ABC Journals
a new era in publishing
Asian Journal of Applied Science and Engineering
Abbreviated key title: Asian j. appl. sci. eng.
Access this article online
http://journals.abc.us.org/index.php/ajase/issue/archive

Prediction of the Upper Flammability Limit of Pure Compounds

Tareq A. Albahri

Chemical Engineering Department, Kuwait University, P.O.Box 5969, Safat 13060, **KUWAIT**

ARTICLE INFO

Received: Mar 03, 2016
Accepted: Mar 17, 2016
Published: Mar 26, 2016

*Corresponding Contact
Email: toalbahri@gmail.com

crossref doi Prefix 10.18034

ABSTRACT

A quantitative structure-property relation model was developed to predict the upper flammability limit (volume percent in air) of pure compounds, based only in their structures. A group contribution method was used to determine the upper flammability limit (UFL) through an artificial neural network. The method was used to identify the structure groups that have a significant contribution to the target property and concluded that 30 atom-type structure groups can represent the UFL for 550 pure substances. The models input parameters are the number of occurrence of each of the 30 structure groups in the molecule. The model predicts UFL with a correlation coefficient of 0.9996 and average absolute deviation of 0.17 vol %. The results were compared with multivariable regression model and other methods in the literature. The model is very useful and convenient for accessing the hazardous risk potential of chemicals for which experimental data is not available.

Key Words: Molecular Modeling, Neural Networks, Quantitative Structure-Property relation, Structure group contribution, Upper flammability limit

Source of Support: EC04/01, Kuwait University

No Conflict of Interest: Declared

How to Cite: Albahri TA. 2016. **Prediction of the Upper Flammability Limit of Pure Compounds** *Asian Journal of Applied Science and Engineering,* 5, 59-70.

INTRODUCTION

The flammability limits provide the range of fuel concentration usually in vol % in air at 298K within which a gaseous mixture can ignite and burn. Below the Lower Flammability Limit (LFL) fuel is not enough to cause ignition. Above the Upper Flammability Limit (UFL), oxygen is insufficient to propagate the reaction away from the ignition source (Sheldon, 1984).

The flammability limits are some of the most important safety specifications used in assessing the overall flammability hazard potential of a chemical substance in storage, processing, and handling. They may be used to determine guidelines for the safe handling of volatile chemicals, and in particular to assess ventilation requirements for gases and vapors. A precise experimental determination of the flammability limits requires the use of

a standardized apparatus and conditions as specified in ASTM standard E681 (ASTM, 1985).The experimental determination of the flammability limits is laborious and not always practical and a prediction method that is desirably convenient and fast must be used to estimate them.

Albahri (2003a) presented a detailed review of the available methods in the literature to determine the flammability limits, and further predicted UFL of hydrocarbon compounds from the molecular structure using the Structure Group Contribution (SGC) method and Multivariable Nonlinear Regression (MNLR) technique. The input parameters to the final equation are the number of occurrence of each of the structure groups in the molecule. Although the model was able to accurately predicted UFL for 464 compounds with a correlation coefficient (R) of 0.96 and an average absolute deviation (AAD) of 1.25 vol %, it is limited to only hydrocarbons.

In previous work, Albahri (2013a, 2014a, 2014b, 2015) developed Artificial Neural Network (ANN) models for estimating various properties of pure substances including the LFL with high accuracy by breaking down their molecular structure into smaller atom-type groups and measuring their contribution to the desired property. This work is a continuation of that effort to predict the UFL using the same technique, with the ultimate future goal to predict the molecular composition and properties of light petroleum fractions like Kerosene and diesel (Albahri, 2005a; Albahri, 2005b).

Recently, Lazzus (2011) developed ANN model based on particle swarm optimization (PSO) to estimate UFL of organic compounds in air from their molecular structure. The input to the ANN was 40 structure groups that define the organic compounds in addition to the molecular weight and the dipole moment. The neural network constructed using 42 neurons in the input layer, six neurons in the hidden layer, and two neurons in the output layer was able to predict the target property with R = 0.9818. The maximum percentage errors were 27.8 %, which is rather high.

Gharagheizi (2009) developed a quantitative structure property relation(QSPR) model to predict the UFL of compounds using a five-parameter multivariable linear equation calculated from the chemical structure of the molecules with an average absolute error (AAE) of 9.7% and R = 0.92 as follows,

$$\begin{aligned} UFL\ (vol\%) = {} & 10.35415(\pm 0.31456) - 1.35486\ (\pm 0.08144)jhetv \\ & - 42.28779(\pm 0.144928)PW5 + 18.59571(\pm 0.62369)SIC0 \\ & - 0.98203(\pm 0.0703)MATS4m + 0.68363(\pm 0.03235)MLOGP \end{aligned} \quad (1)$$

where jhetv is Balaban-type index from van der Waal's weighted distance matrix, PW5 is path/walk 5 Randic shape index, SIC0 is the structural information content (neighborhood symmetry of zero-order), MATS4m is Moran autocorrelation-lag 4 weighted by atomic masses, and MLOGP is Moriguchi octanol-water partition coefficient (log P).Not only does the method use five molecular descriptors that are difficult to determine but also is less accurate than the method of Lazzus (2011).

Gharagheizi (2010) further developed a molecular model for estimation of UFL of pure compounds using a three-layer feed-forward neural network algorithm the input parameters of which are the number of occurrences of 113 predefined functional groups. The ANN model constructed using 113 neurons in the input layer, 4 in the hidden layer, and 1 in the output layer predicts the UFL with AAE = 7.07%, AAD = 0.882 and R= 0.9469.

Gharagheizi et al. (2012) further developed a corresponding states equation for estimation of the UFL temperature of chemical compounds with AAE = 1.7%,

$$UFLT(K) = 2\omega + 11.799 + T_c(0.1401\omega - 0.1401(0.494\omega - 0.3851)^2 + 0.4943) \quad (2)$$

This equation requires the acentric factor ω and the critical temperature, T_c in K, which is not convenient. Our goal is to predict the UFL from the molecular structure of the compound alone.

A careful examination of the UFL of hundreds of pure compounds reveals its dependency on the molecular structure of the compound. In this work, we investigate this dependence of UFL on the molecular structure using a SGC approach, which is a proven and very powerful tool for predicting many physical and chemical properties of pure compounds (Reid et al., 1987). Many commercially available software packages such as AIChE-Cranium (Molecular Knowledge Systems, 1998) and ASTM-CHETAH (ASTM, 2001; Seaton et al., 1974) estimate the properties of pure compounds from their chemical structure.

Numerous structure group contribution methods, including the work of Ambrose, Joback, Fedors, Thinh et al., Benson, Yoneda, Qrrick-Erbar, Grunberg-Nissan, and Chueh-Swanson, are available in the literature (Reid et al., 1987). The major differences between these methods involve the choice of the structure groups and the way in which these groups contribute to the properties overall. The group contribution methods of Benson and Yoneda allow for the effects of next-nearest neighbors, but the calculations are difficult to master. Joback's method is easier to use and has wide applicability. The property calculations may be simple algebraic additions of the group contribution values or a very complex set of equations, such as Thinh et al.'s equations for calculating the heat capacity (Reid et al., 1987; Joback, 1984). In general, the complex methods are more accurate, but the simple methods provide reliable estimates and are easier to use.

METHOD

A careful examination of the UFLs of hundreds of compounds reveals the complex dependence of this property on the structure of the molecule (API, 1987; Dewan, 2006).We must account for all of the parameters when selecting appropriate structure groups that can represent the UFL property.

Structure Group Contribution (SGC)

UFL is a macroscopic property of compounds that relates to the molecular structure and determines the magnitude and predominant types of the intermolecular forces. The concept of structure suggests that a macroscopic property can be calculated from group contributions. The relevant characteristics of the structure are related to the atoms, such as atomic groups, bond type etc. We assign numerical values through regressions of available experimental data using statistical methods and determine the property through algebraic equations that sum the contributions of the parts of the molecule.

Of the many SGC estimation methods available in the literature, the Joback (1984) definitions of group contributions were selected as a starting point based on their simplicity, generality and accuracy (Reid et al., 1987).This combination was tested and modified with the functional groups in the molecules that result in the best R and AAD while using ANN. The Joback method of group contributions is explained in detail elsewhere (Reid, 1987; Joback, 1984).Ever since, additional experimental values and

efficient computational techniques have been developed. We reevaluated Joback's scheme based on the experimental data as explained above and determined the values of the group contributions. From the original Joback scheme, the groups in Table 1 are adopted to account only for those that have a significant influence on the UFL property. For example, no significant distinction in the UFL existed for the cis and transstructural orientations in olefins or cyclic compounds. Therefore, this distinction was avoided when choosing structure groups. Knowing the location of the alkyl substitutions on the benzene ring or in the ortho, meta, and para positions in aromatics, the location of the alkyl branches along the chain for isoparaffins and isoolefins, the location of the double bond along the chain in olefins, and the alkyl substitutions for naphthenes was unnecessary. Our attempts to enhance the model results by using two sets of structure groups, one for the aromatic ring in aromatics and another for the cyclic ring in naphthenes did not significantly improve the model predictions and correlation with the experimental data; therefore, we avoided such distinctions.

SGC-MNLR model

In a traditional SGC approach, the group contributions are usually incorporated in some form of algebraic equation related to other properties, such as the boiling point, molecular weight, or correlation constants to estimate the desired property. Many equations have been proposed ranging from linear algebraic-summation of the SGC values to nonlinear and polynomial functions (Reid, 1987). Albahri (2003a , 2003 b, 2012, 2013 b) tested several correlations and concluded that the target property is best predicted using the following polynomial form,

$$\phi = a + b\,(\sum n_i\varphi_i) + c\,(\sum n_i\varphi_i)^2 + d\,(\sum n_i\varphi_i)^3 + e\,(\sum n_i\varphi_i)^4 \qquad (3)$$

where the correlation constants a, b, c, d, and e are to be determined through regression of the experimental data. ϕ is the target property of interest.n_i is the number of occurrences of each structure group in the molecule. φ_i is the contribution value of each structure group that can be determined through regression of the experimental data by MNLR, and $\sum n_i\varphi_i$is the sum of the SGCs for each molecule toward the target property.

We used the Generalized Reduced Gradient (GRG2) optimization algorithm (Lasdon et al., 1978) in the solver function of Microsoft Excel™ to determine the parameters (structure group contribution values and the constants for Eq.3) while using MNLR. The objective was to minimize the sum of the absolute deviations between the calculated and experimental UFL for the compounds by changing the values of the structure group contributions and the constants for Eq.3. That resulted in the lowest absolute AAD and the best R.

SGC-ANN model

Lippmann (1987) and Widrow and Lehr (1990) presented the theoretical basis of neural computing. ANNs are computational models inspired by the human central nervous system capable of machine learning and pattern recognition. ANNs are systems of inter-connected mathematical functions called "neurons." They can compute values from inputs once the inter-connecting weights between the neurons are determined. Neural network computations have several advantages over traditional MNLR methods regarding the speed of computation, learning ability, and fault tolerance (Lippmann, 1987; Widrow and Lehr, 1990).ANN technique has been widely applied to various engineering areas such as modeling of chemical processes (Elkamel et al., 1999)and prediction of the thermodynamic and the transport properties (Lee and Chen, 1993; Ismail et al., 1996; Bunz et al., 1999).Moreover, Albahri (2003a, 2003b, 2013a,

2014, 2015) used combination SGC and ANN models to predict thermodynamic properties of pure compounds with various ANN architectures and input parameters.

In this work, we estimate the UFL of pure compounds from their molecular structure using a SGC approach. Toward that purpose, ANN model was constructed using MATLAB (the Mathworks Inc., 2001) code to test several possible SGC definitions (Reid et al., 1987; Albahri, 2003a; Albahri, 2003b; Albahri, 2013a; Albahri, 2014; Albahri, 2015) and propose the most appropriate structure groups and their contribution values for predicting the UFL property. The model was used to investigate the structure groups that have a significant contribution to UFL and to reveal the groups that provided the best correlation with the experimental data. Furthermore, the final groups were tested for their ability to predict the UFL of a new set of compounds that were not used while developing the ANN model.

We tried several ANN architectures and selected the one that best simulated UFL. The trials included using ANN with a single hidden-layer or a double hidden-layer, while varying the number of neurons in the hidden and input layers. They also included using various functional or structure groups, constraining the connection weights amongst the neurons between certain limits, and determining the appropriate number of time steps (epochs) for the runs. The final network structure is shown in Figure 1.

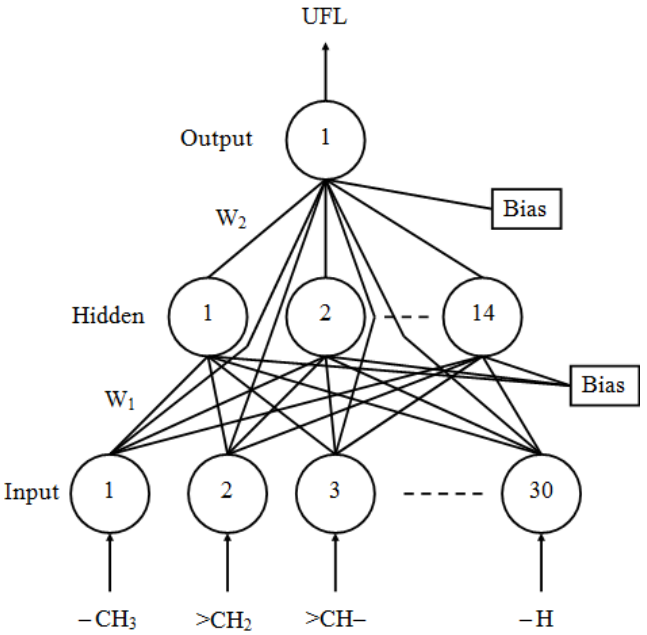

Figure 1: Schematic representation of the 30-14-1ANN used to predict the UFL of pure compoundsfrom their structure groups

It consists input, output, and hidden layers. The number of neurons in the input layer is equal to the number of structure groups investigated. The number of neurons in the hidden layer is varied to determine the optimum. The output layer has one neuron representing the predicted UFL property. A hyperbolic tan-sigmoid (tansig) function was selected as the transfer function for each neuron in the hidden layer and a linear (purelin) function for the neuron in the output layer, both of which are built in functions in the

Matlab library. From MATLAB documentation the functional form for the tansig is the hyperbolic tangent shown in Eq. (4) and purelin is simply y=x. The inputs to the network algorithm are the number of occurrence of the structure groups in the given substance. We assign an input value of zero to the group that does not exist in a molecule.

$$y = \frac{2}{[1 + \exp(-2x)]} \quad (4)$$

When comparing the different structure group definitions and arriving at the final list, the data set contained 550 pure substances that are used as inputs to the neural network. This probing data set is taken from the property databank of API-TDB (API, 1987) and AIChE-DIPPR (Dewan, 2006) to represent the compounds that likely exist in Kerosene and diesel which are the target of our future simulation. These data on UFL were determined using the experimental method described by ASTM standard E681-85 (ASTM, 1985). Each one of these 550 input sets included a group-number vector representing the number of structure groups in a given substance. The connection weights of the network were adjusted iteratively between -1 and 1 through a back-propagation algorithm with the generalized delta rule to minimize the mean square error between the desired and the actual outputs. During learning, we recorded the AAD, maximum deviation, and R of the predictions along with the corresponding epochs. We found that 300 epochs were sufficient to achieve convergence in learning where the deviation between the actual and the desired responses is not significant. Therefore, the training was terminated at that number of time steps. Convergence occurred in less than a minute for all cases when using a desktop computer.

To avoid over-fitting when determining the number of neurons in the hidden layer and ensure that the neural network is trained properly with the experimental data, the maximum number of neurons in the hidden layer is determined using the following constraint for the three-layered network architecture shown in Figure 1,

$$H < (D-1) / (I - 2) \quad (5)$$

Where, H is the number of neurons in the hidden layer rounded down to the nearest whole number. D is the number of experimental data points used when training the neural network, and I is the number of inputs (total number of groups) to the network. Only 90% of the available experimental data were used during network training, while the remaining 10% were used to test the trained network. The actual number of hidden-layer neurons was determined by stepping down one number at a time until the best results were obtained, as indicated by the AAE and R, for both the training and testing data sets. We found that using 14 neurons in the hidden layer is the best for that purpose and therefore the optimal network architecture was 30 neurons in the input layer, 14 neurons in the hidden layer, and one neuron in the output layer, as shown in Figure 1.

A mathematical description of the computation from the ANN model is

$$O = \sigma_O \left[\sum_i w_i^{H \to O} \sigma_H \left(\sum_j w_{ji}^{I \to H} I_j + b_i \right) + b_O \right] \quad (6)$$

Where O is the output, σ_H is the activation (transfer) function for the hidden layer, and σ_O is the activation function for the output layer, $w_i^{H \to O}$ is the weight connecting neuron i from the hidden layer to the output, $w_{ji}^{I \to H}$ is the weight connecting input j to hidden neuron i, I_j is the value of input j, b_i is the bias value associated with hidden neuron i, and b_O is the bias associated with the

output. One should be able to compute UFL from this equation and the sigmoid function provided above, along with the weights and biases determined in this manuscript.

After determining the final network architecture, the predictive ability of the ANN was demonstrated by training it with the experimental values of only495of the compounds then using the trained networks to predict the UFL of the 55 remaining compounds. Due to the limited number of molecules in our data set, we used 90% of the data for training and 10% for testing to ensure that the network is trained well for the various structure groups. The compounds in the testing data set were randomly chosen based on the abundance of their counterparts from the class of compounds they represent in the training data set. This choice was necessary because ANN can predict UFL only for the compounds upon which it was trained. The accuracy of these predictions was then compared with the experimental data.

RESULTS AND DISCUSSION

SGC-MNLR model

Using the experimental UFL values for 550 pure compounds we calculated the constants for Eq.3 and the values for the various structure group contributions in Table 1.

Table 1: Atom-type structure groups corresponding to the input nodes of the ANN in Figure 1 and SGC values for the MNLR model for estimating UFL of pure compounds

Serial No.	Groups	UFL_i
1	– CH_3	1.114692
2	> CH_2	0.339248
3	> CH –	0.138901
4	> C <	-2.50000
5	= CH_2	2.305876
6	= CH –	2.024712
7	= C <	1.346964
8	= C =	6.542315
9	≡ CH	18.1876
10	≡ C –	2.243466
11	> CH_2 (ring)	0.570317
12	> CH – (ring)	0.70827
13	> C < (ring)	0.403864
14	= CH – (ring)	0.672134
15	= C < (ring)	0.654589
16	– F	4.231781
17	– Cl	2.254908
18	– OH (alcohol)	6.41015
19	– O – (non-ring)	13.04636
20	>C = O (non-ring) Ketone	5.417596
21	O = CH – (aldehyde)	13.023
22	– COOH (acid)	0.047503
23	= O	0.001000
24	– NH_2	8.603798
25	> N – (nonring)	1.900000
26	– O – (ring)	3.179851
27	> C = O (ring)	9.414214
28	> NH (ring)	6.019325
29	– N = (ring)	2.458781
30	– H	12.51823

An optimization algorithm was used for that purpose. The algorithm minimizes the sum of the difference between the calculated and experimental UFL. The regression of the algorithm was less than 1 min on a personal computer. The final equation obtained is,

$$UFL = 3.563 + 0.5237\ (\sum n_i UFL_i) + 1.572.10^{-9}\ (\sum n_i UFL_i)^2$$

$$+ 6.375.10^{-8}\ (\sum n_i UFL_i)^3 + 3.266.10^{-5}\ (\sum n_i UFL_i)^4 \qquad (7)$$

Where UFL is the upper flammability limit in vol %, UFL_i is the atom-type structure group contribution to UFL listed in Table 1, n_i is the number of occurrence of each structure group in the molecule, and $\sum n_i\ UFL_i$ is the sum of the atom-type SGCs to the UFL for each molecule. The calculation procedure for UFL using Eq. 7 and the SGC values in Table 1 is illustrated in Albahri (2003a). The results for the SGC-MNLR model to predict UFL using Eq.7 and the structure group contribution values from Table 1 are summarized in Table 2.

Table 2: Comparison of the present ANN and MNLR models with others from the literature for estimating the UFL of pure compounds

Method	Model	Data Set	AAD (Vol.%)	Ave % error	Max % error	Correlation Coefficient
This work	SGC-ANN	Overall 550	0.17	2.04	16.09	0.9996
This work	SGC-ANN	Training 495	0.17	2.09	16.09	0.9996
This Work	SGC-ANN	Testing 55	0.16	1.64	10.84	0.9996
This Work	SGC-MNLR	Overall 550	3.36	22.96	345.74	0.5200
Lazzus (2011)	ANN	Overall 418	-	7.10	27.80	0.9818
Gharagheizi (2009)	QSPR correlation	Overall 856	0.8483	9.70	30.73	0.9200
Gharagheizi (2010)	ANN	Overall 867	0.882	7.07	> 30.92	0.9469

The predictions did not correlate very well with the experimental data with R= 0.52, AAD = 3.36 vol % and AAE = 22.96%. Our efforts to vary the initial guesses did not improve the results at all. The limitations are primarily associated with using a simple mathematical expression (Eq. 7) that was not able capture the complex nature of UFL for the various families of the chemical compounds. Minor improvements were obtained using other optimization tools, such as JMP statistical software, MATLAB or GAMS. Previous work shows that when MNLR fails a nonlinear modeling environment such as ANN can produce better predictions (Pan et al., 2007; Patel et al., 2009; Albahri, 2013a; Albahri, 2014; Albahri, 2015).

SGC-ANN model

While using the probing set of data on UFL, several structure groups derived from the Joback definitions of group contributions were tested and modified. During this stage, R was used to reveal the structure group definitions that significantly contribute to UFL. The set of groups that can best correlate the experimental data (with R = 0.9996) is shown in Table 1comprising 30 groups. The AAD in the predicted UFLs for all types of compounds ranging in UFL from 4.67 to 96.19 vol % was 0.17% as shown in Table 2, which is very accurate. In addition to the proposed structure groups in Table 1, several other groups were investigated. Although better results were obtained with a larger number of structure groups, the improvement was not significant. For instance, to improve the model's predictions we sought to add the following structure groups to distinguish the groups that are part of an aromatic ring from those that are not: =CH−, $>CH_2$, >C= (both fused and non-fused). We also sought to account for the structural orientation of the >C= group in the ortho, meta, and para positions on the aromatic ring (Albahri, 2003a; Albahri, 2003b; Albahri, 2012; Albahri, 2013b). Adding these seven groups improved the

predictions marginally; AAD improved by only 0.005vol. % while R improved by only 0.0001. Many other prospects have been considered but are too numerous to present.

To assess the accuracy of the predictions, the available experimental data were then separated into training and testing data sets consisting of 495 and 55compounds, respectively. The AAD, AAE and R between the correlated UFL and the experimental data used for training the network were calculated. The results from the trained network are summarized in Table 2, showing that the AAD= 0.17vol%, AAE = 2.09 % and R = 0.9996 for UFL calculations for this mode. The correlation between the ANN model and the training data set is very good. The predictions of the trained networks were then cross-validated against a testing set of data for 55 compounds that were not used during the training process. The AAD, AAE and R between the predicted UFL and the experimental data used when testing the network were calculated. The predictions compared favorably against this new data set with an AAD= 0.16 vol %, AAE=1.64%, and R = 0.9996.The predictions for the ANN model for the testing set of data were excellent. The predictions were comparable to the trained networks in terms of AAD, AAE, and R, as shown in Table 2. Figure 2 is a parity plot showing the accuracy of the model correlations for both training and testing data sets.

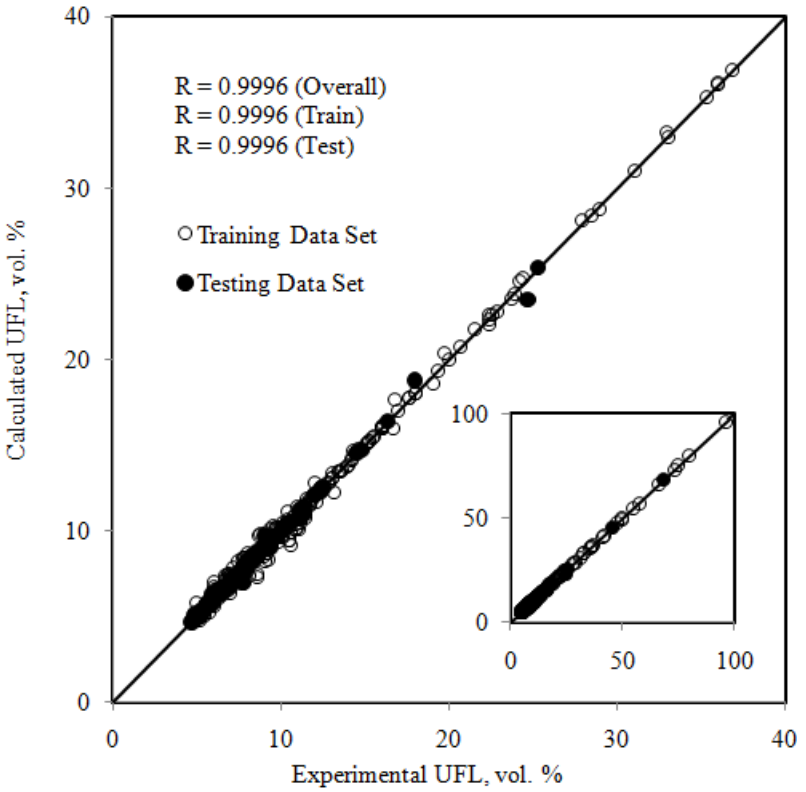

Figure 2: Parity plot showing the accuracy of the model correlation for the UFL of a training set containing 495 pure compounds and the prediction for a testing set containing 55 compounds using the SGC-ANN model

Figure 3 shows the percentage errors for the overall data set where 90% of data is below the 0.5 vol % absolute deviation, 7% of data is within 0.5-1 vol % absolute deviation, and only 3% of data is between 1-1.5% absolute deviations.

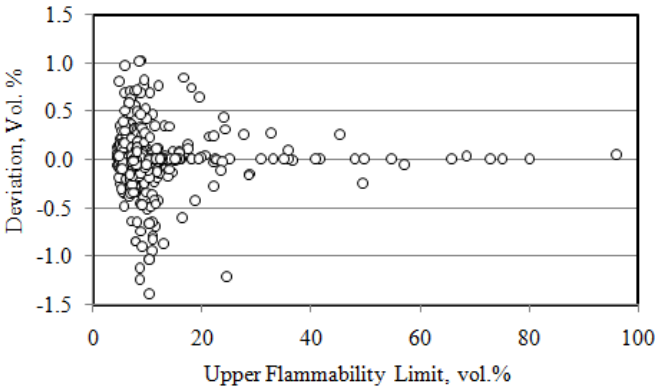

Figure 3: Percentage errors for the UFL for the whole data set of 550 pure compounds when using the SGC-ANN model

Table 3: Parameters (weights and biases) of the hidden layer for the ANN architecture described in Figure 1 for predicting UFL

W_1	Hidden neuron													
Input neuron	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	0.1149	-0.9467	0.1931	0.6017	0.8969	-0.5836	0.3754	0.0384	0.7693	0.4395	-0.4833	-0.0912	-0.5627	0.931
2	-0.2862	-0.8788	0.5127	0.7513	-0.402	-0.5221	0.7975	-0.6421	0.7834	0.4158	0.6178	-0.5264	-0.0771	-0.8319
3	0.6965	0.1209	-0.3628	-0.2381	0.6914	-0.2683	-0.7857	-0.5223	-0.8006	0.1374	-0.6718	0.04	-0.6574	0.536
4	-0.2478	0.6302	-0.9103	0.1863	0.8443	-0.4432	0.7166	-0.6933	-0.7725	-0.5473	0.7696	-0.791	-0.4048	0.9919
5	-0.5713	-0.0031	-0.4207	0.3288	0.861	0.5594	0.3062	-0.6422	-0.719	-0.9993	-0.3764	0.5735	0.8334	-0.9584
6	0.6336	0.2361	0.1209	-0.4871	0.6053	-0.4969	0.4674	0.2778	-0.5081	0.2103	0.1865	0.2888	-0.5436	0.2711
7	-0.654	0.0772	0.2478	0.3538	0.3334	0.7908	-0.8977	-0.3299	0.496	-0.3969	0.7582	0.3236	-0.7308	-0.8525
8	-0.9682	-0.5317	0.0326	-0.0795	0.382	0.1731	0.017	-0.7408	-0.5452	-0.248	-0.3973	0.4905	-0.3194	0.2744
9	0.9652	0.0056	0.8989	0.6242	0.7849	-0.812	0.5752	0.7104	-0.6107	-0.7784	0.4668	0.4016	0.2603	0.5044
10	0.3237	0.7474	-0.4575	-0.1533	-0.5395	-0.9746	-0.7719	0.61	-0.284	-0.0697	0.7353	-0.4914	0.6215	0.311
11	0.776	-0.0231	0.9892	-0.2411	0.059	-0.6688	0.0036	-0.1354	0.2851	0.3577	0.8127	-0.5579	-0.6674	0.1292
12	0.9307	-0.9282	0.7434	-0.9004	0.0367	-0.6458	0.3975	-0.4339	0.771	-0.7474	0.0384	0.7155	0.7595	-0.3275
13	-0.1242	-0.0562	-0.7041	-0.693	0.0611	0.4739	-0.1867	-0.2464	-0.3816	0.7593	0.2246	-0.4686	0.8207	0.279
14	-0.5359	0.3534	0.333	-0.6951	-0.8976	-0.4991	-0.7068	-0.7494	0.6272	-0.6585	-0.8309	0.4235	0.0627	-0.4447
15	-0.2609	-0.9493	0.7815	0.6965	-0.462	0.1458	-0.628	0.1642	-0.3002	0.3267	0.6462	0.1224	0.8249	0.5795
16	-0.6891	0.649	-0.6187	0.2645	0.3177	0.571	-0.2215	-0.1016	-0.0301	0.2227	-0.5758	-0.9016	0.4982	0.027
17	-0.5688	-0.7728	-0.688	-0.176	-0.1734	-0.9387	0.8143	-0.6091	-0.2055	-0.3892	-0.589	0.7182	-0.3043	0.4647
18	-0.1767	-0.1953	0.0111	-0.6293	0.0465	0.2963	-0.8916	0.5862	0.5393	0.4074	-0.0677	-0.7557	0.5506	-0.6096
19	-0.1084	-0.9492	-0.384	0.7143	0.6302	-0.3503	0.7016	-0.0353	0.1081	0.2388	0.2877	0.2112	0.3178	0.0202
20	0.4244	0.0296	0.2126	0.8881	0.6055	-0.3825	0.1616	-0.6435	-0.4365	0.6242	0.2983	-0.8806	0.1057	-0.0902
21	0.8033	-0.4245	-0.8734	-0.0445	0.9093	0.8864	0.1128	0.2651	0.4846	-0.8112	-0.8866	0.0756	-0.8466	-0.0966
22	-0.6037	0.5596	0.2381	-0.922	0.7347	0.5492	0.7502	0.4499	-0.212	-0.3479	0.0072	0.1169	0.4587	0.5557
23	-0.0315	0.589	-0.0582	-0.5657	0.1497	0.3494	0.3258	0.77	0.4801	0.489	0.6508	0.8889	0.0067	0.2563
24	0.5804	-0.1001	0.0489	-0.6252	-0.6943	-0.5902	-0.7271	-0.624	-0.0765	0.5935	-0.389	-0.4983	0.7019	-0.9783
25	0.176	0.0821	0.3083	-0.355	-0.5054	-0.1762	-0.3708	0.3001	0.7788	-0.3231	0.1119	-0.69	-0.5682	-0.4391
26	0.1127	-0.0281	0.908	-0.5102	-0.0401	0.0557	0.5395	-0.5342	0.7279	0.8697	-0.8649	0.4844	0.768	0.6221
27	0.8407	-0.5866	0.3499	0.8128	-0.2936	0.1983	0.2128	-0.8642	0.8566	0.823	0.3425	-0.1093	0.3451	0.2179
28	-0.397	0.6939	-0.7789	-0.3967	-0.7567	-0.2152	-0.3072	0.773	0.6017	-0.4976	-0.8123	-0.8946	0.1278	0.4843
29	0.6086	0.2681	-0.4994	-0.6769	0.2849	0.9362	0.5822	0.7487	-0.3378	-0.4762	0.0648	-0.6097	-0.4962	-0.5623
30	0.3011	-0.8716	-0.5436	0.3186	-0.3554	-0.4058	0.4067	0.7907	-0.6555	-0.8893	-0.666	-0.6045	0.7063	-0.7223
Bias out	1	2	3	4	5	6	7	8	9	10	11	12	13	14
-0.6744	-3.0575	2.5871	-2.1167	-1.6463	-1.176	0.7056	-0.2352	0.2352	0.7056	1.176	-1.6463	-2.1167	-2.5871	3.0575
W_2	Hidden neuron connection to output neuron													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
	0.234	-0.462	-0.5587	0.4258	0.098	0.8827	-0.3403	0.409	0.8869	0.1632	0.7603	0.4992	-0.2408	0.4511

The connection weights and biases are shown in Table 3. It is fair to mention that the data sets used in the literature to model the UFL are so different that these models cannot be compared to one another. However, Table 2 shows that our proposed SGC-ANN model gives good results compared to the other models. Lazzus (2011) model, for example, predicts UFL with a maximum deviation of 27.8vol%which is rather high. Gharagheizi (2009) five-parameter multi-linear equation is less accurate (AAE = 9.7% and R = 0.92) and the five intricate parameters (structural descriptors) pose an additional effort to determine practically. Gharagheizi (2010) ANN model, which predicts the UFL with R = 0.9469, AAE = 7.07% and AAD = 0.882, is further complicated by the use of 113 intricate groups making the method difficult to apply. Compared to other artificial intelligence methods our method has an advantage regarding combined accuracy and simplicity. It utilizes simple atom-type structure group definitions and requires only the compounds molecular structure, which is always known.

Using the SGC-MNLR model has proven to be very successful in the past for predicting the properties of pure hydrocarbons with R as high as 0.99 (Albahri, 2003a; Albahri, 2003b; Albahri, 2012; Albahri, 2013b). However, when the method was applied to other classes of compounds with various functional groups it was less successful with R ranging from 0.79 to 0.90 while ANN has consistently provided better alternative with a high accuracy (Pan et al., 2007; Patel et al., 2009; Albahri, 2013a; Albahri, 2014; Albahri, 2015).

CONCLUSION AND PROSPECT

We developed two models based on MNLR and ANN algorithms using a SGC approach to predict the UFL of pure compounds from their molecular structures. While it was not possible to predict UFL with the SGC-MNLR model, the back-propagation SGC-ANN model offered a significant improvement and an advantage in terms of accuracy (R = 0.9996 and AAE = 0.17 vol %) although it requires specific computation resources. The SGC-ANN model is simpler than other artificial intelligence methods in the literature. Our definition of atomic-type structure groups is fewer, simpler, and easier to use, while providing good correlations with the experimental data. Two major advantages of this work is the accurate identification of the structure groups that affect and contribute to the UFL of molecules, which enhances our understanding of the fundamentals of flammability, and the ease of accessing the SGC data, which might be hard to estimate or calculate using other models.

ACKNOWLEDGMENT

This work was supported by as part of project EC04/01, Kuwait University on the molecular simulation, characterization and properties estimation of petroleum fractions.

REFERENCES

Albahri, T. A. (2003a), "Flammability Characteristics of Pure Hydrocarbons", *Chemical Engineering Science,* Vol. 58no. 16, pp. 3629-3641.

Albahri, T. A. (2003b), "Structural Group Contribution Method for Predicting the Octane Number of Pure Hydrocarbon Liquids', *Industrial and Engineering Chemistry Research,* Vol. 42 no. 3, pp. 657-662.

Albahri, T. A. (2005a), "Molecularly Explicit Characterization Model (MECM) for Light Petroleum Fractions", *Industrial and Engineering Chemistry Research,* Vol. 44 No. 24, pp. 9286-9298.

Albahri, T. A. (2005b), "Enhanced Method for Predicting the Properties of Light Petroleum Fractions", *FUEL,* Vol. 85No. 5-6, pp. 748-754.

Albahri, T. A. (2012), "Prediction of the Aniline Point Temperature of Pure Hydrocarbon Liquids and their Mixtures from Molecular Structure", *J Molecular Liquids,*Vol. 174, pp. 80-85.

Albahri, T. A. (2013a), "Prediction of the lower flammability limit percent in air of pure compounds from their molecular structures", *Fire Safety J,* Vol. 59, pp. 188-201.

Albahri, T. A. (2013b), "Method for predicting the standard net heat of combustion for pure hydrocarbons from their molecular structure", *Energy Conversion Management,* Vol. 76, pp. 1143-1149.

Albahri, T. A. (2014), "Accurate Prediction of the Solubility Parameter of Pure Compounds from their Molecular Structures", *Fluid Phase Equilibria,* Vol. 379, pp. 96–103.

Albahri, T. A. (2015), "MNLR and ANN Structural Group Contribution Methods for Predicting the Flash Point Temperature of Pure Compounds in the Transportation Fuels Range", *Process Safetyand Environmental Protection,*Vol. 93, pp. 182–191.

Albahri, T. A. and George, R. S. (2003), "Artificial Neural Network Investigation of the Structural Group Contribution Method for Predicting Pure Components Auto Ignition Temperature", *Industrial and Engineering Chemistry Research,* Vol. 42 No. 22, pp. 5708-5714.

API (1987), *Technical Data book - Petroleum Refining*. The American Petroleum Institute, Washington, DC.

ASTM (1985), "ASTM E681-85 Standard Test Method for Concentration Limits of Flammability of Chemicals (Vapors and Gases)", ASTM International, West Conshohocken, PA, www.astm.org

ASTM (2001), *Chemical Thermodynamics and Energy Release, CHETAH V7.3;* American Society for Testing and Materials, ASTM International, West Conshohocken, PA, www.astm.org.

Bunz A. P., Braun, B. andJanowsky R. (1999), "Quantitative structure-property relationships and neural networks: correlation and prediction of physical properties of pure components and mixtures from molecular structure', *Fluid Phase Equilibria,* Vol. 158, pp. 367-374.

Dewan, A. (2006), "Experimental Project Results from the Design Institute for Physical Properties (DIPPR) of the American Institute of Chemical Engineers. 8", J Chemical Engineering Data, Vol. 51 No. 6, pp. 1953 – 1953.

Elkamel A., Al-Ajmi, A. andFahim, M. (1999), "Modeling the Hydrocracking Process Using Artificial Neural Networks', *Petroleum Science and Technology,* Vol. 17 No. 9-10, pp. 931-954.

Gharagheizi, F. (2009), "Prediction of upper flammability limit percent of pure compounds from their molecular structure", *J. Hazardous Materials,* Vol. 167 No. 1-3, pp. 507-510.

Gharagheizi, F. (2010), "Chemical Structure-Based Model for Estimation of the Upper Flammability Limit of Pure Compounds", *Energy and Fuels,* Vol. 24No. 7, pp. 3867-3871.

Gharagheizi, F., Ilani-Kashkouli, P. andMohammadi, A. H. (2012), "Corresponding States Method for Estimation of Upper Flammability Limit Temperature of Chemical Compounds", *Industrial and Engineering Chemistry Research,*Vol. 51 No. 17, pp. 6265-6269.

Ismail, A., Soliman, M. S. andFahim, M. A. (1996), "Prediction of the viscosity of heavy petroleum fractions and crude oils by neural networks', *Sekiyu Gakkashi,* Vol. 39 No. 6, pp. 383-388.

Joback, K. G. (1984), Thesis, Massachusetts Institute of Technology, Cambridge Massachusetts.

Lasdon, L. S., Waren, A. D., Jain, A. and Rather, M. (1978), "Design and Testing of a Generalized Reduced Gradient Code for Nonlinear Programming". *ACM,* Vol.4 No. 1, pp. 34-50.

Lazzus, J. A. (2011), "Neural network/Particle swarm method to predict flammability limits in air of organic compounds", *ThermochimicaActa,*Vol. 512 No. 1-2, pp. 150-156.

Lee, M. J. and Chen, J. T. (1993), "Fluid Property Prediction with the Aid of Neural Networks", *Industrial and Engineering Chemistry Research,* Vol. 32No. 5, pp. 995-997.

Lippmann, R. P. (1987), "An Introduction to Computing with Neural Nets. *IEEE ASSP Magazine,* Vol. 3 No. 2, pp. 4-22.

Molecular Knowledge Systems, Inc. (1998), *Cranium V1.0.3, Property Estimation Software* (http://www.molknow.com/).

Pan, Y., Jiang, J. and Wang, Z. (2007), "Quantitative Structure-Property Relationship Studies for Predicting Flash Points of Alkanes Using Group Bond Contribution Method with Back-propagation Neural Networks", *J Hazardous Materials,* Vol. 147 No. 1-2, pp. 424-430.

Patel, S. J., Ng, D. andMannan, M. S. (2009), "QSPR Flash Point Prediction of Solvents Using Topological Indices for Application in Computer Aided Molecular Design", *Industrial Engineering Chemistry Research,* Vol. 48No. 15, pp. 7378–7387.

Reid, R. C., Prausnitz, J. M. and Polling, B. E. (1987), *The properties of gases and liquids*, Hill; New York, N.Y.

Seaton, W. H., Freedman, E. andTreweek, D. N. (1974), "*CHETAH-The ASTM Chemical Thermodynamic and Energy Release Evaluation Program, ASTM DS 51"*, American Society for Testing and Materials, ASTM International, West Conshohocken, PA, www.astm.org

Sheldon, M. (1984), "Study of the Flammability Limits of Gases and Vapors", *Fire Prevention,* Vol. 14, pp. 23-31.

The MathWorks Inc. (2001), *MATLAB V6.1, Neural Network Toolbox* (http://www.mathworks.com/).

Widrow, B. and Lehr, M. A. (1990), "30 Years of Adaptive Neural Networks: Perceptron, Madaline, and Backpropagation', *Proc. IEEE,*Vol. 78 No. 9, pp. 1415-1442.

--0--