Available online www.jocpr.com

Journal of Chemical and Pharmaceutical Research, 2014, 6(4):880-887

Research Article

ISSN : 0975-7384
CODEN(USA) : JCPRC5

An improved approach to the determination of unknown organic compounds by measuring combustion heat based on non-linear regression analysis

Zhi Yang[1], Hao Li[2], Senyi Mao[1], Xifeng Liu[3], Peng Wang[*1] and Mei Hong[*1]

[1]College of Software Engineering, Sichuan University, Chengdu, Sichuan, China
[2]College of Chemistry, Sichuan University, Chengdu, Sichuan, China
[3]Department of Chemistry, Michigan State University, East Lansing, Michigan, United States

---

ABSTRACT

A novel method is proposed to determine the unknown organic compounds by measuring the combustion heat based on non-linear regression analysis. According to the combustion heat determined using oxygen bomb calorimeter, the equation of determination can be transformed into the determination of Q/M value. After the Q/M value is calculated by means of the equation, we used nonlinear least squares data-fitting method to obtain a best fit curve and to predict Q/M value. Levenberg-Marquardt algorithm is used during the fitting process. On the basis of the accomplishments above, we designed software called Organic Predictor that can quickly and easily determine the unknown organic compounds.

Keywords: Unknown organic compounds, *Q/M* value, combustion heat, non-linear regression; Levenberg-Marquardt algorithm, database system, organic predictor.

INTRODUCTION

Combustion heat [1-5] is the energy released in the form of heat when a compound undergoes complete combustion with oxygen under standard conditions. In the research work, the combustion heat of the organic compounds is mainly measured with the oxygen bomb calorimeter. Based on the previous studies [1], a new approach of determining the structure of the organic compounds by the measurement of heat combustion was created. On the basis of the new approach, we replaced the method of polynomial fitting by the non-linear optimization [6], and built a demonstrate program with a database holding all experimental data, equations and compound information. Under the circumstance of knowing the category of the analyte and determining that it is a pure substance, to determine the unknown organic compound, we can measure the combustion heat with a oxygen bomb calorimeter and obtain the *Q/M* value by the transformed equation [1]. Inputting the *Q/M* value to the demonstration software, the possible structure of the analyte can be revealed on the software interface.

Heat of combustion, as one of the important properties of organic compounds, also has inevitable relevance to the corresponding structure. In our previous studies, we created a method to analyse the unknown organic compound without the routine analytic instrument [1], the unknown organic pure substance is determined by the oxygen bomb calorimeter with the acquired value of *Q/M*, where *Q* represents the combustion heat of the analyte, *M* represents the relative molecular mass of the analyte. Recently, some researchers paid their attention to the field of electrotopological state index in prediction of liquid hydrocarbon combustion heat [8-9]. Besides, some researchers have made a deep study on the group contribution method to calculating combustion heat of liquid alkanes [9-14].

Linear regression is a more powerful and simpler method for analyzing data described by linear models. However, researchers often have problems modeled by non-linear parameters. Therefore, non-linear regression techniques must be used in spite of considerable complexity [18]. Non-linear regression is a form of regression analysis in

which observational data are modeled by a function which is a nonlinear combination of the model parameters and depends on one or more independent variables. The data are fitted by a method of successive approximations. Mathematically, a non-linear regression problem is often transformed into a least-squares problem. There are many methods developed for solving non-linear least squares problems, for instance, the Gauss-Newton method, the Levenberg-Marquardt method, and Powell's Dog Leg method [15, 19-20].

Determination

The most common and effective way of determining the combustion heat is utilizing the oxygen bomb calorimeter. The main mathematical formula is showed as follows:

$$mQ / M + \varepsilon b + qc = K\Delta T. \tag{1}$$

Where Q represents the molar constant-volume combustion heat of the determinand; m represents the mass of the determinand; M represents the molar mass of the determinand; ε represents the heat value of the Ignition wire; b represents the consumption quality of the Ignition wire; q represents the heat value of the combustion of cotton; c represents the consumption quality of the cotton; K represents the constant of the oxygen bomb calorimeter; ΔT represents the change value of the system temperature.

Transposing the equation (1), we can obtain:

$$Q / M = (K\Delta T - \varepsilon b - qc) / m \tag{2}$$

According to the equation Eq. 2, it is obvious that the right of the equation has nothing to do with the determinand. Therefore, we only need to take the left of the equation into consideration. Since the fractional error of the oxygen bomb calorimeter with the correct operation is less than 0.2%, we can obtain the precise value of *Q/M*. Through the study of the *Q/M* value's regulation, researchers found that every single organic compound owns a single value. Therefore, the unknown organic compound can be determined by the *Q/M* value. With the help of the database, we can find the matched value and determine what the unknown organic matter is.

Non-Linear Least Squares Fitting

After a series of experimental data were obtained and input to the software we built, the equation predicting the *Q/M* value of the organic compound is automatically calculated with the approach called non-linear least squares fitting.

Given the carbon atom number of certain homolog x_i, we used an exponential equation to predict the corresponding *Q/M* value.

$$y = Ae^{\lambda x_i} + b \tag{3}$$

Given a series of observed values y_i', which is obtained in experiment, the Prediction Error Sum of Square is given as follows:

$$PESS = \sum_{i=0}^{n} (y'_i - y_i)^2 \tag{4}$$

To find the curve best fits the experimental data, we should determine the value of parameters in Eq. 3 that have the minimum PESS.

Levenberg-Marquardt Algorithm

Unlike linear fitting or polynomial fitting, the parameters in the exponential equation cannot be precisely calculated, but gradually closed in from an initial guess of all the parameters, using the Levenberg-Marquardt Algorithm [16]. According to "Methods for non-linear least squares problem" Section 3.2 [15], the pseudo code of the algorithm is given as follows:

Algorithm. Levenberg - Marquardt method
begin
$k := 0;\quad \nu := 2;\quad \mathrm{x} := \mathrm{x}_0$
$\mathbf{A} := \mathbf{J}(\mathrm{x})^{\mathrm{T}}\mathbf{J}(\mathrm{x});\quad \mathbf{g} := \mathbf{J}(\mathrm{x})^{\mathrm{T}}\,\mathrm{f}(\mathrm{x})$
$found := (\| \mathbf{g} \|_\infty \le \varepsilon_1);\quad \mu := \tau * \max\{a_{ii}\}$
while (**not** *found*) **and** ($\mathrm{k} < \mathrm{k}_{max}$)
$k := k + 1;\quad \mathrm{Solve}(\mathbf{A}+\mu\mathbf{I})\mathbf{h}_{\mathrm{lm}} = -\mathbf{g}$
if $\| \mathbf{h}_{\mathrm{lm}} \| \le \varepsilon_2(\| \mathrm{x} \| + \varepsilon_2)$
$found := \mathbf{true}$
else
$\mathrm{x}_{\mathrm{new}} := \mathrm{x} + \mathbf{h}_{\mathrm{lm}}$
$\tilde{n} := (F(\mathrm{x}) - F(\mathrm{x}_{\mathrm{new}})) / (L(\mathbf{0}) - L(\mathbf{h}_{\mathrm{lm}}))$
if $\tilde{n} > 0$ {step acceptable}
$\mathrm{x} := \mathrm{x}_{\mathrm{new}}$
$\mathbf{A} := \mathbf{J}(\mathrm{x})^{\mathrm{T}}\mathbf{J}(\mathrm{x});\quad \mathbf{g} := \mathbf{J}(\mathrm{x})^{\mathrm{T}}\,\mathrm{f}(\mathrm{x})$
$found := (\| \mathbf{g} \|_\infty \le \varepsilon_1)$
$\mu := \mu * \max\{\frac{1}{3}, 1 - (2\tilde{n} - 1)^3\};\quad \nu := 2$
else
$\mu := \mu * \nu \qquad \nu := 2 * \nu$
end

Fig.1: Levenberg-Marquardt Algorithm

The complete definition of variables and functions are given in the referenced paper.

The Levenberg-Marquardt Algorithm is considered a puissant algorithm, which means the parameters can still be properly calculated even if the initial guess is significantly distant from the correct values of the parameters.

EXPERIMENTAL SECTION

Data processing

With the Levenberg-Marquardt Algorithm, the parameters of the best fit exponential equation of a certain kind of homolog can be calculated by inputting several carbon atom numbers and corresponding *Q/M* values.

In our previous study, we used a polynomial equation as the fitting curve, in which case the error of prediction becomes so large that it is completely useless for determination with the increase of carbon number. However, there is a significant improvement after the Levenberg-Marquardt Algorithm is performed on an exponential equation.

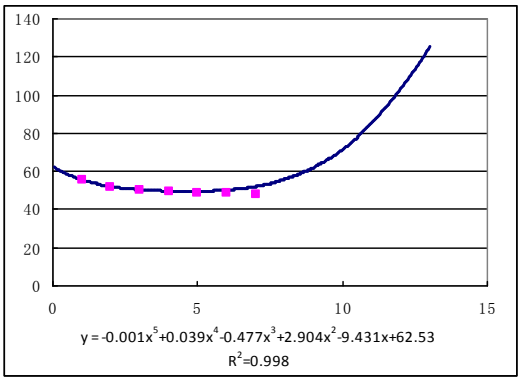
The *Q/M* value of alkane, alkene and alkyne are shown as follows (the values of combustion heat were extracted from WINGCH Database [21] or calculated by a novel calculation method [11]):

Table 1 The *Q/M* values of alkane, alkene and alkyne

Carbon numbers	Q/M value of alkane	Q/M value of alkene	Q/M value of alkyne
1	55.59	--	--
2	51.96	50.34	49.94
3	50.41	48.96	48.40
4	49.58	48.50	48.06
5	48.70	48.18	47.83
6	48.36	47.48	44.98
7	48.12	47.56	44.78
8	47.94	47.45	44.65
9	47.80	44.37	44.52
10	47.68	44.31	44.43
11	47.54	44.25	44.36
12	46.95	44.20	44.30
13	46.90	44.16	44.26
14	46.88	44.12	44.23
15	46.84	44.08	44.19
16	46.81	44.05	44.15
17	46.78	44.03	44.11
18	46.76	44.01	44.08
19	46.74	43.38	44.05
20	46.72	43.36	44.03

Shown on the Table 1, the difference between two *Q/M* values in the same category is reducing. And the change of it is slow and regular. Thus, it is unlikely that the *Q/M* value would increase rapidly as the carbon number increases.

With the data of *Q/M*, a best fit curve of each column can be calculated using the L-M algorithm. Comparing with the linear regression method that was used in the previous study, the figures are shown as follows:



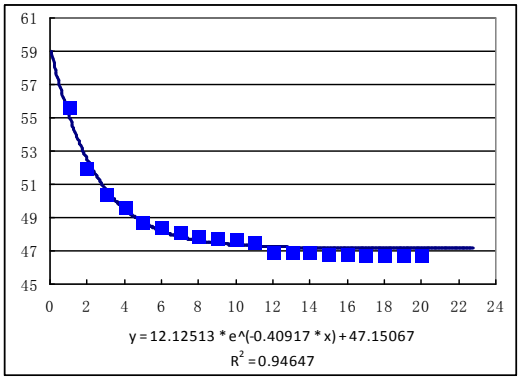
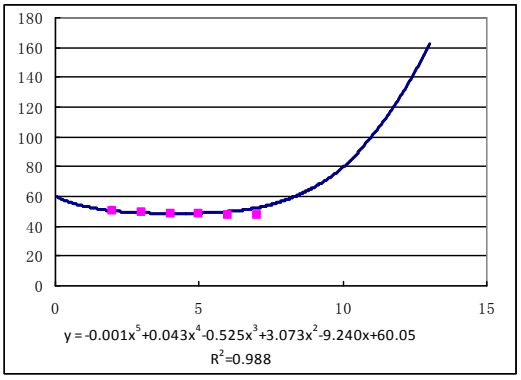


Fig.2 The polynomial *Q/M* curve of alkane **Fig.3 The exponential *Q/M* curve of alkane**



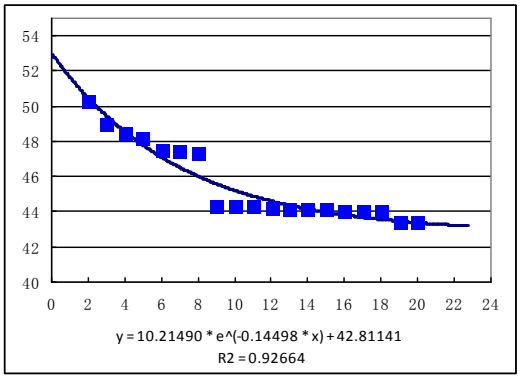
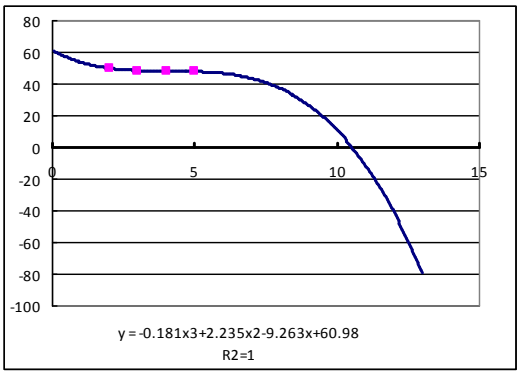


Fig.4 The polynomial *Q/M* curve of alkene **Fig.5 The exponential *Q/M* curve of alkene.**



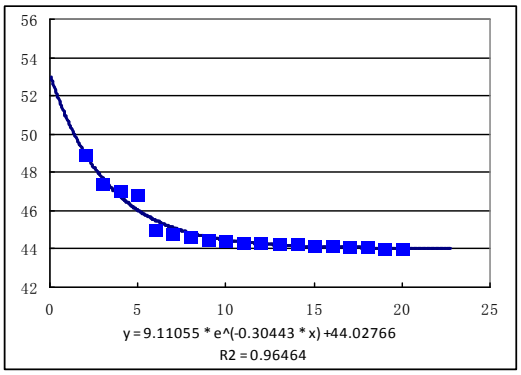


Fig.6 The polynomial *Q/M* curve of alkyne **Fig.7 The exponential *Q/M* curve of alkyne**

Particularly, in Fig.5, a part of the *Q/M* value prediction is below zero, which is impossible because the combustion heat in these cases and the relative molecular mass are greater than zero. Comparing to polynomial fitting, which is used in our previous study, the best fit curve using exponential fitting is visibly improved in the dimension of curve trend.

Database

To store the data needed in the process of prediction, a database with three tables is built. The structure of the database is shown in the form of Enhanced Entity Relation Diagram as follows:

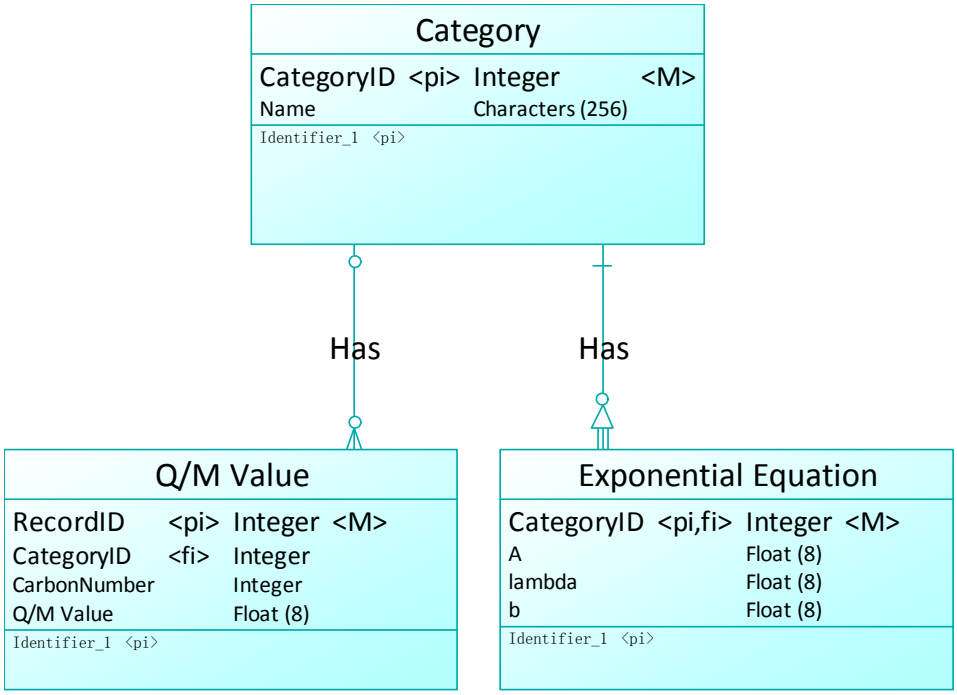


Fig.8 Enhanced Entity Relation Diagram of database

The database contains three main tables. The category table stores the information of organic compound categories, including a unique index 'CategoryID', and the name of corresponding category. The *Q/M* value table stores the information of one *Q/M* value calculated from a experimental record. Each record contains four information, unique index 'RecordID', corresponding 'CategoryID', carbon number, and *Q/M* value. The 'Exponential Equation ' table stores parameters of the best fit equations. Each record represents one exponential equation in the form of Eq. 3.

When an inquiry is called, the programme will first check the existence of input category. If exists, all experimental record of input category are copied to a temporary table. Afterward, the programme will keep filling the temporary table with predicted *Q/M* values, until the deviation between the predicted value and the input value is increasing. After filling the temporary table, the table is shown on the result dialog as Fig.9.

Measurement & Procedure

Firstly, the determinand should be confirmed to be the pure substance and straight chain configuration, as well as the type of the organic matter. Secondly, the determination would be lighted in the oxygen bomb calorimeter strictly in accordance with the operation of heat combustion determination. Finally, the value of ΔT would be acquired. By putting the ΔT and the other data into the equation (1), the *Q/M* value of the determinand would be obtained. To make sure the accuracy of determination, operator may determine the same analyte 2 to 4 times, and calculate the average as the inputted *Q/M* value.

After obtaining the value of *Q/M*, operators can input the *Q/M* value and the homolog category of the determinand into the blank of the software. With clicking the button "Inquire", the probable organic compounds would be listed on the surface with the order of probability.

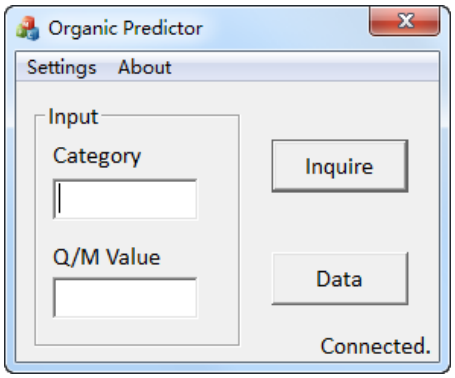


Fig.9 Surface of the Organic Predictor

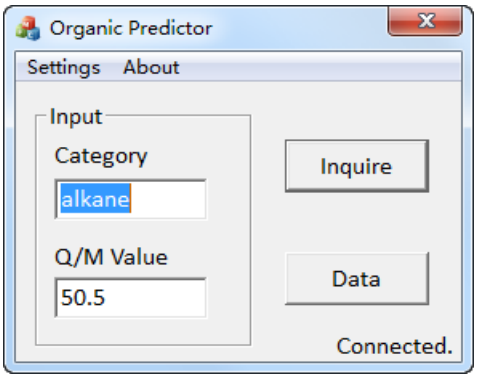


Fig.10 Input type and *Q/M* Value of the determinand

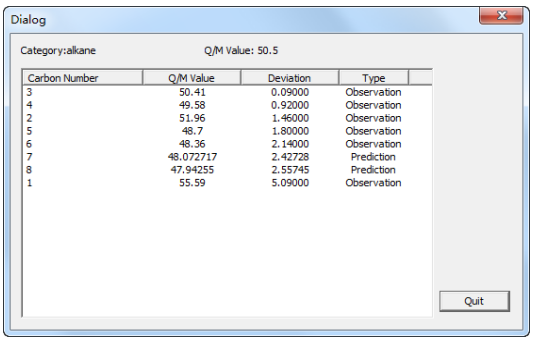


Fig.11 Result of a demonstrate inquiry

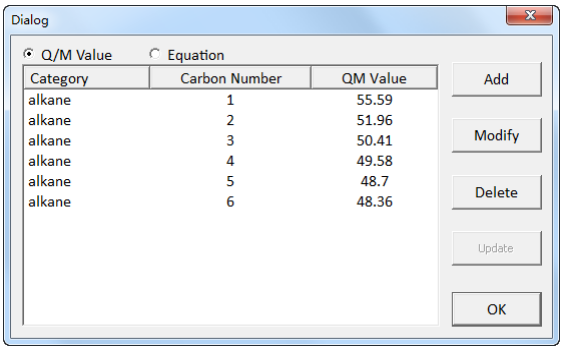


Fig.12 Experimental records of *Q/M* values

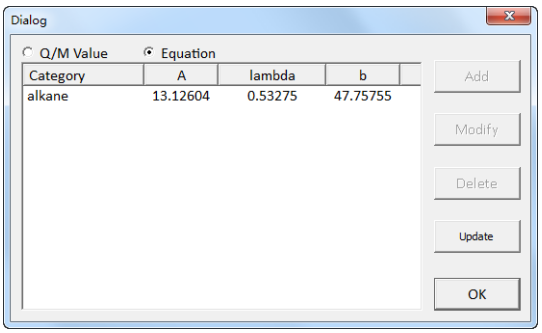


Fig.12 Best fit equation(s)

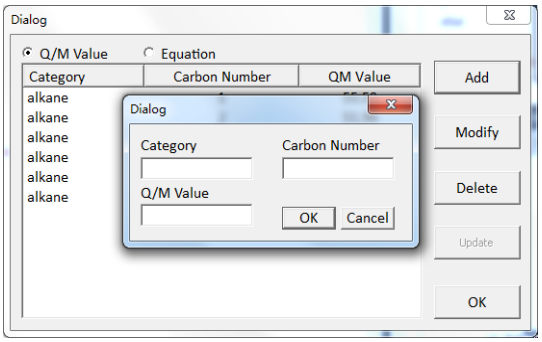


Fig.13 Add a record to database

RESULTS AND DISCUSSION

Accuracy of Determination

The regulation of the change can be obtained through the data analysis of *Q/M*. Each *Q/M* value of homolog has a confirmed value. The different between the adjacent organic compounds in the same homolog are tiny, and it's getting less with the increasing carbon number. To ensure the accuracy of the determination, we modified *Q/M* to be

(Q/M)[10], providing enlarged differences between the adjacent data. Even if the carbon number would be quite large, the difference of the adjacent *Q/M* value would also be clear.

According to the standard operation of the determination of combustion heat [22-27] and the present technique of oxygen bomb calorimeter[28-30], the error of the determination of combustion heat can be controlled within ±0.2KJ/mol~±1.3KJ/mol, the range of difference is corresponded with the standard of the actual determination.

Applications & Limitations

The modified method can be used in the determination of the unknown organic compounds with the confirmed information (straight chain, pure substance, type of the organic matter). Compared to the common methods of the determination, this method is both economized and time-saving. Therefore, it has potential applications on the lab management and the quick classification of the organic substance.

On the other hand, the limitations of the method are obvious: according to our latest studies, the determinand should be in the form of straight chain configuration and the pure substance. In addition, because of the limitation of data, operators should confirm the type of the organic matter before the determination. To eliminate these barriers, we will do further studies.

In the future work, we will first expand the data base through more experiments and adopt a more precise fitting method. Second, we will investigate the differences of the *Q/M* values between the isomers so that operators can determine the non-straight-chain organic compounds. Third, we plan to study the differences of the *Q/M* values between different types of the organic matters, such as the different values between alkane and alkene in order to determine the *Q/M* value without the knowledge of the types of the organic matters. Next, we intend to expand the proposed method to obtain the mixed organic compounds by improving the calculation procedures. The machine learn algorithms will also be used to predict the *Q/M* value and to identify the type of the organic matters given the *Q/M* value.

Download the Software

The demonstrate software is uploaded to the Internet(URL: http://pan.baidu.com/s/1gSi0S). To guarantee the copyright, all the papers use our software need to cite our paper as reference. If the URL given above is expired, please contact us by e-mail. We will be glad to send the software to any interested individual.

Acknowledgements

This work is an extension version of the achievements presented in 4th International Conference on Mechanical Science and Engineering (ICMSE2014)，and the previous work is published in *Journal of Applied Sciences*. We appreciate the great platform organized by International Association Management Science & Engineering Technology (IAMSET).

REFERENCES

[1] H Li, SJ Yang, YW Wang, ZZ Zhao, YZ Li. *Journal of Applied Sciences*, **2013**, 13(23), 5621-5624.
[2] HY Cao , JC Jiang, Y Pan , R Wang , Y Cui. *J. Loss Prevent Proc.*, **2009**, 22(2), 222-227.
[3] A Srisaipet, K Kittiratanapiboon, N Jeyashoke, K Krisnangkura. *KMUTT Research & Development Journal*, **2007**,30(3), 513-522.
[4] JG Burger. *Society of Petroleum Engineers*, **1972**, 12(5), 410-422.
[5] DC Rakopoulos, CD Rakopoulos, RG Papagiannakis, DC Kyritsis. *Fuel*, **2011**, 90(5), 1855–1867.
[6] JJ Moré, BS Garbow, KE Hillstrom. *ACM T. Math. Software*, **1981**, 7(1),17-41.
[8] JC Jiang. *Gas Chemical Industry*, **1996**, 21(6),53-56.
[9] LH Hall, LB Kier. *J. Chem. Inf. Comput. Sci.*,**1995**, 35(4), 1039-1045.
[10]K Wang. *Journal of Ningxia University (NATURAL SCIENCE EDITION)*,**2000**, 21(3),239-241.
[11] J Peng. *Fire Science and Technology*,**1999**,4, 9-11.
[12] J Taskinen, J Yliruusi. *Adv. Drug Deliv. Rev.*, **2003**, 55(2),1163-1183.
[13] TA Albahri. *Chem. Eng. Sci.*, **2003**, 58(2), 3629-3641.
[14]L Türker. *J. MolStruct (Theochem)*, **2004**, 680(2): 29-31.
[15] K Madsen, HB Nielsen, O Tingleff. Methods for Non-Linear Least Squares Problems. Technical University of Denmark, Copenhagen, **1999**;.
[16] J Fan. *J. Comput. Math.*, **2003**, 21(5), 22-28.
[18] GA Milliken. *Technometrics*, **1990**, 32(2), 219-220.
[19] JE Dennis, JJ Moré. *SIAM Review*, **1977**, 19(1), 46-89.
[20] M Al-Baali, R Fletcher. *J. Oper. Res. Soc.*, **1985**, 405-421.

[21] WINGCH Data Base[DB/OL]: http://www.basechem.org/
[22] WD Good , DW Scott , G Waddington. *J. Phys. Chem*. **1956**, 60 (8), 1080–1089.
[23] L Bjellerup. *ACTA Chemica Scandinavica,* **1961**, 12(5),121-140.
[24] HA Skinner and A Snelson. *Trans. Faraday Soc*. **1960**,56(3),1776-1783.
[25] RS Jessup. Precise Measurement of Heat of Combustion with a Bomb Calorimeter, National Bureau of Standards, U.S., **1960**;.
[26] EW Washburn. *J. Res. Nat. Bur. Stand. (US)*, **1933**, 10, 525-58.
[27] WN Hubbard , C Katz , G Waddington. *J. Phys. Chem*.,**1954**, 58(2), 142–152.
[28] RN Walters, SM Hackett, RE Lyon. *Fire Mater*., **2000**, 24(5), 245-252.
[29] SM Akers , LJ Conkle , N Stephanie. *J. Chem. Educ*., **2006**, 83 (2):260-271.
[30] XW Yang, SP Chen, SL Gao, HY Li and QZ Shi. *Instrum. Sci. Technol.*, **2002**, 30 (3), 123-126.