Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Engineering 45 (2012) 512 – 517

Procedia Engineering

www.elsevier.com/locate/procedia

2012 International Symposium on Safety Science and Technology

A model for predicting the auto-ignition temperature using quantitative structure property relationship approach

Fang-Yi TSAI[a], Chan-Cheng CHEN[b,*], Horng-Jang LIAW[a]

[a]*Department of Occupational Safety and Health, China Medical University, No.91 Hsueh-Shih Rd. , Taichung City 404, Taiwan, China*
[b]*Department of Safety Health and Environmental Engineering, Kaohsiung First University of Science and Technology, No 2 Jhuoyue Rd. , Nanzih District, Kaohsiung City 811, Taiwan , China*

Abstract

While flammable materials are operated in process industries, the electric equipments should be explosion-proof to reduce the possibility of a fire or an explosion. Auto-ignition temperature (AIT) of a flammable material is the primary characteristic in determining the specifications of these explosion-proof equipments. However, due to limitations on experiments, the AIT of a compound reported in different data compilations is very diverse, and the difference between different compilations was found to be higher than 300 K in many cases. Thus, an effective method to predict the AIT of flammable materials is indispensible in this regard. In this study, a model to predict the AIT of organic compounds is built by using the quantitative structure property relationship (QSPR) approach. This model is built from a set of 820 organic compounds, which are collected from the DIPPR database supported by American Institute of Chemical Engineer. This model is of four molecular descriptors: mean electrotopological state, the aromatic ratio, rotatable bond fraction, and atom-centered fragments. It is found that the R value of the proposed model is 0.900, the average error in percentage is of 6.0%, and the average absolute error is about 36.0 K. While comparing with other works in the literature, this model is built from the largest data set and gives satisfactory performance. As compared with the known experimental errors in measuring the AIT, the proposed model also offers a reasonable estimate of the AIT. Thus, the proposed model can estimate the AIT of a compound for which its AIT is as yet not readily available within a reasonable accuracy.

Keywords: auto-ignition temperature; quantitative structure property relationship; stepwise regression.

1. Introduction

As the technology progresses, more and more flammable materials are operated in process industries. Thus, explosion-proof electric equipments are required in these industries to safely handle flammable materials. The allowable maximum surface temperature of these electric equipments is one of the important characteristics to classify the such equipments and the required specification on these equipments depends on the auto-ignition temperature (AIT) of the flammable materials being operated. For example, article 500.8 of NFPA 70 (also known as the National Electric Code) provides that “Class I equipment shall not have any exposed surface that operates at a temperature in excess of the ignition temperature of the specific gas or vapors.”[1] AIT is defined as the lowest temperature at which the substance will produce hot-flame ignition in air at atmospheric pressure without the aid of an external ignition source such as spark or flame [2]. Obviously, the ability of a substance to spontaneously ignite is important to people who handle, transport, and store these flammable materials. However, although the AIT data are indispensable to safely handle and operate flammable materials, the AIT data reported

* Corresponding author. Tel.: +8-867-601-1000 ext.2311; fax: +-8-867-601-1061.
E-mail address: chch_chen@nkfust.edu.tw

1877-7058
doi:10.1016/j.proeng.2012.08.195

in different data compilations are very much diverse. The difference between different data compilations might be up to more than 300 K for many flammable liquids. Such diversity is attributed to many experimental factors and has been discussed in the literature [3]. Besides this diversity, determining the AIT of a chemical by experimental approach is very laborious and is not always feasible[4]. In this regard, the ability to estimate the AIT of flammable materials by mathematical model will be a cost-efficient and critical aid to this discipline.

One of the important approaches to predict the AIT of a flammable material is the quantitative structure property relationship (QSPR) approach [5-12]. In this category, many molecule-based parameters, which are often called as "molecular descriptors", are directly calculated from the molecular structure of a compound, and then the relationship between the target property and these molecular descriptors are developed. As this approach does not require any existing properties of a compound, the developed model could be can easily apply to the case of predicting the FP of a novel substance. Thus, this approach has been adopted in many research to predict the AIT of flammable materials. Suzuki et al. had proposed a five-descriptor multiple linear regression (MLR) model for predicting the AIT of hydrocarbons. This model was built from a data set of fifty hydrocarbons, and the R value was reported to be 0.941 [5]. Suzuki extended aforementioned model to be a 6-descriptor model by examining 21 descriptors and the data set used to build the model also expanded to 250 hydrocarbons. The R value of the new model was reported to be 0.952 [6]. Egolf and Jurs had pointed the fitting performance is very limited if all hydrocarbons are considered to be a group, so they divided hydrocarbons into four categories: (1) low-temperature hydrocarbons; (2) high-temperature hydrocarbons; (3) alcohols; and (4) esters. These four models are a 8-descriptor model builds from 58 compounds, a 5-descriptor model builds from 46 compounds, a 4-descriptor model builds from 28 compounds and a 4-descriptor model builds from 25 compounds. Their R values are reported to be 0.975, 0.939, 0.970 and 0.963, respectively [7]. Mitchell and Jurs had divided the organic compounds into more categories to enhance the predictive performance of their model. They proposed AIT models for low-temperature hydrocarbons, high-temperature hydrocarbons, nitrogen compound, oxygen/sulfur compounds and alcohol/ester compounds, and artificial neural network models instead of the multiple linear regression models are adopted in this work [8]. Kim et al. built a 9-descriptor model from a data set of 157 organic compounds, and the R value was found to be of 0.959 [9]. Pan et al. (2008) explored the performance of the supported vector machine (SVM) approach in predicting the AIT of the flammable materials [10]. Pan et al. (2009) had proposed and compare a SVM model with the MLR model from a data set of 356 compounds [11]. Bagheri et al. had proposed a 3-descriptor model to predict AIT, but their model is applicable to the organic sulfur only [12]. Although many models for predicting the AIT of flammable materials are proposed in the literature, most of them are built from a data set of lower than 300 compounds, and this will make the model built up by regression approach to be not robust. Thus, a model which is built from a larger data set is still required in this regard.

2. Methodology

In present work, the AIT of 820 organic compounds are collected from DIPPR database supported by American Institute of Chemical Engineer (AIChE) [13]. The collected compounds show wide variability in both their AIT and molecular weight. Their AIT ranges from 444 K to 874 K and their molecular weight ranges from 27.0 to 681.6. Figs.1 and 2 show the distribution of the AIT and molecular weight for the explored compounds, respectively. It could be found from Fig.2 that most of the explored compounds are of the molecular weight below 300.

In this work, molecular structure of the explored compounds are drawn into Hyperchem software and pre-optimized using MM+ and then AM1 molecular mechanics force field [14]. Since the values of some types of molecular descriptors depend on bonds lengths and bonds angles, the optimized chemical structures are necessary to avoid errors in calculating these descriptors. In next step the Dragon software are used to calculate the all the molecular descriptors for all explored compounds according to their optimized chemical structures [15]. Dragon software can calculate up to 3224 descriptors for every molecule. However, some of these molecular descriptors will give the same numerical values for all compounds in our data set. After dropping these molecular descriptors, there are 1707 molecular descriptors.

The remaining 1707 molecular descriptors are then considered as the candidates of the regressor variables in a multiple linear regression (MLR) model. As we know, when the MLR model, which is depicted in equation (1), is built up from a large number of regressors (i.e., molecular descriptors in this study), there might be interactions between these regressors, and we should properly assess the correlations between the regressors. Otherwise, it is possible to include redundant regressors that confuse the identification of significant effects for a model. Thus, a key problem in developing a QSPR model is to find a model can predict the desired property with the least number of molecular descriptors as well as with the highest accuracy.

The MLR model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \tag{1}$$

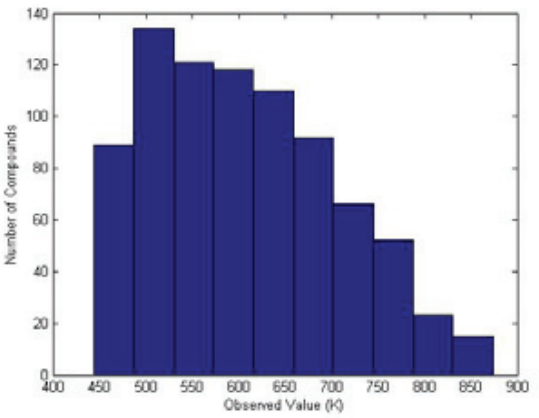

Fig. 1. Distribution of Auto-Ignition Temperatures.

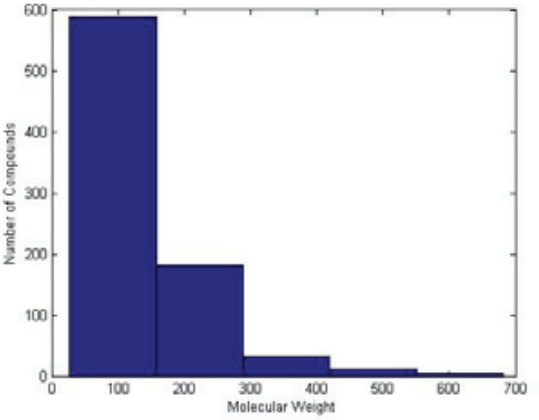

Fig. 2. Distribution of Molecular Weights.

Selecting a subset of regressors to create a model with smaller number of regressors is the problem of feature selection. Selection criteria usually involve the minimization of a specific measure of the predictive error for models which are fit to different subsets of the regressors. Algorithms are then applied to search for a specific subset of regressors that optimally model measured responses, subject to constraints such as required or excluded features and the size of the subset. In the literature many algorithms have been proposed to accomplish this work. Stepwise regression, which is adopted in this work, is a systematic method for adding and removing regressors from a MLR model based on their statistical significance in a regression. The stepwise regression method begins with an initial model and then compares the explanatory power of incrementally larger and smaller models. At each step, the p-value of an F-statistic is computed to test models with and without a potential regressor. If a regressor is not currently in the model, the null hypothesis is that the regressor would have a zero coefficient if it is added to the model. If there is sufficient evidence to reject the null hypothesis, the regressor is added to the model. Conversely, if a regressor is currently in the model, the null hypothesis is that the regressor has a zero coefficient. If there is insufficient evidence to reject the null hypothesis, the regressor is removed from the model. However, depending on the regressors included in the initial model and the order in which regressors are moved in and out, the method may build up different models from the same set of potential regressors. In this sense, models obtained by the stepwise regression method are locally optimal, but may not be globally optimal. To overcome this drawback, the random search technique is introduced to automatically set up the initial model in the algorithm in present work, and the details could be found in our previous work [16].

3. Results and discussions

A 4-descriptor MLR model for predicting the AIT of the flammable organic compounds is proposed in present work, the suggested model is shown in equation (2).

$$\begin{aligned} AIT(K) &= 495.39(\pm 8.07) + 57.79MS(\pm 3.15) + 194.80ARR(\pm 6.33) \\ &- 388.70RBF(\pm 18.08) + 49.06C-040(\pm 3.17) \end{aligned} \tag{2}$$

The meaning of these molecular descriptors are summarized in Table 1 and briefly discussed in the following. *MS*, which means electrotopological state, is calculated by dividing the sum of electrotoplogical state of the *i*th atom in the molecule by the number of nonhydrogen atoms in the molecule. The aromatic ratio (*ARR*) is the ratio of the number of aromatic bonds over the total number of non-H bonds. The rotatable bonds fraction (*RBF*) is the number of rotatable bonds divided by the number of the bonds in a molecule. *C-040* means the number of the description in a molecule. The description includes R-C(=X)-X / R-C#X / X=C=X, in which R represents any group linked through carbon, X represents any electronegative atom (O, N, S, P, Se, halogens), - represents a single bond, = represents a double bond, and # represents a triple bond. The details of these molecular descriptors could be found in the book written by Todeschini and Consonni [17].

Table 1. Molecular Descriptors for the Proposed Models

No.	Type	Molecular descriptors	Definition
1	Constitutional descriptors	Ms	Mean electrotopological state
2	Constitutional descriptors	ARR	Aromatic ratio
3	Constitutional descriptors	RBF	Rotatable bond fraction
4	Atom-centred fragments	C-040	R-C(=X)-X / R-C#X / X=C=X

Table 2 shows the statistical details for this model, and it could be found form this table that the chosen molecular descriptors are highly significant for the proposed model, and the R value and the average error of this model are 0.900 and 36.0 K, respectively. The predicted values and experimental values are compared in Fig.3, and the distribution of the predicted error is shown in Fig.4. As it could be seen most predicted errors are below 10% which is comparable with the experimental accuracy. Moreover, no obvious outlier are found in this study.

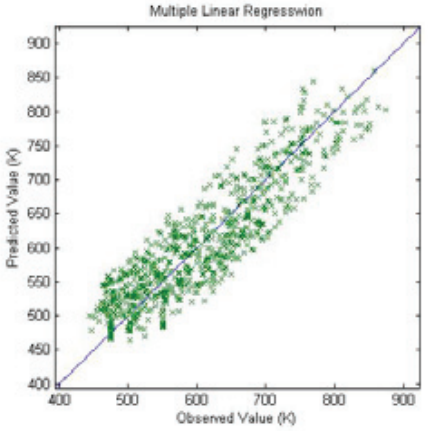

Fig. 3. Plot of Predicted AIT v.s. experimental AIT.

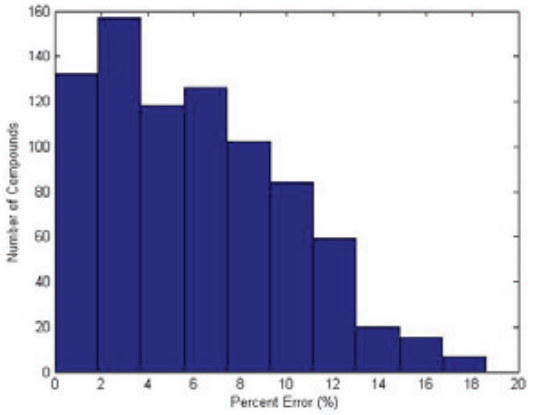

Fig. 4. Distribution of predictive percentage errors.

Table 2. Statistical details of the proposed MLR model

Descriptor	Coefficient	Standard error of coefficient	*t*-test	*p*-Value
Intercept	495.39	8.07	61.39	6.83E-308
MS	57.79	3.15	18.33	3.96E-63
ARR	194.80	6.33	30.75	1.82E-138
RBF	-388.70	18.08	-21.50	1.51E-81
C-040	49.06	3.17	15.46	1.87E-47

R = 0.900, S = 43.0, n = 820, maximum error = 89.1K, average absolute error = 36.0K

$$\text{average error (K)} = \sum_{k=1}^{n} \left| Y_k - \hat{Y}_k \right| \Big/ n$$

Table 3 compares the performance of the proposed model with those of the other research. It could be found that the proposed model includes most amount of the AIT data in the model building step, and it gives satisfactory performance for practical applications with the least molecular descriptors. As it can be seen in this table that the amount of the AIT data in most research is less than 300, except for the work by Pan et al. [11]. Their SVM model, in fact, is a nonlinear model and it is of many parameters which are opaque to the users, so it is not compared with the present work. As for their 9-descriptor MLR model, the R-value is 0.932 and the average absolute error is about 30.1 K. These two performance indices seem to a little better than those of the proposed one, 0.900 and 36.0 K, but it should be noted that the proposed model is of 4-descriptor and is built from a data set of 820 organic compounds rather than of 9-descriptor and from a data set of 356

compounds for their model. Thus, the proposed method gives a method to predict the AIT of organic compounds with satisfactory performance for most widely applications.

Table 3. Comparison of the Predictive Performance between Existing QSPR Models with the propose one

Study	Model type	Compounds	Descriptors number	n	Temperature rang (K)	R	s	RMSE (K)	AAE (K)
Suzuki et al.	MLR	Hydrocarbons	5	50[a]		0.941	39.4	-	28.8
Egolf and Jurs	MLR	Hydrocarbons	8	58[b]	475-674	0.975	12.0	-	-
			5	46[b]	678-835	0.939	16.0	-	-
		Alcohols	4	28[b]	548-872	0.970	24.0	-	-
		Esters	4	25[b]	531-843	0.963	20.0	-	-
Suzuki	MLR	Organic	6	89[a]	471.15~907.15	0.951	33.0	-	33.9
Mitchell and Jurs	MLR	Hydrocarbon	5	47[c]	423.15~623.15	0.872	-	19.0	-
			6	46[c]	623.15~873.15	0.888	-	23.7	-
		Nitrogen	6	36[c]	423.15~1023.15	0.953	-	41.1	-
		Oxygen/Sulfur	7	132[c]		0.824	-	55.6	-
		Alcohol/Ether	6	67[c]		0.924	-	35.0	-
Kim et al.	MLR	Hydrocarbon and Heteroatom	9	157[d]	460.15~813.15	0.959	-	25.9	-
Pan et al.(2008)	MLR	Alkane	6	40[e,f,g]	475.15~771.15	0.836	0.3	50.7	-
	BPNN					0.975	-	21.5	-
	SVM					0.984	-	16.4	-
	MLR	Organic		52[h]	453.15~935.15	0.930	0.2	40.6	-
	BPNN					0.959	-	32.7	-
	SVM					0.963	-	29.8	-
Pan et al.(2009)	MLR	Organic	9	356[e,f,h]	475.15~771.15	0.932	38.5	38.0	31.0
	SVM					0.949	-	33.2	27.6
Bagheri et al.	MLR	Organic Sulfur	3	46[j]	644.15~1011.15	0.962	-	17.6	-
This work	MLR	Organic	4	820[j]	444~874	0.900	43.07	42.8	36.0

The AIT Data Sources

a	Collected from the literature
b	DIPPR Project 931: Data Prediction Methods.
c	The Chemsafe Database
d	Not available
e	The University of Oxford Department of Chemistry MSDS web (http://ptcl.chem.ox.ac.uk/MSDS/)
f	The Chemical Database (http://ull.chemistry.uakron.edu/erd/)
g	MSDS Xchange (http://www.msdsxchange.com/english/index.cfm)
h	ICSCs(International Chemical Safety Cards)(http://www.inchem.org/pages/icsc.html)
j	Refs [13]

4. Conclusions

In this work, a 4-descriptor QSPR model for predicting AIT of flammable materials are proposed. The proposed model gives the performance of R = 0.900 and average absolute error = 36.0K. As compared with other existing works, this model is built from the largest data set, with the least descriptors, and give performance with acceptable accuracy while comparing with the experimental accuracy. As the QSPR approach just requires the information of the molecular structure itself to predict the desired properties, this model provides a way to predict the AIT of a developing compounds within reasonable accuracy for assessing their flammability characteristics.

Acknowledgements

The authors would like to thank the National Science Council of the ROC for supporting this study financially under Grant No. NSC 101-2221-E-327 -015 -MY3.

References

[1] Britton, L. G., Cashdollar, K. L., Fenlon, W., Frurip, D., Going, J., Harrison, B. K., Niemeier, J., 2005. Ural E. A. The Role of ASTM E27 Methods in Hazard Assessment Part II: Flammability and Ignitability, Process Safety Progress 24(1), p. 12.
[2] ASTM International, 2000. ASTM Standard Test Method E659-78, The American Society for Testing and Materials, PA: West Conshohocken.
[3] Chen, C. C., Liaw, H. J., Shu, C.M., Hsieh, Y. C., 2010. Autoignition Temperature Data for Methanol, Ethanol, Propanol, 2-Butanol, 1-Butanol and 2-Methyl-2,4-Pentanediol, Journal of Chemical & Engineering Data 55, p. 5059.
[4] Chen, C. C., Hsieh, Y. C., 2010. Effect of experimental conditions on measuring auto-ignition temperatures of liquid chemicals, Industrial & Engineering Chemistry Research 49, p. 5925.
[5] Suzuki, T., Ohtaguchi, K., Koide, K., 1992. Correlation and prediction of autoignition temperatures of hydrocarbons using molecular properties, Journal of Chemical Engineering of Japan 25, p. 606.
[6] Suzuki, T.,1994. Quantitative structure-property relationships for auto-ignition temperatures of organic compounds, Fire and Materials 18, p. 81.
[7] Egolf, L. M., Jurs, P. C., 1992. Estimation of autoignition temperatures of hydrocarbons, alcohols, and esters from molecular structure, Industrial & Engineering Chemistry Research 31, p. 1798.
[8] Mitchell, B. E., Jurs, P.C., 1997. Prediction of auto-ignition temperatures of organic compounds from Molecular Structure, Journal of Chemical Information and Computer Science 37, p. 538.
[9] Kim, Y. S., Lee, S. K., Kim, J. H., Kim, J. S., No, K. T., 2002. Prediction of autoignition temperatures (AITs) for hydrocarbons and compounds containing heteroatoms by the quantitative structure–property relationship, Journal of the Chemical Society, Perkin Transactions 2, p. 2087.
[10] Pan, Y., Jiang, J., Wang, R., Cao, H., 2008. Advantages of support vector machine in QSPR studies for predicting auto-ignition temperatures of organic compounds, Chemometrics and Intelligent Laboratory Systems 92, p. 169.
[11] Pan, Y., Jiang, J., Wang, R., Cao, H., Cui, Y.,2009. Predicting the auto-ignition temperatures of organic compounds from molecular structure using support vector machine, Journal of Hazardous Materials 164, p. 1242.
[12] Bagheri, M., Borhani, T. N. G., Zahedi, G., 2012. Estimation of flash point and autoignition temperature of organic sulfur chemicals, Energy Conversion and Management 58, p. 185.
[13] AIChE, 2009. Evaluated process design data, Design Institute for Physical Properties Relationships (DIPPR) Project 801.
[14] Hypercube Inc., 2008. HyperChem Release 8.0 for Windows. Molecular Modeling System.
[15] Talete srl., 2007. Dragon for Windows (Software for molecular Descriptor Calculations) Version 5.5.
[16] Chen, C. C., Liaw, H. J., Tsai, Y. J., 2010. Prediction of Flash Point of Organosilicon Compounds Using Quantitative Structure Property Relationship Approach, Industrial & Engineering Chemistry Research 49, p.12702
[17] Todeschini, R. and Consonni, V., 2009. Handbook of Molecular Descriptors Vol. 11, Wiley, Weinheim, New York, Chichester, Brisbane, Singapore, Toronto.