence of the randomized models.

The predicted residuals and observed values for the developed model are shown in Figure 4. It can be seen that the calculated residuals are randomly distributed on both sides of the zero baseline, which demonstrates that no systematic errors exist in the proposed model.

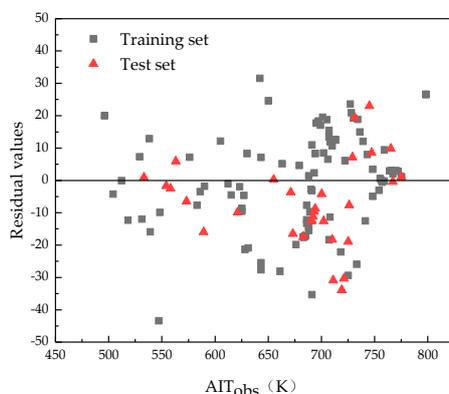


Figure 4. Plot of the residuals versus the observed AIT values for the MLR model.

From all of the above validation results, it can reasonably be concluded that the proposed MLR model has satisfactory robustness and predictability. So, it can be reliably and conveniently employed to predict the AITs of binary miscible liquid mixtures, solely from their molecular structures and mole fractions.

2.3. Applicability Domain of the Proposed Model

A Williams plot for the proposed QSPR model is shown in Figure 5. The AD is established inside a squared area within ± 3 standard deviations and a leverage threshold h^* of 0.212. In Figure 5, there are three possible outliers (namely, #63, #78, and #110) in the dataset with higher leverage values ($h > h^*$). The structures of these mixtures are obviously different from the others. However, their AITs can still be satisfactorily predicted by the present model within the standard deviation. Thus, the predictions are considered to be acceptable. Therefore, the developed model can be expected to reliably predict the AITs for the binary miscible liquid mixtures falling within the corresponding applicability ranges. However, it should be stated that there is also a limitation to the AD of the model in terms of chemical diversity, since the studied dataset only contained 10 different pure compounds. However, it is rather difficult to find a further larger set of AIT data for binary mixtures in the open literature that contains more and different pure compounds.

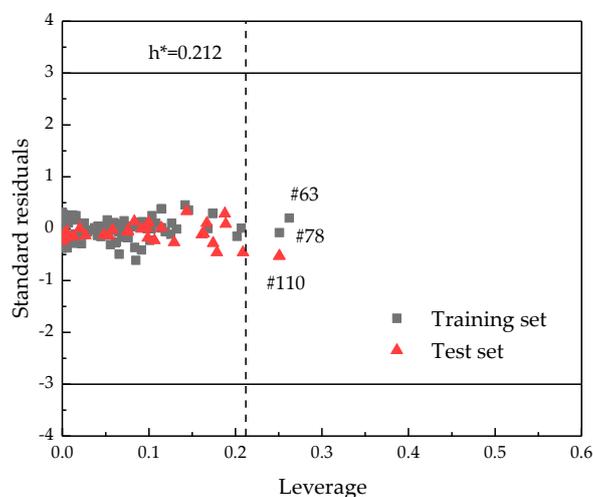


Figure 5. A Williams plot describing the applicability domain of the Quantitative Structure-Property Relationship (QSPR) model ($h^* = 0.212$).

3. Materials and Methods

3.1. Dataset

The dataset consists of 132 binary miscible liquid mixtures and originates from Lan et al.'s work [8], the detail of which can be found in the Supplementary Table S1. The pure compound components include alcohols, acids, esters, benzenes, ketones, and alkanes. All of the AIT values were obtained by experimental tests according to the ASTM E659-78 test standard (American Society for Testing and Materials). The AIT values of the whole dataset range from 496.15 K to 798.15 K. As is well-known, with a larger dataset, a better predictive model could be developed; however, it is rather difficult to find a larger set of AIT data for binary mixtures in the open literature in terms of chemical diversity.

3.2. Descriptor Calculation and Reduction

An important step in a QSPR study is the characterization of the molecular structures. In this study, the binary mixtures were represented by a variety of SiRMS descriptors. In the framework of SiRMS, any molecule can be represented as a system of different fragments (simplexes) of fixed composition, structure, chirality, and symmetry simplexes [18]. All of the possible topological structure types of simplexes are shown in Table 3.

Bounded and unbounded two-dimensional (2D) simplexes were used. Bounded simplexes were used to describe pure compounds, while unbounded simplexes can describe both the pure compounds and the mixtures. Thus, during descriptor generation, a special mark is used to distinguish them. The details of the procedure for calculation of 2D simplex descriptors for mixtures in this study are as follows. Firstly, the 2D chemical structures of each pure substance were drawn in MarvinSketch (version 15.6.29.0, ChemAxon, Budapest, Hungary) [19], and optimized based on the "clean in 2D" method by this software. For binary mixtures, the program generated the simplexes of individual species and mixture simplexes with atoms from two compounds. Then, each atom of the fragment obtained a calculated value by the *cxcalc* tool [19] and the atoms were divided into the corresponding groups: (i) partial charge $A \leq -0.05 < B \leq 0 < C \leq 0.05 < D$; (ii) lipophilicity $A \leq -0.5 < B \leq 0 < C \leq 0.5 < D$; and (iii) refraction $A \leq 1.5 < B \leq 3 < C \leq 8 < D$. Three characteristics of atom H-bond formation ability were specified: A (acceptor of hydrogen in H-bond), D (donor of hydrogen in H-bond), and I (indifferent atom). In this work, fragments with four atoms were considered to reduce the probability of the model over-fitting and ensure its predictivity and AD [20]. The described SiRMS descriptors can be implemented in the open-source software (version 1.1.2, GitHub, San Francisco, California, America) [21] written on Python 3, which is available on the Github repository.

Descriptors of constituent parts (compounds 1 and 2) are weighted according to their molar fraction, which was calculated as follows:

$$D_s = x_1 D_1 + x_2 D_2. \quad (3)$$

Meanwhile, mixture descriptors are multiplied on the doubled minimal weight according to Equation (4).

$$D_M = 2x_1 D_{1+2} \quad (4)$$

where x_1 and x_2 are molar fractions of compounds 1 and 2 ($x_1 < x_2$ and $x_1 + x_2 = 1$), respectively, and D_1 , D_2 , and D_{1+2} are descriptor values for individual compounds 1 and 2, and for their mixtures, respectively. Furthermore, the volume ratio obtained from the literature [8] needs to be converted to a molar ratio first, since the calculation rules are based on the molar ratio.

A concatenation of D_s and D_M represents the mixture descriptors of the whole dataset. Finally, a total set of 434 simplex descriptors was achieved.

3.3. Descriptor Selection and Model Development

The key step in QSPR modeling is to find the optimal descriptors that make a significant contribution to the AITs of binary miscible liquid mixtures. The well-known genetic algorithm (GA) is a powerful optimization method to solve this problem and has been successfully applied to feature selection in previous QSPR studies [22–24]. In this study, genetic algorithm along with multiple linear regression (GA-MLR) was used to find the optimal subset that accurately represented the relationships between molecular structures and AITs of binary liquid mixtures. The GA-MLR was performed by the MATLAB M-file written in our laboratory. The fitness function of this method corresponds to the root mean square error of cross-validation (rmsecv).

The selection program is started with one descriptor, and the best one-parameter regression model, with the minimal rmsecv value, should be obtained. Then, the number of desired variables should be increased to two, three, four, etc. and the corresponding best multi-parameter regression models with the desired number of descriptors should be found. When the number of descriptors was increased and the rmsecv did not significantly improve, it can be determined that the optimum subset of descriptors that produce the best MLR model has been achieved [25].

3.4. Model Validation

Model validation is a necessary step to ensure the reliability of the developed QSPR models. In this study, both internal and external validation methods were employed to validate the developed QSPR model.

Cross-validation (CV) is one of the most common methods for internal validation. A good CV result often represents a good robustness and high internal predictive capability of QSPR models. In this study, leave-one-out (LOO) cross-validation (Q^2_{LOO}) was employed, which is calculated with the following equation:

$$Q^2_{\text{LOO}} = 1 - \frac{\sum_{i=1}^{\text{training}} (y_i - y_0)^2}{\sum_{i=1}^{\text{training}} (y_i - \bar{y})^2} \quad (5)$$

where y_i , y_0 , and \bar{y} are respectively the observed, predicted, and mean observed AIT values of the mixtures in the training set.

External validation is significant and necessary to determine both the predictive capability and generalizability of a developed model for new mixtures. There are three widely used strategies, including “Points out”, “Mixtures out”, and “Compounds out” for dataset partition. Among these three strategies, the “Compounds out” strategy is the most rigorous one and it will fully reflect the ability of models to predict mixtures with a new compound [14,17]. Thus, in this study, the external validation was carried out by randomly splitting the available dataset into a training set (75% of the dataset), and an external test set (25% of the dataset) based on the “Compounds out” partition strategy. The training set is used for descriptor selection and model development, while the test set is used for model validation. The predictive capability of a QSPR model can be judged by an external Q^2_{EXT} , which is defined as follows:

$$Q^2_{\text{EXT}} = 1 - \frac{\sum_{i=1}^{\text{test}} (y_i - y_0)^2}{\sum_{i=1}^{\text{test}} (y_i - \bar{y}_{\text{tr}})^2} \quad (6)$$

where y_i and y_0 are the observed and predicted AIT values of the mixtures in the test set, respectively, and \bar{y}_{tr} is the mean observed AIT values of the mixtures in the training set.

Additionally, a Y-randomization test was employed to further ensure the robustness of the model. The dependent-variable vector (Y vector) was scrambled randomly, while all independent data variables were unchanged, and the robustness of the developed model was tested. The process was repeated 50–100 times. In each model, the highest R^2 value obtained by descriptor selection is recorded as the highest random R^2 of randomization. In addition, the mean highest random (mhr) R^2 and its standard deviation (SD) were calculated by averaging over the repetitions. If all R^2 values of

the randomized models are lower than that of the original model, and the difference between R^2 of the original model and $mhr R^2$ is higher than 2.3 SD for significance at the 1% level, then higher than 3 SD for the 0.1% level. It can be concluded that there is no chance correlation in the model development, and the model can be considered as an acceptable model [26].

The squared correlation coefficient (R^2) is used to determine the calibration capability of the model. The average absolute error (AAE) and root mean square error (RMSE) were employed to evaluate the predictive capability of the developed models, which are calculated as follows:

$$AAE = \frac{\sum_{i=1}^n |y_i - y_0|}{n} \quad (7)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - y_0)^2}{n}} \quad (8)$$

where y_i is the observed value, y_0 is the predicted value, and n is the number of mixtures in the dataset.

3.5. Applicability Domain

According to Organization for Economic Cooperation and Development (OECD) principle 3 [27], the AD should be defined once a QSPR model is obtained. The AD of the model is a theoretical region of the chemical space, which is defined by the descriptors and modeled response. Statistical models can provide reliable predictions for the mixtures in this region. If all of the AIT values are within the AD range, it can be considered that the model is reliable. In this study, the Williams plot was depicted to analyze the AD.

For the x -axis, the leverage value (h_i) describes the impacts of the objects on the model, which is defined as:

$$h_i = x_i (X^T X)^{-1} x_i^T, \quad (9)$$

where x_i is the descriptor column-vector of the considered mixtures and X is the descriptor matrix derived from the training set descriptor values. The warning leverage value (h^*) is calculated as follows:

$$h^* = \frac{3(p+1)}{n} \quad (10)$$

where p is the number of model parameters and n is the number of training mixtures.

If the h_i of a mixture is greater than the h^* , it can be considered as outside of the AD range of the model. For the y -axis, a Williams plot presented the Euclidean distances of the mixtures to the model measured by the cross-validated standardized residuals. The mixture is classified as an outlier when the cross-validated standardized residual is greater than 3 standard deviation units.

4. Conclusions

In this work, for the first time, a QSPR model has been developed for predicting the AITs of binary miscible liquid mixtures from the molecular structures. To the best of our knowledge, the largest existing database of AITs for binary mixtures was employed for modeling. The most rigorous "compounds out" method was used to divide the training set and the test set. The SiRMS methodology was employed to describe the structure characteristics of binary mixtures. The best-resulted QSPR model was a six-parameter linear equation. The model validation results showed the satisfactory robustness and predictivity of the model. The developed model would be expected to provide a new way to reliably predict the AIT values of existing or new binary miscible liquid mixtures, belonging to their AD. Furthermore, the method provides some guidance for prioritizing the design of safer liquid mixtures with desired properties.

Supplementary Materials: Supplementary materials can be found at <http://www.mdpi.com/1422-0067/20/9/2084/s1>. Table S1: A complete list of the compositions of the 132 binary miscible liquid mixtures and their predicted and observed AIT values, as well as the values of the six employed SiRMS descriptors in the model.

Author Contributions: Data curation, S.S.; Formal analysis, S.S. and X.J.; Funding acquisition, Y.P.; Methodology, S.S. and Y.P.; Software, S.S. and X.J.; Validation, Y.N.; Writing—original draft, S.S. and Y.P.; Writing—review & editing, J.J. and Y.P.

Acknowledgments: This research was supported by the National Natural Science Fund of China (No. 21576136, 21436006) and the National Program on Key Basic Research Project of China (2017YFC0804801, 2016YFC0801502). Yong Pan acknowledges the sponsorship of the Qing Lan Project.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

QSPR	Quantitative structure-property relationship
AIT	Auto-ignition temperature
SiRMS	Simplex representation of molecular structure
GA	Genetic algorithm
MLR	Multiple linear regression
AD	Applicability domain
ARE	Average relative error
AAE	Average absolute error
GA-MLR	Genetic algorithm along with multiple linear regression
rmsecv	root mean square error of cross-validation
CV	Cross-validation
Q^2_{LOO}	Leave-one-out cross-validation
mhr	mean highest random
RMSE	Root mean square error
hi	Leverage value
h^*	Warning leverage value
ME	Mean effect
SD	Standard deviation

References

1. Bagheri, M.; Borhani, T.N.G.; Zahedi, G. Estimation of flash point and autoignition temperature of organic sulfur chemicals. *Energy Convers. Manag.* **2012**, *58*, 185–196. [[CrossRef](#)]
2. Pan, Y.; Jiang, J.C.; Wang, R.; Cao, H.Y.; Zhao, J.B. Prediction of auto-ignition temperatures of hydrocarbons by neural network based on atom-type electrotopological-state indices. *J. Hazard. Mater.* **2008**, *157*, 510–517. [[CrossRef](#)] [[PubMed](#)]
3. Pan, Y.; Jiang, J.C.; Wang, R.; Cao, H.Y.; Cui, Y. Predicting the auto-ignition temperatures of organic compounds from molecular structure using support vector machine. *J. Hazard. Mater.* **2009**, *164*, 1242–1249. [[CrossRef](#)] [[PubMed](#)]
4. Frutiger, J.; Marcarie, C.; Abildskov, J.; Sin, G. Group-contribution based property estimation and uncertainty analysis for flammability-related properties. *J. Hazard. Mater.* **2016**, *318*, 783–793. [[CrossRef](#)]
5. Chen, C.C.; Liaw, H.J.; Kuo, Y.Y. Prediction of autoignition temperatures of organic compounds by the structural group contribution approach. *J. Hazard. Mater.* **2009**, *162*, 746–762. [[CrossRef](#)] [[PubMed](#)]
6. Rota, R.; Zanoelo, E.F. Prediction of the auto-ignition hazard of industrial mixtures using detailed kinetic modeling. *Ind. Eng. Chem. Res.* **2003**, *42*, 2940–2945. [[CrossRef](#)]
7. Peper, S.; Dohrn, R.; Konejung, K. Methods for the prediction of thermophysical properties of polyurethane raw material mixtures. *Fluid Phase Equilib.* **2016**, *424*, 137–151. [[CrossRef](#)]
8. Lan, J.X.; Jiang, J.C.; Pan, Y.; Dou, Z.; Wang, Q.S. Experimental measurements and numerical calculation of auto-ignition temperatures for binary miscible liquid mixtures. *Process. Saf. Environ. Prot.* **2018**, *113*, 22–29. [[CrossRef](#)]
9. Luan, F.; Xu, X.; Liu, H.; Cordeiro, M.N.D.S. Prediction of the baseline toxicity of non-polar narcotic chemical mixtures by QSAR approach. *Chemosphere* **2013**, *90*, 1980–1986. [[CrossRef](#)] [[PubMed](#)]

10. Zare-Shahabadi, V.; Lotfizadeh, M.; Gandomani, A.R.A.; Papari, M.M. Determination of boiling points of azeotropic mixtures using quantitative structure-property relationship (QSPR) strategy. *J. Mol. Liq.* **2013**, *188*, 222–229. [CrossRef]
11. Gaudin, T.; Rotureau, P.; Fayet, G. Mixture descriptors toward the development of quantitative structure-property relationship models for the flash points of organic mixtures. *Ind. Eng. Chem. Res.* **2015**, *54*, 6596–6604. [CrossRef]
12. Zhou, L.L.; Wang, B.B.; Jiang, J.C.; Pan, Y.; Wang, Q.S. Predicting the gas-liquid critical temperature of binary mixtures based on the quantitative structure property relationship. *Chemom. Intell. Lab. Syst.* **2017**, *167*, 190–195. [CrossRef]
13. Sobati, M.A.; Aboali, D.; Maghbooli, B.; Najafi, H. A new structure-based model for estimation of true critical volume of multi-component mixtures. *Chemom. Intell. Lab. Syst.* **2016**, *155*, 109–119. [CrossRef]
14. 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. Inf.* **2012**, *31*, 202–221. [CrossRef] [PubMed]
15. Muratov, E.N.; Kuz'Min, V.E.; Artemenko, A.G.; Kovdienko, N.A.; Gorb, L.; Hill, F.; Leszczynski, J. New QSPR equations for prediction of aqueous solubility for military compounds. *Chemosphere* **2010**, *79*, 887–890. [CrossRef]
16. Polishchuk, P.; Madzhidov, T.; Gimadiev, T.; Bodrov, A.; Nugmanov, R.; Varnek, A. Structure–reactivity modeling using mixture-based representation of chemical reactions. *J. Comput.-Aided Mol. Des.* **2017**, *31*, 829–839. [CrossRef] [PubMed]
17. 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. [CrossRef] [PubMed]
18. Kuz'Min, V.E.; Artemenko, A.G.; Polishchuk, P.G.; Muratov, E.N.; Hromov, A.I.; Liahovskiy, A.V.; Andronati, S.A.; Makan, S.Y. Hierarchic system of QSAR models (1D–4D) on the base of simplex representation of molecular structure. *J. Mol. Model.* **2005**, *11*, 457–467.
19. ChemAxon. Available online: <https://chemaxon.com> (accessed on 23 March 2019).
20. Muratov, E.N.; Artemenko, A.G.; Varlamova, E.V.; Polishchuk, P.G.; Lozitsky, V.P.; Fedchuk, A.S.; Lozitska, R.L.; Gridina, T.L.; Koroleva, L.S.; Sil'nikov, V.N.; et al. Per aspera ad astra: Application of Simplex QSAR approach in antiviral research. *Future Med. Chem.* **2010**, *2*, 1205–1226. [CrossRef] [PubMed]
21. GitHub. Available online: <https://github.com/DrrDom/sirms/releases/tag/v1.1.2> (accessed on 23 March 2019).
22. Pan, Y.; Jiang, J.C.; Wang, R.; Cao, H.Y.; Cui, Y. A novel QSPR model for prediction of lower flammability limits of organic compounds based on support vector machine. *J. Hazard. Mater.* **2009**, *168*, 962–969. [CrossRef]
23. Zhao, X.Y.; Pan, Y.; Jiang, J.C.; Xu, S.Y.; Jiang, J.J.; Ding, L. Thermal hazard of ionic liquids: Modeling thermal decomposition temperatures of imidazolium ionic liquids via QSPR method. *Ind. Eng. Chem. Res.* **2017**, *56*, 4185–4195. [CrossRef]
24. Cassani, S.; Kovarich, S.; Papa, E.; Roy, P.P.; Wal, L.v.d.; Gramatica, P. Daphnia and fish toxicity of (benzo)triazoles: Validated QSAR models, and interspecies quantitative activity-activity modelling. *J. Hazard. Mater.* **2013**, *258*, 50–60. [CrossRef]
25. Pan, Y.; Jiang, J.C.; Wang, R.; Zhu, X.; Zhang, Y.Y. A novel method for predicting the flash points of organosilicon compounds from molecular structures. *Fire Mater.* **2013**, *37*, 130–139. [CrossRef]
26. Rücker, C.; Rücker, G.; Meringer, M. Y-Randomization and its variants in QSPR/QSAR. *J. Chem. Inf. Model.* **2007**, *47*, 2345–2357. [CrossRef]
27. *Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models*; ENV/JM/MONO(2007)2; OECD Environment Health and Safety Publications, Series on Testing and Assessment, No. 69; Organization for Economic Cooperation and Development (OECD): Paris, France, 2007.

