A new approach to interpreting relationship between Rock-Eval S_2 and TOC data for source rock evaluation based on regression analyses

Ali Shekarifard
School of Chemical Engineering, College of Engineering, University of Tehran, Tehran, Iran
e-mail: ashekary@ut.ac.ir

(*received: 05/02/2015 ; accepted: 01/06/2015*)

Abstract

To evaluate the relationship between total organic carbon (TOC) and Rock-Eval S_2 (petroleum potential) of petroleum source rocks, a total of 180 outcrop samples from the black organic matter–rich facies of Mesozoic strata from a locality of southwest of Iran were investigated using Rock-Eval VI pyrolysis and Leco Carbon Analyzer. The linear regression is applied to determine the correlation between Leco TOC and Rock-Eval S_1 and S_2. The accuracy of the proposed model by this method has approximately 95% conformity according to the Rock-Eval S_2 and Leco TOC data (TOC = 0.492 + 0.174 S_2). Then, by using the P value method, it was determined that TOC is a function of S_2 and S_1 only causes the fluctuations. By means of sensitivity analysis of TOC with respect to S_1 and S_2, it was shown that TOC has a linear relationship with S_2 and does not have any noticeable correspondence to S_1. The result of this study can be used to evaluate petroleum potential (S_2) of organic matter–rich facies by using TOC obtained by Leco Carbon Analyzer. For the studied samples from the organic matter-rich facies, organic carbon richness is a quality and quantity index of petroleum potential.

Keywords: *Leco TOC, Organic Matter–Rich Facies, P value, Petroleum Potential index, Rock-Eval VI pyrolysis.*

Introduction

The black organic matter–rich facies from the investigated locality are hosted within a carbonate succession, with the age of Jurassic to Cretaceous.

The Rock-Eval pyrolysis technique is the best and the most practical method for assessing petroleum generation potential of possible source rocks. When applying Rock-Eval pyrolysis, a graph of S_2 (hydrocarbons generated by the pyrolysis, measured in mg HC/g rock) versus TOC (total organic carbon content of the rock by wt%) is used an efficient tool (Langford *et al*., 1990; Yalçın Erik *et al*., 2006). Using this tool, we can obtain the correction for hydrogen index (mg HC/g TOC), determine the type of organic matter present, evaluate the adsorption of hydrocarbon by the rock matrix, process the organic component of different sets of samples, and can determine their petroleum generation potentials.

The amount of organic matter in rocks is described as TOC, and the relative ability of a source rock to generate petroleum is defined by the quality and quantity of its organic matter (Hunt, 1996). In dried sediments, total carbon content is obtained. Leco Carbon Analyzer (Bernie *et al*., 2010) is one of the most important methods evaluating the TOC.

The aim of this article is to evaluate the relationship between petroleum potential (S_1 + S_2) and TOC of Jurassic–Cretaceous organic matter–rich units from a locality of southwest of Iran. The results of this study reveal a correlation between TOC of the Leco Carbon Analyzer and S_1 and S_2 of Rock-Eval pyrolysis of organic matter–rich units using linear regression method, which determines the appropriateness of rejecting the null hypothesis.

First, a brief description of the Rock-Eval pyrolysis is given and then the regression method is described. Using the regression analysis, a model is proposed based on S_2. In this study, we evaluated the effects of two parameters of the model on the response values (Leco TOC).

Methods and Materials

In this study, 180 outcrop samples were collected from the organic matter–rich sediments (black units) of southwest of Iran from Zagros Basin. To measure and understand the relationship between TOC and S_1 and S_2 values and to assess the geochemical characteristics for correlating TOC and S_1 and S_2, the techniques of Rock-Eval VI and Leco analyses were used.

Leco Carbon Analyzer

Total organic carbon content of the samples was determined using an elemental analyzer Leco. As this apparatus burns rocks up to 1100°C, the determination of total carbon is better compared with those calculated by Rock-Eval pyrolysis, especially for carbon-rich (coals) samples. To remove the effect of carbonate content of the samples on the result of TOC, all samples were treated by HCl. The Leco Carbon Analyzer shows that the organic matter–rich facies have 0.48–26.4 wt% TOC.

Rock-Eval pyrolysis

The Rock-Eval VI pyrolysis technique is a rapid efficient method to broadly evaluate the properties of petroleum source rocks (Espitalié *et al*. 1985a, 1985b, 1986). Using Rock-Eval pyrolysis, parameters such as quantity and type of organic matter in a sedimentary rock and the level of organic maturation were obtained. The geochemical parameters obtained by Rock-Eval VI are as follows: i) S_1 fraction (mg HC/g sample; mostly composed of small volatile molecules), in which the amount of hydrocarbon released at 300 °C is measured; ii) the S_2 fraction (mg HC/g; composed in part of larger thermally cracking molecules of hydrocarbons derived from kerogen (e.g., algal cell wall detritus)), in which peak S_2 is the amount of hydrocarbon released during temperature-programmed pyrolysis (300–600°C); and iii) the S_3 fraction (mg CO_2/g; derived from oxygen-containing organic molecules).

Regression analysis

Regression is a highly useful statistical technique for developing a quantitative relationship between a dependent variable (response) and one or more independent variables (factors). The experimental data are used on the pertinent variables to develop a numerical relationship showing the influence of the independent variables on a dependent variable of the system. Throughout engineering, regression may be applied to correlating data in a wide variety of problems ranging from the simple correlation of physical properties to the analysis of a complex industrial system. If the relationship among the pertinent variables is not known from theory, a function may be assumed and fitted to experimental data on the system. Often, a linear function is assumed. In other cases where a linear function does not fit the experimental data, the engineer might try a polynomial or an exponential function. In a large number of cases, theory produces incomplete models. Regression analysis is used in such cases to determine unknown coefficients in a theoretical equation from available experimental data.

Simple Linear Regression

In the simplest case, the proposed functional relationship between two variables is:

$$Y = \beta_0 + \beta_1 X + \varepsilon \qquad (1)$$

In this model, Y is the dependent variable, X is the independent variable, and e is a random error (or residual amount), which is the amount of variation in Y not accounted for by the linear relationship. The parameters β_0 and β_1 are called the regression coefficients that are need to be estimated. The variable X is not a random variable and takes fixed values. It will be assumed that the errors ε are independent and have a normal distribution with mean 0 and variance δ^2, regardless of what fixed value of X is being considered. Taking the expectation of both sides of Equation (2), we have:

$$E(Y) = \beta_0 + \beta_1 X \qquad (2)$$

where the expected value of the errors is zero. For a fixed value of X, the expectation in Equation (2) is usually denoted by:

$$E(Y) = E(Y/X) = \mu_{Y/X}$$

Thus, we can write:

$$E(Y) = E(Y / X) = \mu_{Y/X} = \beta_0 + \beta_1 X \qquad (3)$$

Equation (3) is called the regression of Y on X. The only random variables in Equation (1) are Y and ε. As the ε is normally distributed, the random variable Y has a normal distribution with mean $\mu_{Y/X} = \beta_0 + \beta_1 X$ and variance δ^2. Geometrical representation of the linear regression is shown in Figure 1.

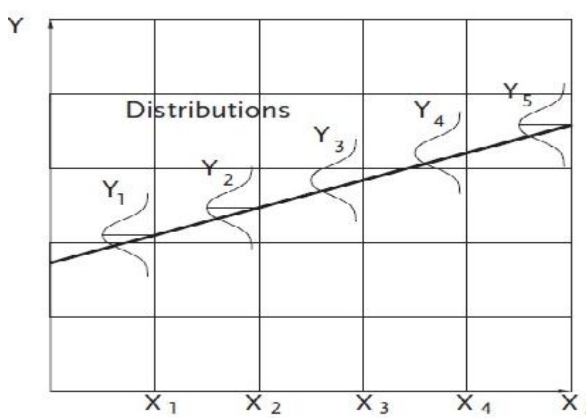

Figure 1. Geometric interpretation of linear regression

To estimate the relationship between Y and X, we have n observations on Y and X, denoted by (X_1,Y_1),

(X_2,Y_2), …, (X_n,Y_n). From Equations (1) and (3), the relationship between Y and X is given as:

$$Y = E(Y/X) + \varepsilon \quad (4)$$

The objective is to estimate β_0 and β_1 and thus $E(Y/X)$ or Y in terms of n observations. If b_0 and b_1 denote estimates of β_0 and β_1, then an estimate of $E(Y)$ is denoted by $\hat{Y} = \hat{E}(Y) = b_0 + b_1 X$.

P value

P value determines the appropriateness of rejecting the null hypothesis. P values range from 0 to 1. The P value is the probability of obtaining a test statistic that is at least as extreme as the calculated value if the null hypothesis is true. Before conducting the analyses, alpha (α) level is set to 0.05. If the P value of a test statistic is less than the alpha value, the null hypothesis is rejected.

Because of their fundamental role in hypothesis testing, P values are used in many areas of statistics, including basic statistics, linear models, reliability, and multivariate analyses among others. The key is to understand what the null and alternative hypotheses represent in each test and then to use the P value to aid in decision making to reject the null hypothesis.

For example, when a 2-sample *t*-test is considered where the difference between the mean strength of steel is tested from two mills based on random samples from each mill, the null hypothesis states that the two population means are equal, whereas the alternative hypothesis states that they are not equal. A P value below the cutoff level suggests that the population means are different.

When regression analyses are performed on steel strength where temperature is one of the explanatory variables, a P value for each regression coefficient is obtained. Here, the default test is to determine if the coefficient for temperature is different from zero. Therefore, the null hypothesis states that the coefficient equals zero, whereas the alternative hypothesis states that it is not equal to zero. A P value below the cutoff level indicates that the coefficient of temperature is significantly different from zero.

The P value is calculated from the observed samples and represents the probability of incorrectly rejecting the null hypothesis when it is actually true. However, it is the probability of obtaining a difference at least as large as the one between the observed value and the hypothesized value through random error alone.

Determination of the model component

In this study, 180 samples were collected from the organic matter–rich units. We will fit a multiple linear regression model:

$$TOC = \beta_0 + \beta_1 S_1 + \beta_2 S_2 + \varepsilon \quad (5)$$

The least squares fit, with the regression coefficients reported to two decimal places, for Leco TOC and petroleum potential (S_1+S_2) of Rock-Eval is as follows:

$$\hat{TOC} = 0.639 + 0.169\,(S_1+S_2) \quad (6)$$

The determination of coefficient, denoted as R^2, indicates how well data points of an experiment fit a model. It provides a measure of the observed outcomes that are replicated in the model (R^2=94.5%), which indicates that 94.5% of experimental data are matched with the regression curve. Adjusted R^2 has been modified for the number of terms in the model. If the model includes unnecessary terms, R^2 can be artificially high. Unlike R^2, adjusted R^2 may be smaller when terms are added to the model. Adjusted R^2 is used to compare models with different numbers of predictors. For the Leco Carbon Analyzer and Rock-Eval data, the adjusted R^2 is 94.4%, which is a decrease of 0.1% (94.6%–94.5%). This model suggests that the variability in Leco TOC increases as S_1+S_2 increases.

The first three columns of Table 1 (10 data with variable content of Leco TOC) present the actual observations of TOC, the predicted or fitted values of $\hat{TOC}_i$, and the residuals. Figure 2 plots the residual values for TOC based on S_1+S_2. Figure 2-a shows the normal probability plot of the residuals, which indicates whether the data are normally distributed, other variables are influencing the response, or outliers exist in the data. Figure 2-b shows the plot of standard residual versus predicted value of $\hat{TOC}_i$, which indicates whether the variance is constant, a nonlinear relationship exists, or outliers exist in the data. Plots of the standard residual versus frequency and observational order that indicate whether the data are skewed or outliers outliers exist in the data and whether there are systematic effects in the data due to time or data collection order are shown in Figures 2-c and 3-b, respectively. More accurately, residual plotting is an an integral part of regression model building. These plots indicate that there is a tendency for the variance of the observed TOC to decrease with the magnitude of TOC.

Sensitivity analysis of TOC with respect to S_1 and S_2

Table 2 lists the estimated coefficients for the predictors. Linear regression examines the relationship between a response and the predictors (Table 2). To determine whether the observed relationship between the response and predictors is statistically significant, we need to:

- Use the first P value (Regression) to analyze whether the regression coefficients are significantly different from zero. If the P value is smaller than a preselected α level, it can be assumed that at least one coefficient is not zero. A commonly used α level is 0.05.
- Use the second P value (Lack of Fit) to determine whether the linear predictors alone are sufficient to explain the variation in Response.

Table 1. Prediction Values, Residuals and other diagnostics

Observations i	TOC_i	Predicted Value $T\hat{O}C_i$	Residual	Standard Residual
1	0.48	0.629	-0.149	-0.11
2	1.260	0.601	0.659	0.47
3	2.340	2.046	0.294	0.21
4	4.880	5.194	-0.314	-0.22
5	8.130	8.272	-0.142	-0.10
6	11.60	14.679	-3.079	-2.16
7	16.70	17.230	-0.530	-0.38
8	18.70	18.017	0.683	0.48
9	23.70	23.301	0.399	0.28
10	26.40	24.35	2.05	1.45

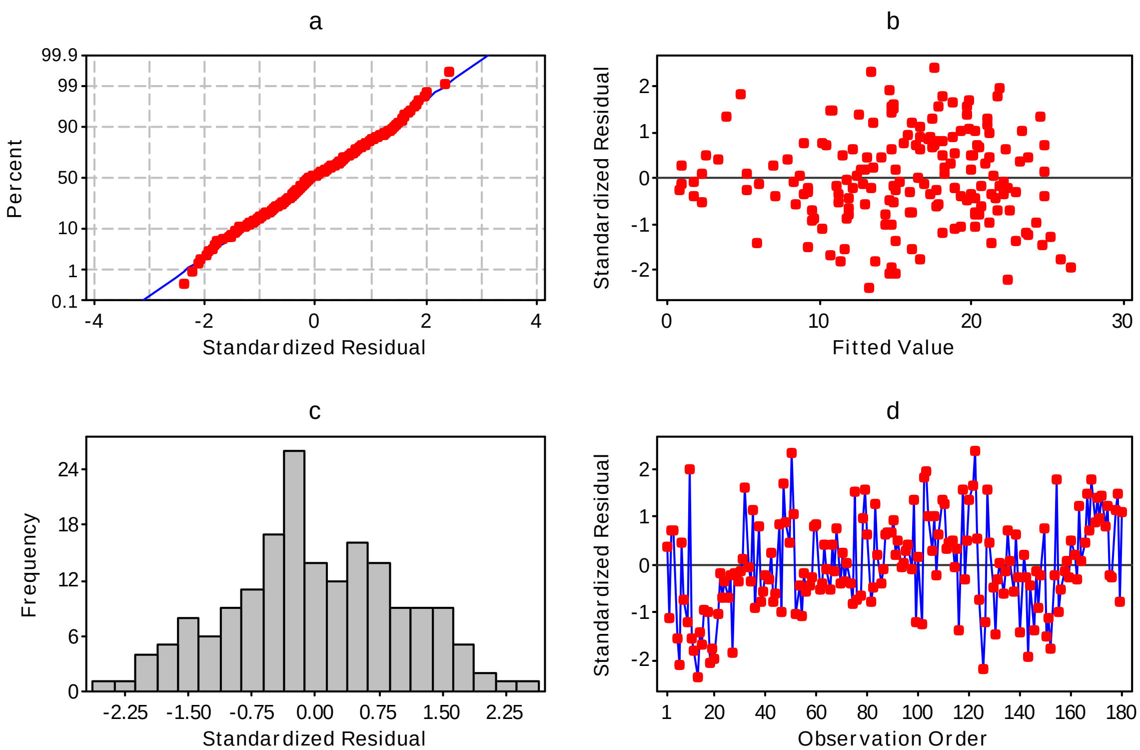

Figure 2. Residual plots for TOC base on S_1 and S_2

If the P value is smaller than a preselected α level, it can be concluded that the linear predictors are not sufficient to explain the variation in response. In this case, higher order terms can be considered. We could include the quadratic terms of the predictors, one at a time, and reanalyze the data.

For the Rock-Eval data, the results can be summarized as follows:

Table 2. Coefficients table

Predictor	Coefficient	P
Constant	0.4168	0.184
S_1	- 0.0934	0.507
S_2	0.177127	0.000

The relationship between the $\bar{TOC}$ and the predictor, S_2 ($P = 0.000$), is significant.

The relationship between the $\bar{TOC}$ and the predictor, S_1 ($P = 0.607$), is not significant because the P value is higher than the preselected α level. In this case, we can refit the model without this predictor, examine the residuals, and then decide whether it should be included. Therefore, new regression model base on TOC and S_2 is as follows:

$$TOC = 0.492 + 0.174\, S_2 \qquad (7)$$

If repeated $\hat{TOC}$ values are observed at certain settings of the predictors, the unexplained variation can be divided variation due to pure error and model inadequacy (Lack of Fit).

In 3-D plot of TOC versus S_1 and S_2 (Fig. 3), which is determined by increasing S_2, the TOC increases, and this progress shows a linear behavior. It should be noted that increasing S_1 causes fluctuations in the rate of TOC, which indicates that S_1 does not have any considerable effect on the TOC.

The S_2 versus TOC plot

The graph of S_2 vs. TOC is an instructive path that displays Rock-Eval and Leco Carbon Analyzer data. The linear regression curve is used on the S_2 vs. TOC graph for calculating hydrogen index (HI) (Langford and Blanc-Valleron, 1990) and shows the petroleum potential and the type of kerogen present. For the organic matter–rich units, the amount of HI is about 550 mg HC/g TOC (pyrolyzable hydrocarbons are about 55% of TOC), which is close to type II kerogen (Fig. 4).

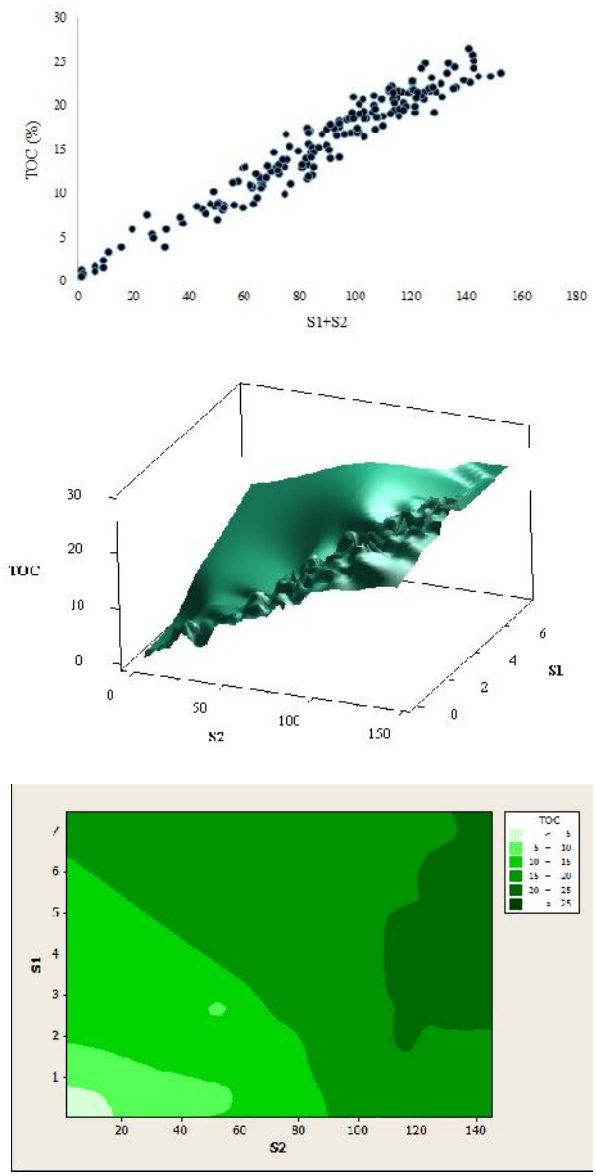

Figure 3. plots of TOC versus S_1 and S_2 for the organic matter-rich facies

In these organic matter–rich units, S_2 and HI with a high proportion show good oil-prone potential.

The regression lines in the S_2 vs. TOC graph pass through the origin because if we have very small quantities of organic matter–yield hydrocarbons during pyrolysis of S_2 vs. TOC the regression lines represent positive intercepts on the x-axis.

This represents that a threshold amount of organic matter must be present before enough hydrocarbons are detected during pyrolysis (Langford and Blanc-Valleron, 1990). For the studied samples, there is no clear intercept of the

S_2–TOC plot, which implies that there is no mineral matrix effect and not much enough inert organic matter of the samples (Fig. 4).

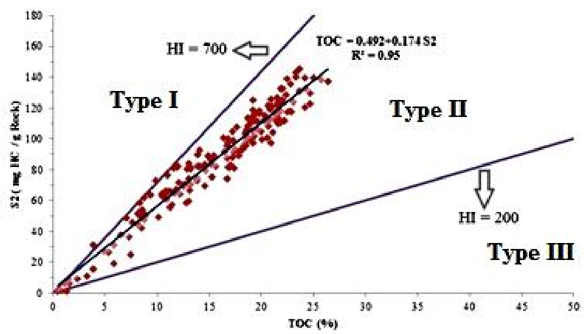

Figure 4. S_2 vs. TOC diagram for the organic matter-rich facies

In the set of samples with the same type of organic matter, HI should have a constant value (Langford and Blanc-Valleron, 1990). However, the observation implies that HI apparently increases with increased TOC (Katz, 1983).

In this study, for majority of samples (TOC > 8%), the regression curve indicates that HI range of the samples is mostly constant and there is no considerable change with increasing TOC (Fig. 5), indicating no substantial variation in the type of organic matter during the deposition of the black organic matter facies in the studied locality.

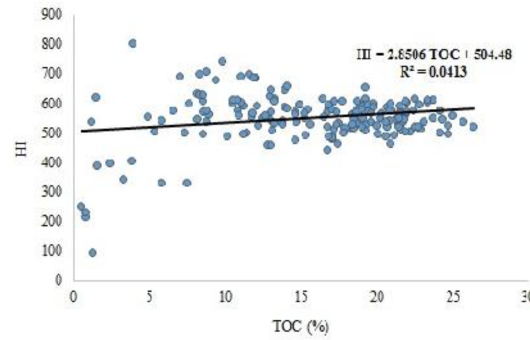

Figure 5. HI vs. TOC graph for investigated units

Conclusion

- A combination of linear regression and P value methods is used to characterize the relationship of Leco TOC and Rock-Eval pyrolysis S_1 and S_2 data from organic matter–rich units. According to the linear regression model, increasing S_2 increases TOC. This model fits approximately 95% to Leco TOC and Rock-Eval pyrolysis data.
- In this article, to determine the parameters that influence TOC, the sensitivity analysis and P value method have been used and a new model based on S_2 was proposed, which has a linear relationship to TOC. In all the models, remaining hydrocarbon potentials (S_2) and TOC have a direct relationship. Using the regression equation, which is based on TOC of the Leco analyzer, petroleum potential of the source rocks can be simply evaluated without Rock-Eval pyrolysis.

References

Bernie B. Bernard, Heather Bernard, and James M. Brooks., 2010. Determination of total carbon, total organic carbon and inorganic carbon in sediments. College Station, Texas 77845.

Espitalié, J., Deroo, G., & Marquis, F. 1985a. La pyrolyse Rock-Eval et ses applications. Patrie I. Rev. Inst. Fr. Pétrole, 40(5): 563-579.

Espitalié, J., Deroo, G., & Marquis, F. 1985b. La pyrolyse Rock-Eval et ses applications. Patrie II. Rev. Inst. Fr. Pétrole, 40(6): 755-784.

Espitalié, J., Deroo, G., & Marquis, F. 1986. La pyrolyse Rock-Eval et ses applications. Patrie III. Rev. Inst. Fr. Pétrole, 41(1): 73-89.

Hunt, John M., 1996. Petroleum Geochemistry and Geology, 2nd edition W.H. Freeman and Company. 743 pp.

Katz, B.J., 1983. Limitations of 'Rock–Eval' pyrolysis for typing organic matter. Org. Geochem. 4: 195–199.

Langford, F.F., Blanc-Valleron, M.M., 1990. Interpreting Rock-Eval pyrolysis data using of pyrolizable hydrocarbons vs. total organic carbon. The American Association of Petroleum Geologists. AAPG Bull. 74 (6): 799–804.

Yalçın Erik, N., Özçelik O., Altunsoy, M., 2006. Interpreting Rock–Eval pyrolysis data using graphs of S2 vs. TOC: Middle Triassic–Lower Jurassic units, eastern part of SE Turkey. Journal of Petroleum Science and Engineering. 53: 34–46.