Integration
geoscience
engineering
partnership
geoConvention 2013

Quantifying Total Organic Carbon (TOC) from Well Logs Using Support Vector Regression

Yexin Liu*, SoftMirrors Ltd., Calgary, Alberta, Canada
yexinliu@softmirrors.com
and
Zhuoheng Chen, Geological Survey of Canada, Calgary, Canada
Kezhen Hu, Geological Survey of Canada, Calgary, Canada
Chris Liu, Cornell University, Ithaca, NY, USA

Summary

There are four basic quantitative methods for determination of total organic carbon (TOC) from well logs, which include: 1) directly from regression of core TOC versus core bulk density; 2) ΔlogR technique presented by Passey (1990); 3) based on the petrophysical response model; 4) an artificial neural network. Currently ΔlogR technique is the most popular approach and widely used in shale gas reservoir evaluation.

In this study we propose an improved ΔlogR equation to calculate apparent logs (density, sonic, gamma ray and neutron) as a function of resistivity log for identifying potential shale plays reservoir intervals, then use support vector regression (SVR) to quantify the relationship between the TOC and apparent well logs and other well attributes. One of the advantages of using SVR method is that it will automatically remove redundant samples and keep only relevant support vectors (samples) to build a non-linear relationship in a hyper space that predicts TOC. The SVR method can also overcome the over-learning problem in artificial neural network (to handle as many as core samples), and integrate ΔlogR attributes with the apparent logs (gamma ray, density, sonic and neutron) as the SVR input attributes. The method has the flexibility of introducing additional variables if data indicate good relationships between TOC and other well log responses.

The method has been tested using real dataset from the Devonian Duvernay Formation of Western Canada Sedimentary Basin (WCSB). The results show that the predicted TOCs match well with the observations from core samples. In addition to the TOC prediction, the proposed approach is also capable of generating synthetic logs or other core properties curves from available well logs.

Introduction

Total organic carbon (TOC) is one of major parameters to assess the potential of shale gas resources. Several methods have been published to quantify total organic carbon (TOC) from well logs. There are four common methods: 1) directly by regression of core TOC with core bulk density; 2) ΔlogR technique (Passey, 1990); 3) based on the petrophysical response model; 4) using an artificial neural network.

The bulk density regression method is a simple method and has produced some satisfactory results when the formation is fairly homogeneous, but in heterogeneous formation such attempts will result in poor correlations. The petrophysical model approach requires accurate porosity values to determine the

amount of TOC/kerogen. In addition, clay quantification can prove problematic in the shale formation. Unfortunately the neutron and density logs can't separate porosity and kerogen content, and gamma ray readings are often affected by variable kerogen content of the shale gas rocks. Usually additional logs such as NMR and SpectroLith logs are required to determine the effective porosity and shale content for shale gas reservoir (Boyer et al. 2006). The artificial neural network can use one or multi-log curves to estimate other log properties (Arbogast and Utley, 2003), but the technique needs extensive trial and error to select partial samples to overcome the over-learning problems. In a highly heterogeneous case, the method is difficult to predict the TOC because some meaningful core samples might not be available for the neural network training.

Currently ΔlogR technique (Passey et al, 1990) is the most common method and widely used for TOC determination in shale gas plays. Basically the three porosity well logs, density, neutron and sonic, can be used to calculate the ΔlogR based on the separation between the deep resistivity curve and the porosity logs, which can be converted into TOC through the level of organic maturity parameter (LOM). Usually LOM needs laboratory experiment to measure. In practice core TOC and calculated ΔlogR assist in calibration to estimate the LOM. In ΔlogR technique porosity effects are minimized. However, sometimes the Passey well logging baselines determination is very challenging. On the other hand, ΔlogR technique can be used for the wells with one porosity log or multi porosity logs. If multi porosity log curves are used in ΔlogR technique, estimated TOC values might be different.

In order to minimize the effect of baseline determination and incorporate multi porosity inputs, we propose the use of improved ΔlogR equations to calculate the apparent logs, and apply support vector regression (SVR) to establish the relationship between well responses and TOC content in shale.

Passey's attributes to determine the TOC

Geological and petrophysical studies suggest that the density, neutron and sonic logs have direct correlation with TOC. In addition to the density, neutron and sonic logs, the Passey's ΔlogR can also be used as the SVR input attributes to predict TOC. The followings are the ΔlogR attributes based on the different well logs (Passey et al, 1990):

$$\Delta logR_{dt} = log(RT/RT_{baseline}) + 0.02(DT - DT_{baseline}) \quad (1)$$

$$\Delta logR_{den} = log(RT/RT_{baseline}) - 2.5(DEN - DEN_{baseline}) \quad (2)$$

$$\Delta logR_{cnl} = log(RT/RT_{baseline}) + 4.0(CNL - CNL_{baseline}) \quad (3)$$

In fact the ΔlogR can be converted into TOC using the level of organic maturity (LOM):

$$TOC = \Delta logR * 10^{(2.297 - 0.1688 * LOM)} \quad (4)$$

Apparent logs to identify the shale gas plays

Based on the Passey equations (1, 2, 3), sometimes it is difficult to determine the petrophysical baseline for estimating ΔlogR. Instead we presented the RMS (root mean square) of standard shale intervals to generate the different apparent logs, which can identify the shale gas plays anomalies.

$$DT_a = (log(RT) * DT_{rms} - \log(RT_{\mathrm{rms}}) * DT_{rms})/\log(RT_{\mathrm{rms}}) + (DT - DT_{rms}) \quad (5)$$

$$DEN_a = (log(RT) * DEN_{rms} - \log(RT_{\mathrm{rms}}) * DEN_{rms})/\log(RT_{\mathrm{rms}}) - (DEN - DEN_{rms}) \quad (6)$$

$$CNL_a = (log(RT) * CNL_{rms} - \log(RT_{\mathrm{rms}}) * CNL_{rms})/\log(RT_{\mathrm{rms}}) + (CNL - CNL_{rms}) \quad (7)$$

$$GR_a = (log(RT) * GR_{rms} - \log(RT_{\mathrm{rms}}) * GR_{rms})/\log(RT_{\mathrm{rms}}) + (GR - GR_{rms}) \quad (8)$$

Where DT, DEN, CNL and GR are sonic, density, neutron and gamma ray well log data

$DT_{rms}, DEN_{rms}, CNL_{rms}, GR_{rms}$ can be calculated from the shale intervals of well logs.

DT_a, DEN_a, CNL_a, GR_a are sonic, density, neutron and gamma ray apparent logs.

Like the Passey equations, it is possible to convert the apparent logs to TOC using the LOM:

$$TOC = A * LOG_a * 10^{(B-C*LOM)} \quad (9)$$

Where LOG_a are the apparent logs from equation (5, 6, 7, and 8). A, B, C can be estimated from the core dataset.

Support Vector Regression Solution

The equations 1, 2, 3 and equations 5, 6, 7 and 8 can be used to estimate the TOC using the equations 4 or 9. Unfortunately the TOC will be different based on these different equations. It is difficult for us to determine which equation is best for the determination. So we presented the Support Vector Regression (SVR) method to generate the non-linear relationship between the core data and different equation samples. In fact SVR has been popular for regression problems because the SVR method has the capability to handle the sparse measurements, which is the case for TOC samples obtained at the core analysis, and it also has excellent performance in dealing with these sparse inputs.

The SVR requires a set of well logging samples $\{x_{i,}, i = 1, \dots, N\}$, such as apparent density, apparent sonic and apparent neutron, along with the corresponding sparse core analysis properties $\{y_i, i = 1, \dots, N\}$ like TOC. The SVR makes the training and prediction based on the following linear regression function (Schölkopf et al, 1999, Liu et al 2003, Smola et al, 2004):

$$f = w^T K(x_i, x) + b \quad (10)$$

on a feature space **F**. Here w is a vector in **F** and $K(x_i, x)$ maps the input x to a vector in **F**. We note $\{K(x_j, x_n)\}$ denotes the kernel functions, such as Gaussian function. This shall be denoted as $\langle x_i, x \rangle$ henceforth.

Utilizing $\varepsilon - SV$ regression, we want a function $f(x) = \langle w, x \rangle + b$ that has at most ε deviation from the actual sparse core analysis properties$\{y_i, i = 1, \dots, N\}$.

Before we present the formula, we note that the standard $\varepsilon - SV$ regression is a hard-margin convex-optimization problem. That is, errors those are greater than ε have infinite cost, and errors less than ε have zero cost. This requires our function f successfully approximating all pairs (x_i, y_i) with ε precision, which may not be feasible. We can relax this restraint with slack variables ξ_i, ξ_i^* that allows us to cope with otherwise infeasible constraint. In light of this, we have the following formulation (Smola and Schölkopf, 2004)

$$minimize \ \frac{1}{2}\|w\|^2 + C \sum_{i=1}^{N} (\xi_i + \xi_i^*) \quad (11)$$

$$subject\ to \begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \quad (12)$$

We note that the w and b together determines the regression function f completely.

The constant $C > 0$ is a parameter that determines how costly training examples that do not satisfy the constraint cost with regard to our optimization equation. A larger C simulates hard-margin regression while a smaller C means that the objection function, f, is allowed to deviate larger than ε with a low cost.

In order to solve (11) and (12), the optimization problem for $\varepsilon - SV$ regression can be expressed in matrix notation as:

$$min\ \frac{1}{2} w^T H w + c^T w \quad (13)$$

Where

$$H = \begin{bmatrix} K(x_i, x)^T & -K(x_i, x)^T \\ -K(x_i, x)^T & K(x_i, x)^T \end{bmatrix}, c = \begin{bmatrix} \varepsilon + Y \\ \varepsilon - Y \end{bmatrix}, w = \begin{bmatrix} \alpha \\ \alpha^* \end{bmatrix}.$$

X and Y are the row matrices with values $\{x_i, i = 1, \dots, N\}$ and $\{y_i, i = 1, \dots, N\}$.

We have now formulated the optimization problem in SVR in terms of a quadratic-optimization matrix form (13). We gain additional insight in the form, noting that our regression function, f, can be described by a linear combination of training examples x_i. And furthermore, we are able to evaluate a regression function f in terms of the training example's inner products, meaning our linear regression, when mapped using $K(x_i, x)$, can take a distinctive non-linear nature.

Validation and TOC prediction

For any statistical method, we need to select the attributes (such as apparent logs and modified Passey's attributes) and their samples to build a reasonable relationship. But the number of attributes and samples for generating a reasonable relationship depends on the complexity of the problem. If the problem is linearly separable, the rule is: the less the better. Generally it is possible to achieve good performance with few attributes and samples, but for non-linear problems, the rule is: the more the better. However, although a larger sample size will give more accurate results, real datasets, such as well logs and core measurements, are not perfect and are never exactly representative of the process.

Validation is crucial to determine the sample selections base on the SVR well training satisfactory. Usually we can build the crossplots to validate the sample distribution. Figure 1 is a crossplot between modified Passey sonic and apparent gamma ray log with sample TOC contours superimposed. The contour lines suggest that the TOC has complicated relationship with modified Passey sonic and apparent gamma log. Several TOC datasets from wells penetrated the Duvernay Formation in Western Canada Sedimentary Basin (WCSB) have been tested and satisfactory TOC prediction results are achieved. As an example, figure 2 compares predicted TOC and measured TOC and they show a good correspondence in the test well. The tracks 1 to 4 display the well logs and track 5 is the TOC prediction curve using SVR and the green dots and blue dots are the core TOC samples. Track 6 is the apparent logs and track 7 is the modified Passey attributes and track 8 shows other log attributes.

Conclusions

We have first put forward methods to generate apparent logs (gamma ray, density, sonic and neutron) from resistivity to identify the shale gas plays anomalies, then used the trained support vector regression (SVR) model to predict TOC from apparent well logs and modified Passey attributes. In the well training phase, the proposed approach can automatically remove the redundant samples and keep few support vectors (samples) to build a hyper-regression relationship. In addition, the proposed method utilizes apparent well logs to minimize the effect of baselines determination. Several real datasets have been tested and the results demonstrated that the proposed approach works well. The proposed methods can be extended to generate the synthetic logs and other core properties from available well logs if necessary.

Reference

1) Arbogast, J. S., Utley, L., 2003, Extracting Hidden Details from Mixed-vintage Log Data Using Automatic, High-speed Inversion Modeling of Resistivity and SP Logs: Supplementing Incomplete Log Suites and Editing Poor Quality Log Data Using Neural Networks, AAPG Annual Convention, May 2003
2) Beaton, A. P., Pawlowicz, J. G., Anderson, S. D. A., Berhane, H., and Rokosh, C. D., 2010, ERCB/AGS Open File Report 2010-04
3) Boyer, C., Kieschnick, J., Rivera, R. S., Lewis, R. E., Waters, G., 2006, Producing Gas from Its Source, Schlumberger Oilfield Review, Autumn, 36-49
4) Bowman, T., 2010, Direct Method for Determining Organic Shale Potential from Porosity and Resistivity Logs to Identify Possible Resource Plays, AAPG Annual Convention, April 2010
5) Liu, Y. X., Sacchi, M. D., 2003, Propagation of Borehole Derived Properties via a Support Vector Machine (SVM), CSEG Recorder, 28, 54 - 58
6) Meyer, B.L., and M.H. Nederlof, 1984, Identification of Source Rocks on Wireline Logs by Density/Resistivity and Sonic Transit/Resistivity Crossplots: AAPG Bulletin, 68, 121-129.
7) Passey, Q. R., Creaney, S., Kulla, J. B., Moretti, F. J., Stroud, J. D., 1990, Practical Model for Organic Richness from Porosity and Resistivity Logs, AAPG Bulletin, 74, 1777-1794
8) Schölkopf, B., Burges, C.J.C., and Smola. A. J., 1999, Advances in Kernel Methods: Support Vector Learning, the MIT Press
9) Smola A. J., and Schölkopf, B.,2004, A Tutorial on Support Vector Regression , Statistics and Computing 14, 199-222

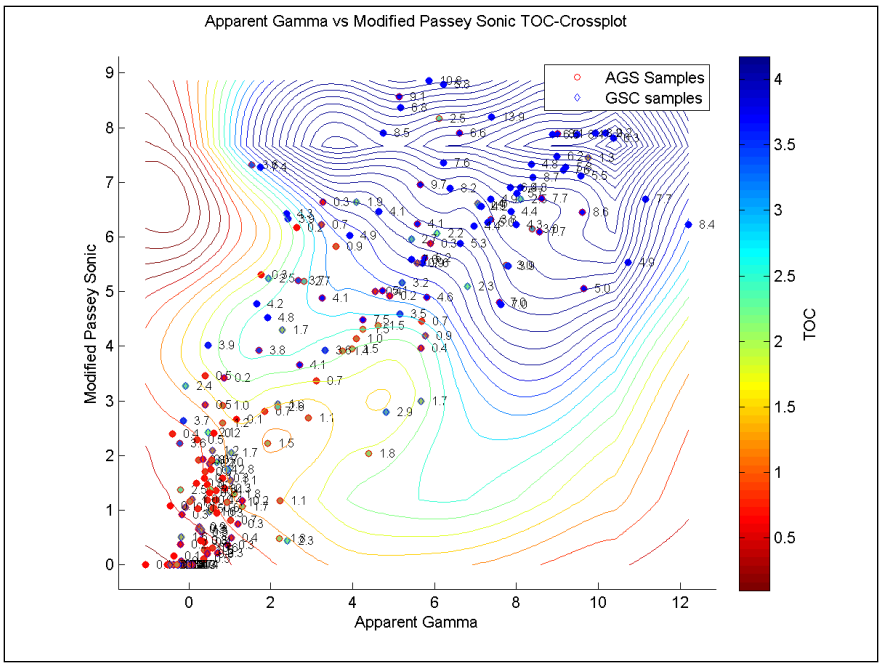

Figure 1: Modified Passey sonic and apparent gamma ray log crossplot demonstrated the complex distribution between TOC and Passey sonic/apparent Gamma Ray log

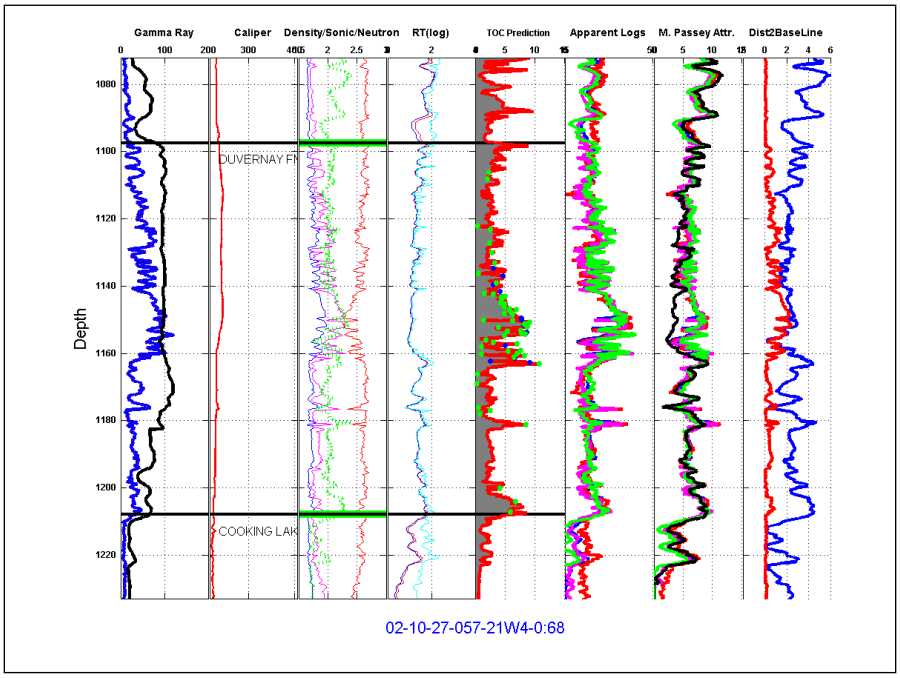

Figure 2: TOC prediction (track 5) from well log attributes and core TOC samples