TOC estimation in the Barnett Shale from triple combo logs using support vector machine

Tao Zhao, Sumit Verma, and Deepak Devegowda, University of Oklahoma*
Vikram Jayaram, Global Geophysical Services

Summary

In this study, we show an application of estimating total organic carbon (TOC) in a Barnett Shale play from the widely available triple combo logs using support vector machine (SVM). Being a nonlinear supervised learning technique, SVM provides superior estimation than the traditional multi-linear regression. Using triple combo logs to automatically estimate TOC content from a limited number of pre-existing TOC measurements, the proposed method delivers convenient and relatively accurate TOC estimation in a resource play where core measurements and mineralogy logs are limited.

Introduction

Total organic carbon (TOC) estimation is a key factor to identify sweet spot in unconventional shale plays. Practically, direct measurement of TOC can only be acquired in a laboratory from core data, or from mineralogy logs (e.g. elemental capture spectroscopy logs), both of which needs special efforts and equipment that are not widely accessible to a general interpreter. Let alone the high processing cost of core datasets. However, there are several ways to estimate TOC from certain well logs. Passey's (1990) method of TOC calculation with well log data is one of the most popular methods. But, it requires interpreters' input on base line selection, which introduces human error. In this study we estimate the relation between basic triple combo well logs and elemental capture spectroscopy sonde (ECS) measured TOC in the Barnett Shale using support vector machine (SVM), which is a nonlinear supervised machine learning technique. Estimation of TOC from core measurements using support vector regression has shown promising result (Liu et al., 2013), whereas in our study we use ECS measured TOC because it is easier to access than core measurements while still providing satisfactory accuracy. Researches on calibrating ECS logs with core cutting also indicate ECS logs can produce readings with consistently high confidence while core measurements may be erroneous due to the change in ambient conditions and human introduced error during the measuring process (Herron et al., 2014). Such estimation is further compared with traditional multi-linear regression to illustrate effectiveness.

Geologic Setting

The target formation in this study is the Lower Barnett Shale (Figure 1), which was deposited in the Mississippian period and dominated by silica-rich mudstones. The shale formation is bonded by Forestburg Limestone and Viola Limestone, which are considered as fracture barriers when doing hydraulic fracturing in the shale formations. In the study area, comparing to the Upper Barnett Shale, the Lower Barnett Shale is more organic rich with TOC varying from 0% to 10% by weight percent according to the ECS logs from well A shown in Figure 2. This measurement agrees with TOC core measurements from Singh (2008).

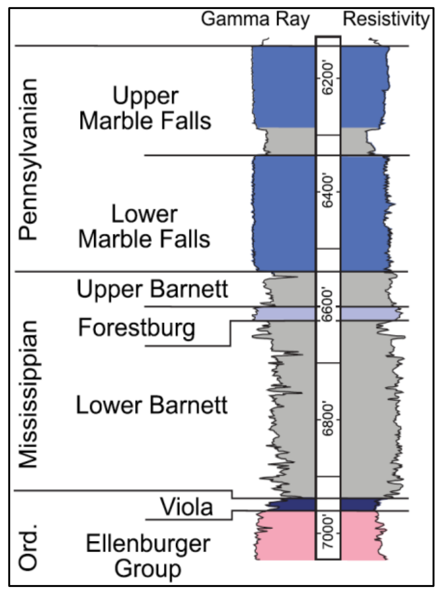

Figure 1. General stratigraphy of the Ordovician to Pennsylvanian section in the Fort Worth Basin through a well near the study area (After Loucks and Ruppel, 2007).

Methodology

Triple combo logs may be the most commonly acquired well logs in a resource play which include measurements of gamma ray, density, porosity, resistivity, caliper, and temperature. Generally, one could expect proportional relation between TOC and gamma ray, porosity and resistivity and inverse proportional relation to density. Based on such relations, crossplots can then be made among these properties to get a qualitative understanding of TOC (Meyer and Nederlof, 1984). If some TOC measurements are available (in our case, ECS logs), one can either use multi-linear regression to estimate TOC from other logs (Verma and Marfurt, 2013), or more sophisticated nonlinear pattern recognition techniques such as SVM to estimate such relation. The latter is assumed to be more effective given the fact that the intrinsic relation between TOC and other well logs is nonlinear (Liu et al., 2013). As a rule of thumb, the less number of input dimensions, the better generalization of the classifier. By testing through many alternative input log combinations, we choose a group of three logs with the highest correlation with TOC log in our study area, which are bulk density, resistivity, and neutron porosity, as the input logs.

We use proximal support vector machine (PSVM) as the implementation of SVM, which provides comparable performance at lower computational cost (Fung and Mangasarian, 2001, 2005; Mangasarian and Wild, 2006). Unlike another popular competing method, artificial neural network (ANN), which converges at local minima therefore provides different prediction results for each run, the SVM solves a convex optimization problem which means exact prediction once the model parameters are fixed. Data from 3-dimensional input space (3 dimensions are density, resistivity, and porosity) are mapped into a higher dimensional feature space where data from different classes can be linearly separated. Such prediction is then compared with the traditional multi-linear regression using the same input logs.

Application and Result

Before stepping into the SVM prediction, a pre-editing is performed on the TOC log to make it compatible with the SVM classifier. Because currently our SVM can only be used for classification but not regression, we have to discretize the TOC long into 20 classes, for 1 being the lowest TOC and 20 being the highest. We find discretizing the TOC log into 20 discrete numbers to be enough to recover the most information in the TOC log.

Figure 3 shows the three input logs, TOC log, discretized TOC log, and the SVM estimated TOC log in the Lower Barnett Shale formation. SVM estimated TOC log is calculated by using 50% of the data as training, then applying the relation to the whole Lower Barnett Shale segment. The correlation between discretized TOC and SVM estimated TOC is 0.96. Once we have derived this relation on the training well A, we apply such relation on a testing well B, which is approximately 33 miles away from well A. Such distance provides a suitable scenario to test the generalization capability of the nonlinear relation derived using SVM. In the same Lower Barnett Shale formation on well B, we observe a correlation of 0.86 between SVM estimated TOC and the discretized TOC (Figure 4), which indicates a reliable estimation considering the distance between these two wells.

To further verify the performance of the SVM prediction, we perform multi-linear regression using the same input logs, to estimate TOC log in the Lower Barnett Shale formation. We still build the regression model on well A, then test it on well B. Table 1 shows the coefficients derived from the multi-linear regression. Figure 5 shows the comparison between SVM prediction and multi-linear regression on both training well A (Figure 5a) and testing well B (Figure 5b). On the training well A, we identify an excellent match between the discretized TOC and SVM estimated TOC when using 50% of the data to build the SVM model; while the multi-linear regression can only give 0.8 correlation with the TOC log. For the testing well B, we have 0.86 correlation between discretized TOC and SVM estimated TOC, and 0.85 correlation between TOC from multi-linear regression and TOC from ECS. To our surprise, multi-linear regression provides comparable correlation on the testing well, which may be credited to the relatively large distance between the two wells. Such large distance reduces the nonlinearity of the relation between TOC and other log measurements, and multi-linear regression can provide a more generalized relation. This is intuitive because for one regional geologic background we would expect local variations, and the regional background gives us a more generalized linear relation, whereas local variations result in more specific nonlinear relation. However, a shift can be easily identified on the multi-linear regression result on well B (in this case, underestimated the TOC), which makes it less favorable especially when characterizing lateral variations within a certain formation. The SVM estimated TOC is free from such problem, therefore is preferred to apply over the resource play area.

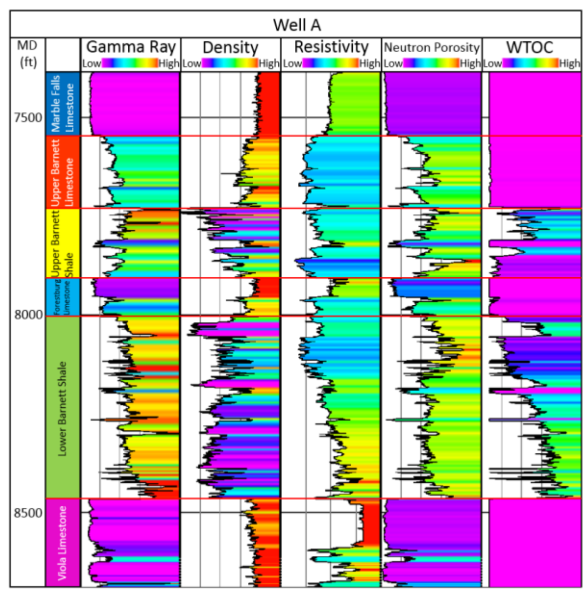

Figure 2. Log displays on well A over the interval from Marble Falls Limestone to Viola Limestone. Bulk density, deep induction resistivity, thermal neutron porosity logs are used as input logs for both SVM and multi-linear regression. WTOC log is measured using ECS sonde. Gamma ray log is displayed as a lithology indicator, which is not used as input for further estimations. One can identify the high TOC content in the Lowe Barnett Shale formation.

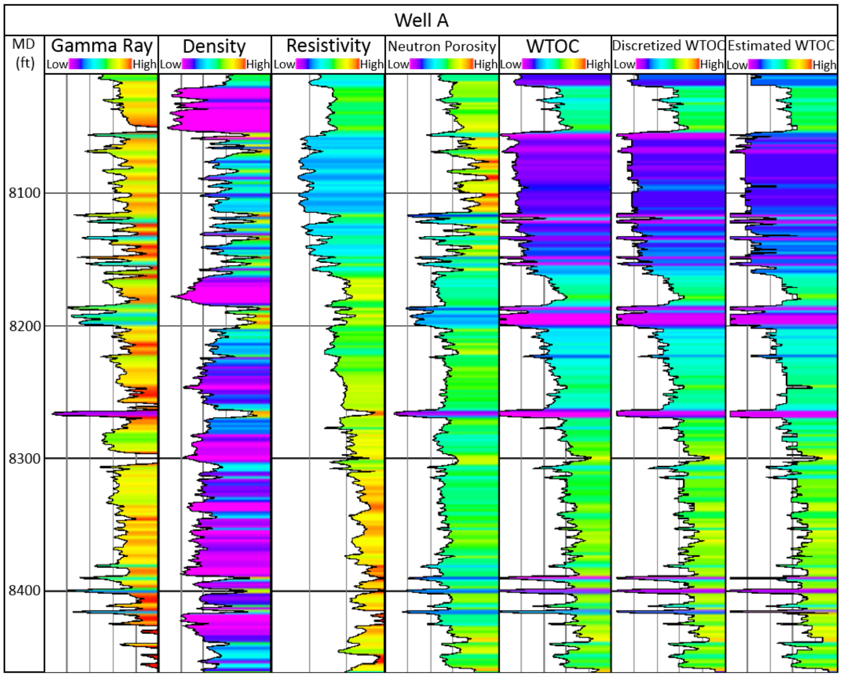

Figure 3. Log displays on training well A over the interval of Lower Barnett Shale. Bulk density, deep induction resistivity, thermal neutron porosity logs are used as input logs for SVM. Gamma ray log is displayed as a lithology indicator, which is not used as input for SVM. Discretized WTOC is a WTOC log discretized into 20 discrete numbers, 1 being the lowest WTOC and 20 being the highest. Estimated WTOC is from SVM using 50% of the data as training. One can identify the high level of correlation between discretized WTOC and estimated WTOC.

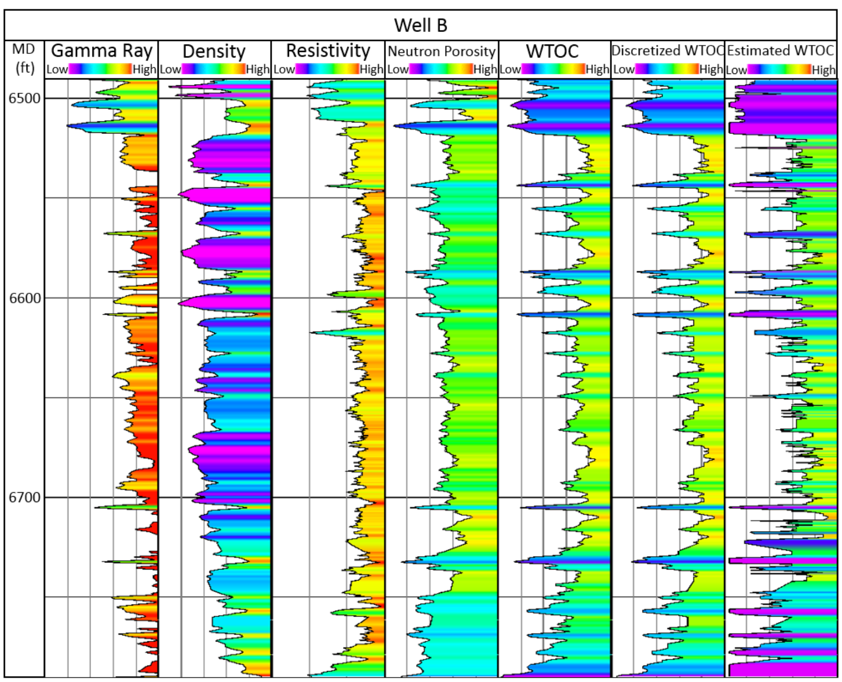

Figure 4. Log displays on testing well B over the interval of Lower Barnett Shale. Estimated WTOC is from the SVM model built on the training well A. One can identify good correlation between discretized WTOC and estimated WTOC.

Variables	Coefficients
Resistivity	8.58828E-06
Density	-0.22159661
Porosity	-0.046354456
Intercept	0.601503779

Table 1. Coefficients for the multi-linear regression derived on training well A.

Conclusion and Future Work

In this study we estimated TOC from basic triple combo logs using SVM, which provides a convenient method to extend the TOC coverage to a greater area. Comparison between SVM and multi-linear regression validates the performance of the proposed method. An extension of this work is to generate a TOC volume by building a relation between TOC logs and seismic attributes (elastic properties and other attributes that are sensitive to TOC) at well locations using ANN or SVM, however the choice of attributes is crucial and requires great insight.

Acknowledgement

We thank Devon Energy for providing the data and permitting for publication. Financial support for this effort and for the development of our own PSVM algorithms was provided by the industry sponsors of the Attribute-Assisted Seismic Processing and Interpretation (AASPI) consortium in the University of Oklahoma. We thank colleagues for their valuable input and suggestions. All the well log displays were made using licenses to Petrel, provided to the University of Oklahoma for research and education courtesy of Schlumberger.

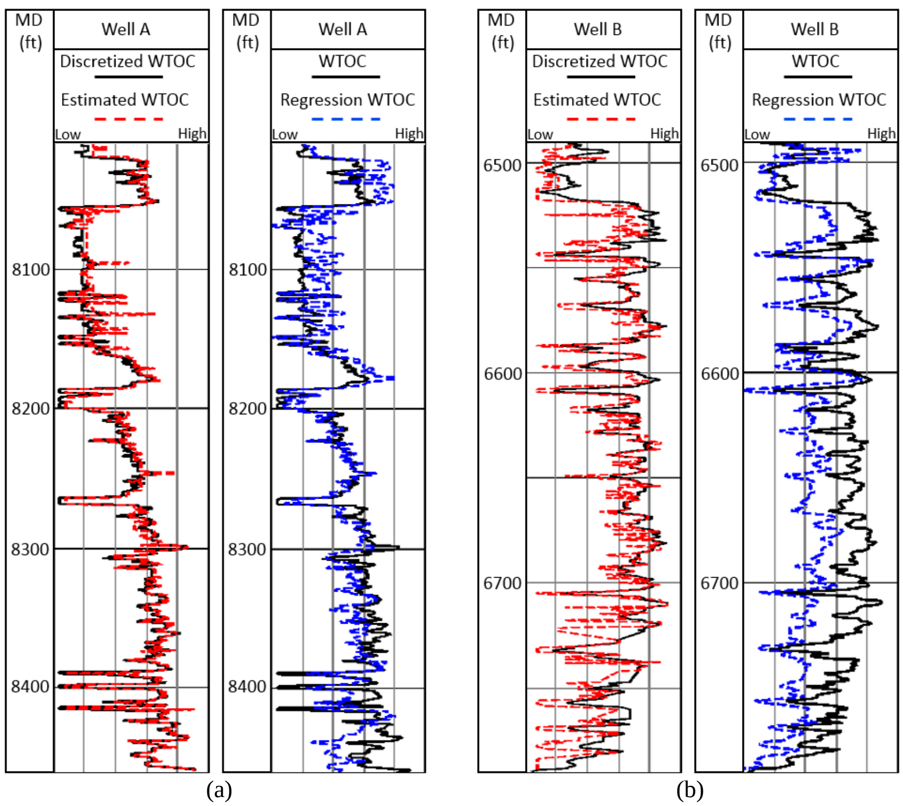

Figure 5. Comparison between SVM estimation and multi-linear regression on **(a)** training well A and **(b)** testing well B. One can identify the better correlation of SVM estimation on both wells, as well as the drift of the WTOC from multi-linear regression on testing well B.