

Prediction of Standard Heat of Combustion using Two-Step Regression

Nor Alafiza Yunus^{*,a,b}, Nik Nurul Norfatiha N Mohd Zahari^a

^aDepartment of Chemical Engineering, Faculty of Chemical and Energy Engineering, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia

^bProcess Systems Engineering Centre (PROSPECT), Research Institute of Sustainable Environment (RISE), Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia
alafiza@utm.my

Heat of combustion is a thermochemical property that is used for assessing the heating value of solid and liquid fuels as well as the calorific value of food and supplements. It is also used to identify fire hazards of hazardous materials. Heat of combustion has many applications across diverse areas including in jet fuel and propellant formulations, the disposal of combustible waste, the study of foods and supplements for humans and animals, as well as in ecological studies. This study proposes a simple and predictive model for predicting standard heat of combustion. This model was developed using a group contribution approach. The group contribution method represents chemicals according to 220 first-order and 130 second-order groups. The first-order groups are simple groups that describe a wide variety of chemicals, whereas the second-order groups describe polyfunctional compounds and are used to differentiate between isomers. In this study, 680 experimental data points comprising the standard heat of combustion for pure chemicals were collected from open literature. This data set represents various types of groups. The group contributions were regressed using linear regression in MATLAB, yielding an R^2 value of 0.9993 with SD, AAE, and ARE values of 71.9892, 53.1008, and 4.6162. The proposed model was found to be predictive and capable of predicting the heat of combustion of various chemicals, which are not only limited to hydrocarbons but also include chemicals that contain groups of alcohol, ester, ether, amine, amide, aromatic, halogen, and sulfur.

1. Introduction

The standard net heat of combustion (ΔH °C) can be defined as an increase in enthalpy when a substance in its standard state (298.15 K and 1 atm) undergoes complete combustion with oxygen. In particular, it is the amount of heat released when a given amount (usually 1 mole) of a combustible pure substance is burned with oxygen to form an incombustible product (e.g., water and carbon dioxide), with both reactants and products at 25 °C and 1 atm (Felder and Rousseau, 2005). Albahri (2013) defines the standard net heat of combustion as a measurement of energy attained from fuel. Knowledge of this thermodynamic property is very important due to its wide applications. Heat of combustion can be used for predicting the performance of explosive and propellant formulations (Albahri, 2014). This measured value is also essential when considering the thermal efficiency of equipment for producing either heat or power. Heat of combustion is also a measure of the energy that is available in fuel. It is used to compare the heating values of fuel, as the fuel that produces a greater amount of heat for a given cost is the more economic choice (Albahri, 2014). Heat of combustion is also applied in measuring the reactive heat in many chemical engineering processes such as the hydrogenation and dehydrogenation of hydrocarbon (Pan et al., 2011).

Determination the heat of combustion using experimental techniques is time-consuming and expensive because pure material and complete reproducible combustion cannot be easily obtained (Keshavarz et al., 2011). An efficient method is required to obtain the heat of combustion of any substances, so that more time and resources can be spent on product manufacturing. The main objective of this study is to propose a model-based method to predict heat of combustion using a group contribution approach.

Several modelling methods for estimation of heat of combustion have been discussed in the literature. Gharagheizi (2008) proposed standard heat of combustion prediction using a Quantitative Structure-Property Relationship (QSPR) for a dataset consisting 1,714 pure chemicals. This method introduced four-parameter empirical correlations, which are the sum of atomic van der Waals volumes, the number of carbon atoms, Broto-moreau autocorrelation of topological structure, and the Eigenvalue sum from an electronegativity weighted distance matrix. Pan et al. (2011) also developed a four-parameter correlation that used 1,322 compounds as input parameters to predict the heat of combustion of organic compounds. Although both methods were fairly accurate with correlation values of 0.995 and 0.991, the molecular descriptors in both methods were not easy to determine and required extensive computational resources.

Keshavarz et al. (2011) provided another method to calculate the gross and net heats of combustion using a group additivity method for important classes of energetic compounds (polynitro arene, polynitro heteroarene, acyclic and cyclic nitramine, and nitrate ester and nitroaliphatic). Gharagheizi et al. (2011) developed a method to predict the standard net heat of combustion (ΔH °C) of pure chemical compounds based on an artificial neural network-group contribution (ANN-GC). The method used 4,590 pure compounds as part of its dataset, and the resulting correlation value of 0.99999 was quite good. Albahri (2013) estimated heat of combustion using a group contribution method. A least square method was used with multivariable nonlinear regression and about 452 pure hydrocarbon substances were included as part of a 32-set atom-type structural group. The proposed method successfully predicted the heat of combustion of hydrocarbon substances with a correlation coefficient of 0.9982.

Constantinuo and Gani (1994) proposed a group contribution method to determine thermodynamic properties (normal boiling point, critical pressure, normal melting point, critical volume, critical temperature, standard Gibbs energy, standard enthalpy of vaporisation at 298 K, and standard enthalpy of formation at 298 K). Estimation was performed at two levels, which include the first-order group as the first level and the second-order group as the second level. Marero and Gani (2001) extended upon this methodology by introducing a higher-level group contribution that considered more complex compounds with a higher number of carbons such as polymeric compounds.

2. Methodology

In this study, the standard heat of combustion was modelled using a group contribution approach. The two-step regression methodology used in this study is illustrated in Figure 1.

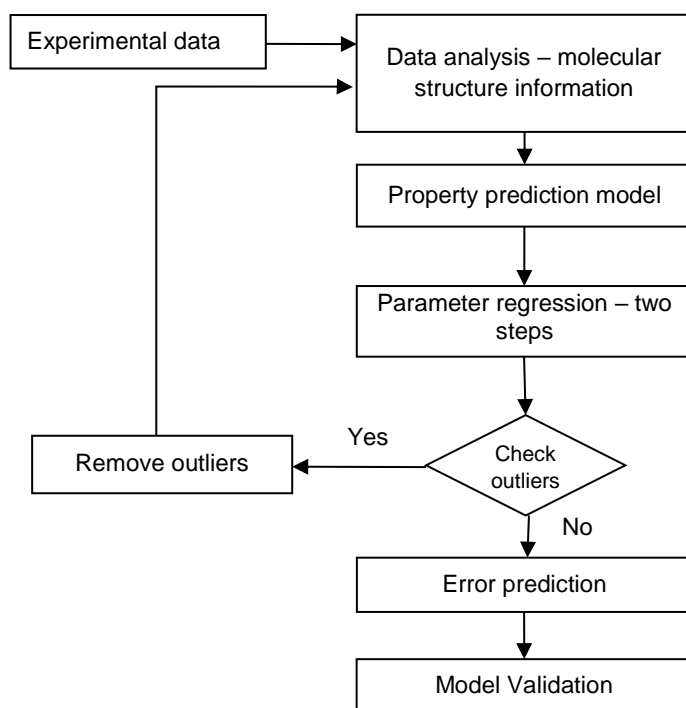


Figure 1: General methodology of the model development in this study

In the experimental data collection for this study, the heat of combustion dataset (at 298 K and atmospheric pressure) was collected from an open database (Linstrom and Mallard, 2011). Data was also collected from

literature review (Gharagheizi, 2008). The 680 data set collected in this study consists compounds from first-order and second-order groups. The process is continued with the identification of groups for the chemicals collected in the first step. Then, 220 groups are categorised as the first-order group such as $-\text{CH}_3$, $-\text{CH}_2$, $-\text{OH}$, and $-\text{COOH}$, while 130 groups make up the second-order group. These groups are defined in detail in Marero and Gani (2001). The third step property prediction model is described as a linear model and is shown in Eq(1).

$$H_c = H_{co} + \sum_i^{NG1} N_i H_{c1i} + w \sum_j^{NG2} M_j H_{c2j} \quad (1)$$

where H_c is heat of combustion, H_{co} is a universal constant, H_{c1i} is the contribution of the first-order group of type- i that occurs N_i times, and H_{c2j} is the contribution of the second-order group of type- j that occurs M_j times. A parameter regression is then carried out in two steps to determine the contribution of each group.

The first step of the regression aims to determine the contribution values of the first-order groups, H_{c1i} and also the universal constant, H_{co} where w is a constant assigned a zero value. In the second stage, w is set to unity where H_{c2j} is determined via regression using the contribution values of the first-order groups, H_{c1i} and H_{co} , already obtained from the previous step. Regression of the dataset is done based on the linear regression of the Levenberg Marquardt curve fitting in MATLAB. After all the contribution values are obtained, heat of combustion (estimated) versus heat of combustion (experimental) is plotted.

The fourth stage involves the analysis of regression results using statistical analysis to obtain the value of standard deviation (SD), relative deviation (RD), average absolute error (AAE), and average relative error (ARE) using Eq(2) - (5), as outlined below:

$$SD = \sqrt{\frac{\sum_i (\zeta_i^{\text{pred}} - \zeta_i^{\text{exp}})^2}{N}} \quad (2)$$

$$RD = \frac{|\zeta_i^{\text{pred}} - \zeta_i^{\text{exp}}|}{\zeta_i^{\text{pred}}} \times 100 \quad (3)$$

$$AAE = \frac{\sum_i |\zeta_i^{\text{pred}} - \zeta_i^{\text{exp}}|}{N} \quad (4)$$

$$ARE = \frac{\sum_i RD_i}{N} \quad (5)$$

The last stage is the model verification. The model was applied using 30 chemicals that were collected separately from a data set of 680 chemicals and then compared with the estimated and experimental values of heat of combustion. A set of additional 30 compounds data test points were tested to prove the effectiveness of the model in estimating the value of heat of combustion.

3. Results and Discussion

The groups of data collected in the data collection stage were analysed first to identify the first-order groups and second-order groups of each data. The analysis of data was done using ProPred in ICAS software. The collected data was then regressed to obtain the contribution values of each group involved, where the first-order group consists 220 groups and the second-order group makes up 130 groups. The regression step starts with first-order group regression followed by second-order group regression.

After the contribution values of the first-order groups and the universal constant are obtained, the value of the predicted heat of combustion is computed. The first result of heat of combustion estimated versus experimental for the 680 chemicals obtained from the first regression process is shown in Figure 2.

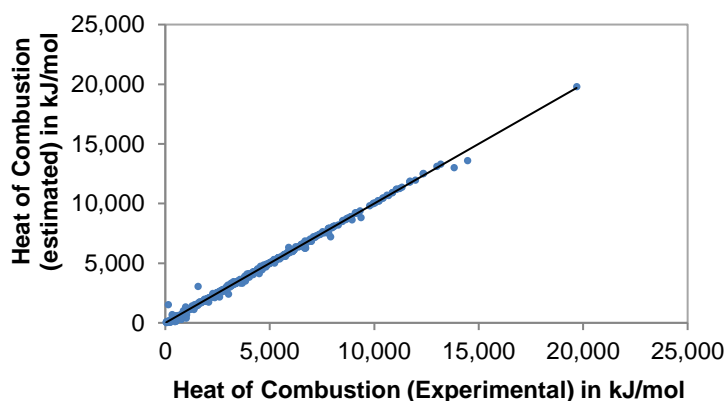


Figure 2: Predicted versus experimental heat of combustion data for 680 data sets in the first-stage regression

The result obtained in Figure 2 shows that some data deviated from the experimental data. The R^2 computed from this regression is 0.9972, which is slightly lower than the unity value. From the first regression step, there are some significant outliers that can be spotted. All these outliers are then identified and analysed for their impact on parameter regression. These outliers can skew the regression results because these outlier data is extensively different than the majority data. About 19 datasets were thus inspected and removed. The parameter regression was repeated for the remaining 661 data sets after the outliers were removed to obtain the actual contribution values for the first-order group, the universal constant, and the predicted heat of combustion. Figure 3 shows the estimated heat of combustion versus experimental heat of combustion after the outlier data was removed.

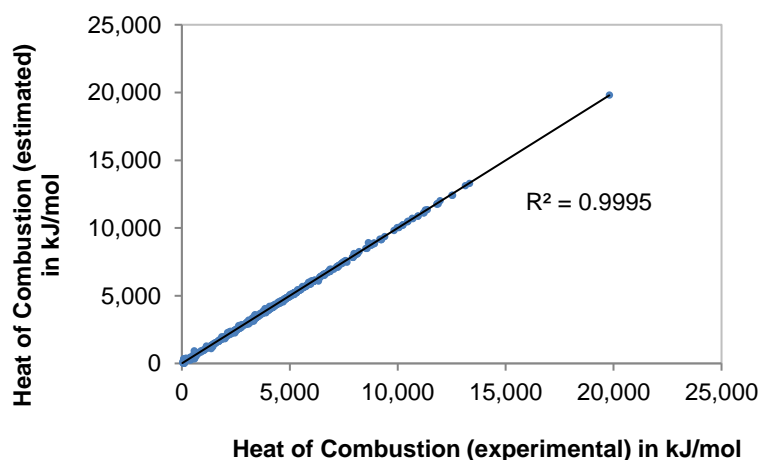


Figure 3: Predicted versus experimental heat of combustion data after removing outliers in the first-stage regression

The graph in Figure 3 shows the good fit between heat combustion and the experimental data after a few outliers were removed. From the regression of 661 data sets after removal of the outliers, the value of R^2 obtained is 0.9995, which is close to unity. In the second stage of regression, the obtained H_{CO} and the first order contribution values, $H_{C_{1i}}$ were substituted in Eq(1), and the contribution values for the second-order groups were regressed using the latest collection of data. Figure 4 shows the comparison between the predicted versus experimental heat of combustion.

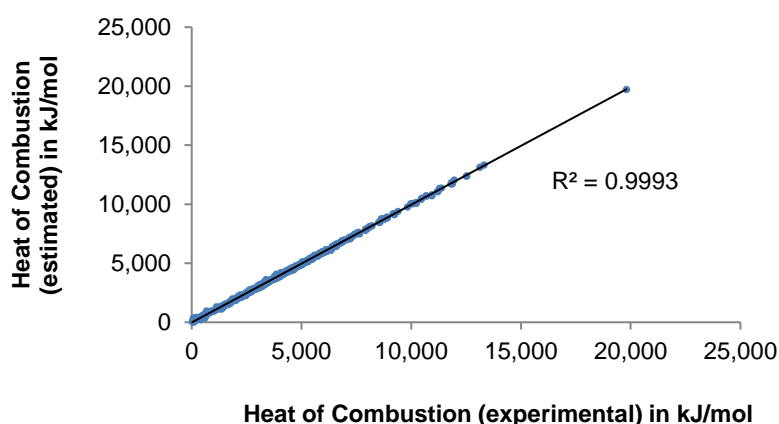


Figure 4: Predicted versus experimental heat of combustion data for the second-stage regression

Figure 4 shows that the result of the predicted heat of combustion fits well with the experimental heat of combustion. The second-step regression should give more accurate results than the first-step regression because it improves upon the overall contribution of the heating value. The R^2 value for the second step is found to be slightly lower than the first step. The accuracy of the model was analysed using statistical analysis through the calculation of SD, RD, AAE, and ARE using Eq(2) - (5). Table 1 shows the statistical analysis result for the second-stage regression of the prediction of heat of combustion.

From Table 1, the statistical analysis shows the computed value of R^2 to be 0.9993, which is still close to unity. The values of AAE, ARE, and SD obtained from the second-step regression process could be considered to be small and within the acceptable range.

Table 1: Statistical analysis for the parameter regression of heat of combustion

Statistical analysis	Value
Coefficient of determination, R^2	0.9993
Average absolute error, AAE	53.1008
Average relative error, ARE	4.6162
Standard deviation, SD	71.9892

The developed model of heat of combustion is then validated using a set of test data consisting 30 compounds. These compounds were collected separately and are a different set from the training set. The data was analysed to identify group and group occurrences. Then, the heat of combustion for all compounds in the test set was estimated using the group contribution parameters obtained. The predicted and experimental heat of combustion for a 30-data test set is plotted in Figure 5.

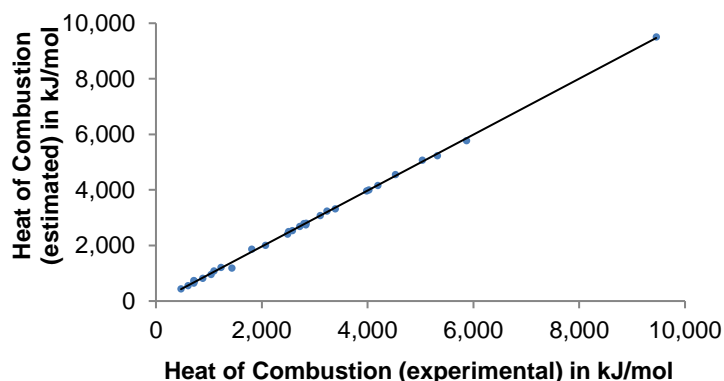


Figure 5: Predicted heat of combustion for 30 components of the data test

From Figure 5, the predicted result of most of the 30 compounds in the data test set in the graph shows a good fit between the predicted and experimental value. From this result, it can be concluded that the model can

successfully predict the value of heat of combustion for any chemical. There are still a few chemicals that show a significant difference in predicted versus experimental heat of combustion. This could be due to some of the group occurrences for that compound, which are still lacking in the data collection, causing the contribution value for that group to be less accurate. More data should be added to the collection to increase the number of group occurrences for the first-order groups and second-order groups.

4. Conclusions

In conclusion, a heat of combustion model was developed in this study to predict the heat of combustion of a select group of chemicals, so as to reduce time and resource consumption when determining heat of combustion. The proposed model used a group contribution method, in which the collected data was first analysed for occurrences of first-order groups and second-order groups and then regressed to obtain their respective contribution values. This regression was done in two steps; the contribution values of the first-order groups and universal constant were obtained in the first step and the contribution values of the second-order groups in the second. In this study, the model successfully determined the heat of combustion of 30 compounds from a test data set. The model was also able to predict the heat of combustion value with small differences when compared with the experimental value. It is therefore proven that the model in this study can predict the heat of combustion of various families of compounds. In future work, it is recommended that an uncertainty analysis be conducted to show the confidence level of the estimated values in this study.

Acknowledgments

The authors would like to thank the Ministry of Higher Education Malaysia and Universiti Teknologi Malaysia (UTM) for providing the research fund for this study (Vote No. Q.J130000.2744.01K93).

Reference

- Albahri T.A., 2013, Method for predicting the standard net heat of combustion for pure hydrocarbons from their molecular structure, *Energy Conversion and Management* 76, 1143-1149
- Albahri T.A., 2014, Accurate prediction of the standard net heat of combustion from molecular structure, *Journal of Loss Prevention in the Process Industries* 32, 377-386
- Constantinou L., Gani R., 1994, New group contribution method for estimating properties of pure compounds, *AIChE Journal* 40 (10), 1697-1710
- Felder R.M., Rousseau R.W., 2005, *Elementary Principles of Chemical Process*, 3rd ed., John Wiley & Sons, Inc, Hoboken NJ, United States.
- Gharagheizi F., 2008, Simple equation for prediction of net heat of combustion of pure chemicals, *Chemometrics and Intelligent Laboratory Systems* 91 (2), 177-180
- Gharagheizi F., Mirkhani S.A., Tofangchi Mahyari A., 2011, Prediction of standard enthalpy of combustion of pure compounds using a very accurate group-contribution-based method, *Energy & Fuels* 25 (6), 2651-2654
- Keshavarz M.H., Saatluo B.E., Hassanzadeh A., 2011, A new method for predicting the heats of combustion of polynitro arene, polynitro heteroarene, acyclic and cyclic nitramine, nitrate ester and nitroaliphatic compounds, *Journal Hazard Mater* 185 (2-3), 1086-1106
- Linstrom P.J., Mallard W.G., 2001, NIST chemistry WebBook, NIST standard reference database number 69, National Institute of Standards and Technology, Gaithersburg MD 20899, USA <www.webbook.nist.gov> accessed 20.09.2015
- Marrero J., Gani R., 2001, Group-contribution based estimation of pure component properties, *Fluid Phase Equilibria* 183-184, 183-208
- Pan Y., Jiang J.C., Wang R., Jiang J.J., 2011, Predicting the net heat of combustion of organic compounds from molecular structures based on ant colony optimization, *Journal of Loss Prevention in the Process Industries* 24 (1), 85-89
- Stahmer K.W., Gerhold M., 2016, The relationship between the explosion indices of dispersed dust and particle surface area and the heat of combustion, *Chemical Engineering Transactions* 48, 295-300