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

The UNIVERSITY of OKLAHOMA
Mewbourne College of Earth and Energy
ConocoPhillips School of Geology and Geophysics

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 using 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 a Barnett Shale play 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 cuttings 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 its effectiveness.

Location of study area

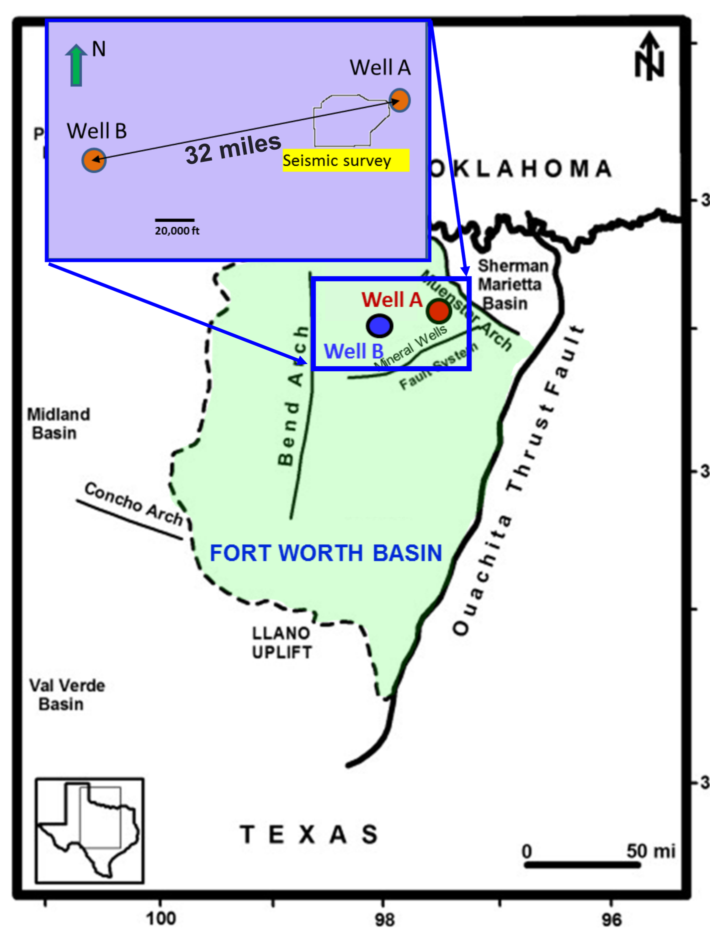


Figure 1. Extent of the Mississippian Barnett Shale, Fort Worth Basin, Texas (Aydemir, 2011). The red and blue dots indicate the approximate locations of cored well A and B used in this study. The zoomed in blue rectangle shows the location of cored wells A and B with respect to a seismic survey.

Stratigraphy of study area

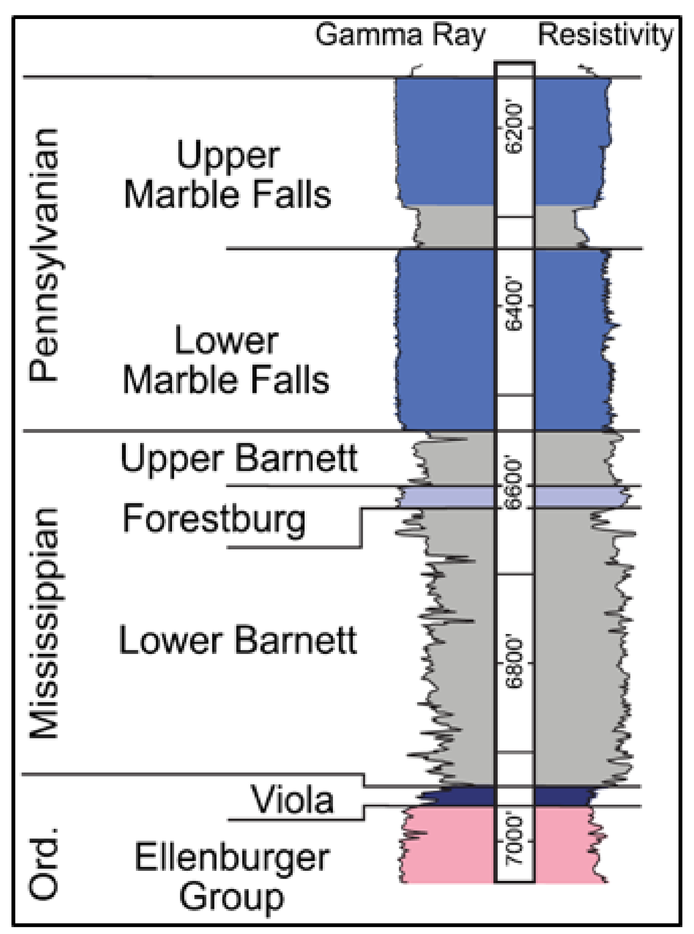


Figure 2. General stratigraphy of the Ordovician to Pennsylvanian section in the Fort Worth Basin through a well near the study area (After Loucks and Ruppel, 2007).

Workflow

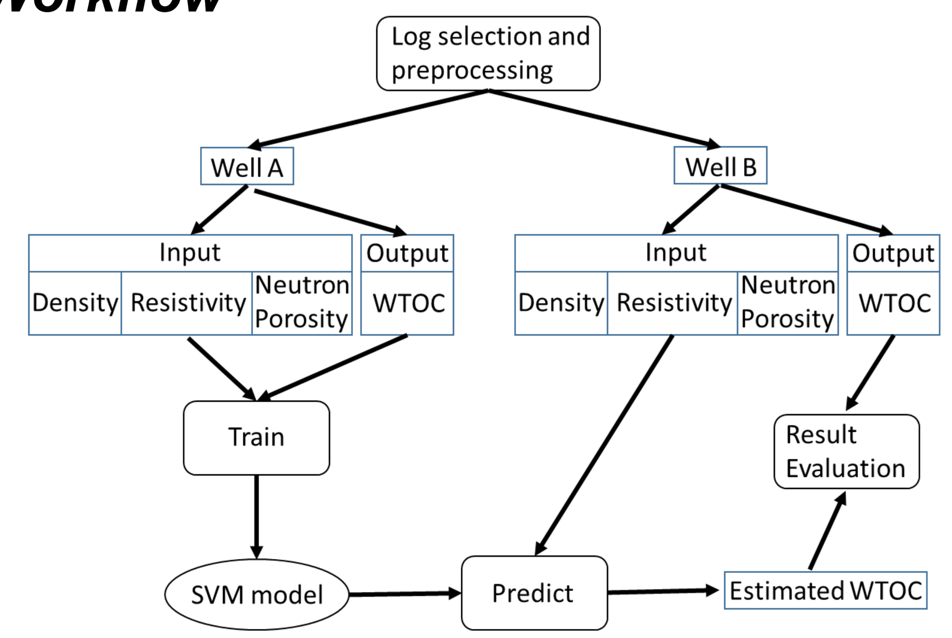


Figure 3. Workflow used in this study for TOC estimation. An SVM model is constructed on a training well A, and applied on a testing well B to evaluate its performance.

Theory of support vector machine

The basic idea of SVM is straightforward. First, we transform the training data vectors into a still higher dimensional "feature" space using nonlinear mapping. Then we find a hyperplane in this feature space that separates the data into two classes with an optimal "margin". The concept of a margin is defined to be the smallest distance between the separation hyperplane (commonly called a decision boundary) and the training vectors (Bishop, 2006) (Figure 4). An optimal margin balances two criteria: maximizing the margin, thereby giving the classifier the best generalization, and minimizing the number of misclassified training vectors if the training data are not linearly separable. The margin can also be described as the distance between the decision boundary and two hyperplanes defined by the data vectors which have the smallest distance to the decision boundary. These two hyperplanes are called the "plus-plane" and the "minus-plane". The vectors which lie exactly on these two hyperplanes that mathematically define or "support" them and are called support vectors. If data are not linearly separable in the original input space, they are projected into the higher dimensional space where linear separation is achieved (Figure 5) using kernel functions.

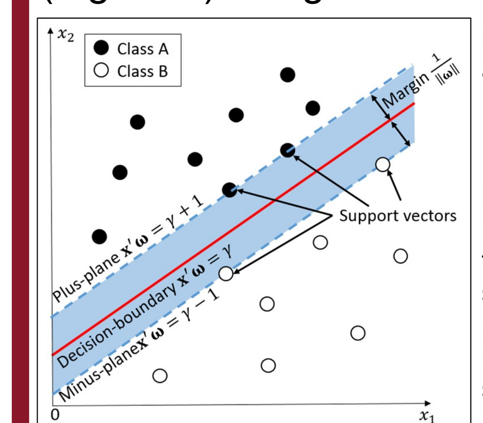


Figure 4. (left) A cartoon illustration of an SVM classifier with two classes.

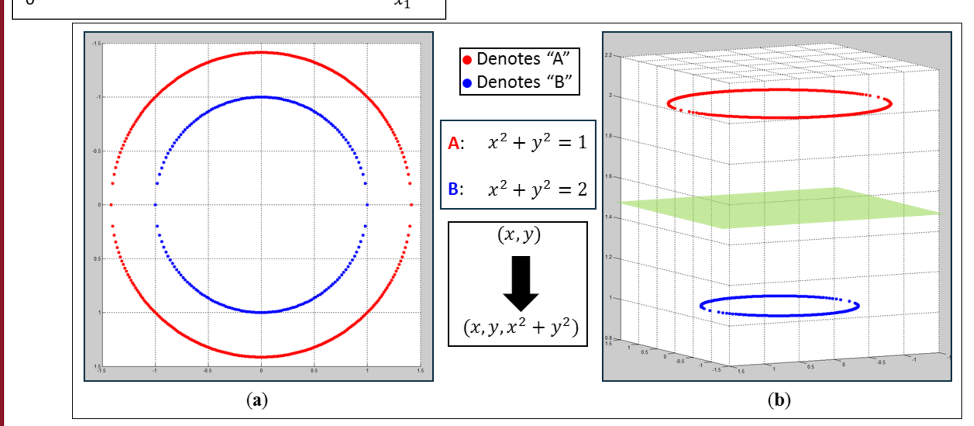


Figure 5. (bottom) (a) Two classes represented by two circles in 2D space. In this space, the two classes are not linearly separable. **(b)** The same two classes as in (a) but in a 3D space. The two classes are linearly separable in this higher dimensional space.

Results

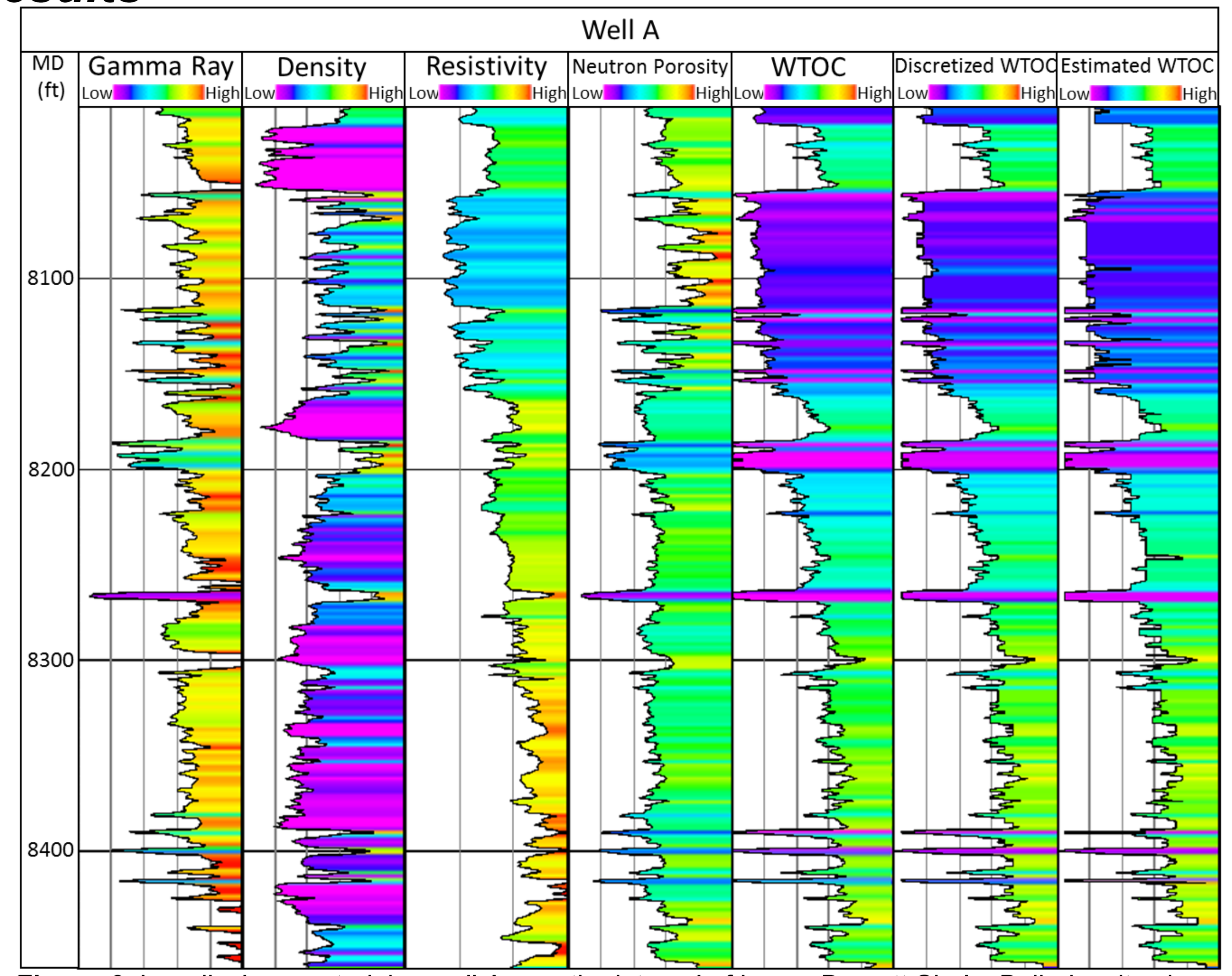


Figure 6. Log displays on training well A over the interval of Lower Barnett Shale. Bulk density, deep induction resistivity, thermal neutron porosity logs are used as input logs for SVM. Gamma ray log is displayed as a lithology indicator, which is not used as input for SVM. Discretized WTOC is a WTOC 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.

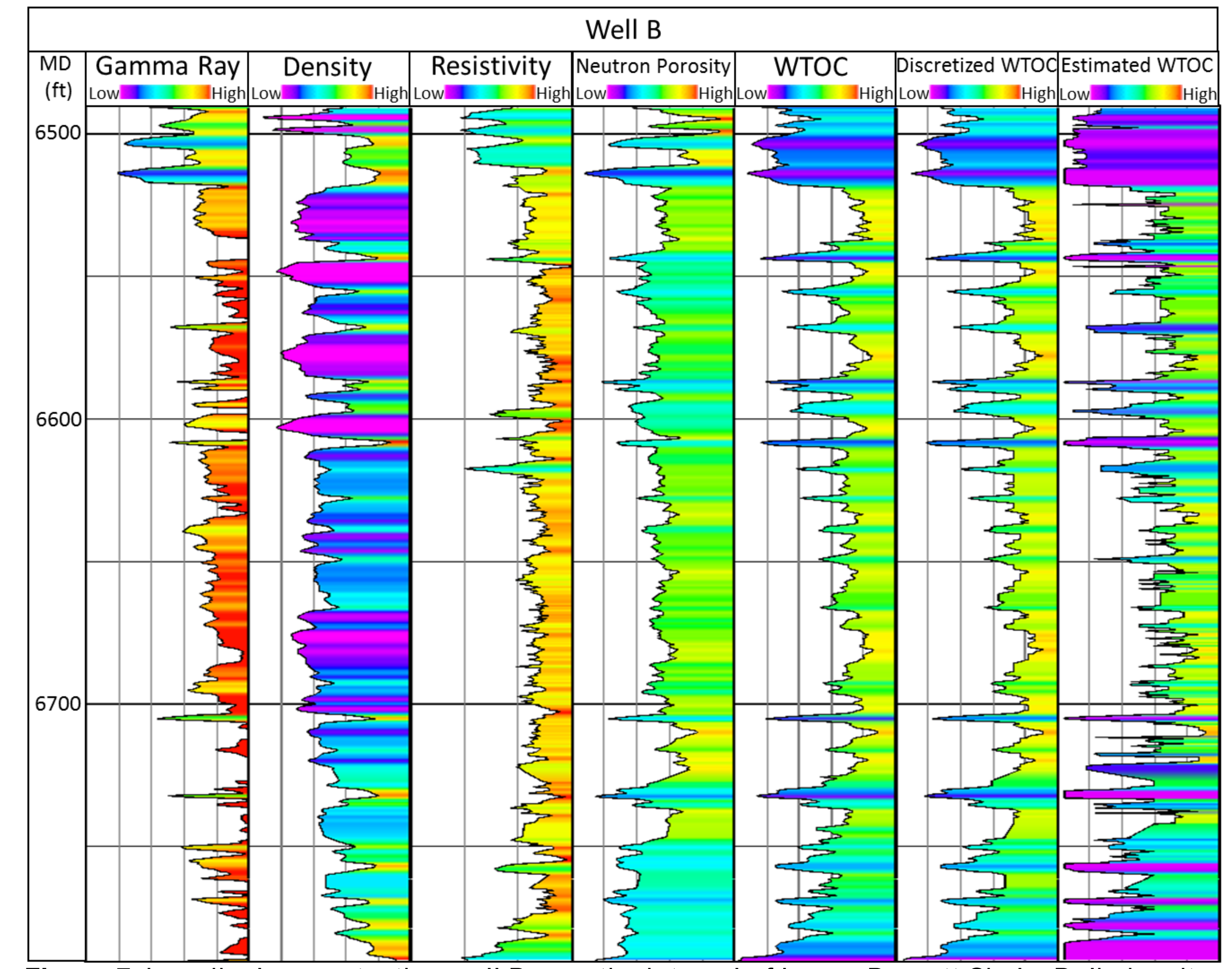
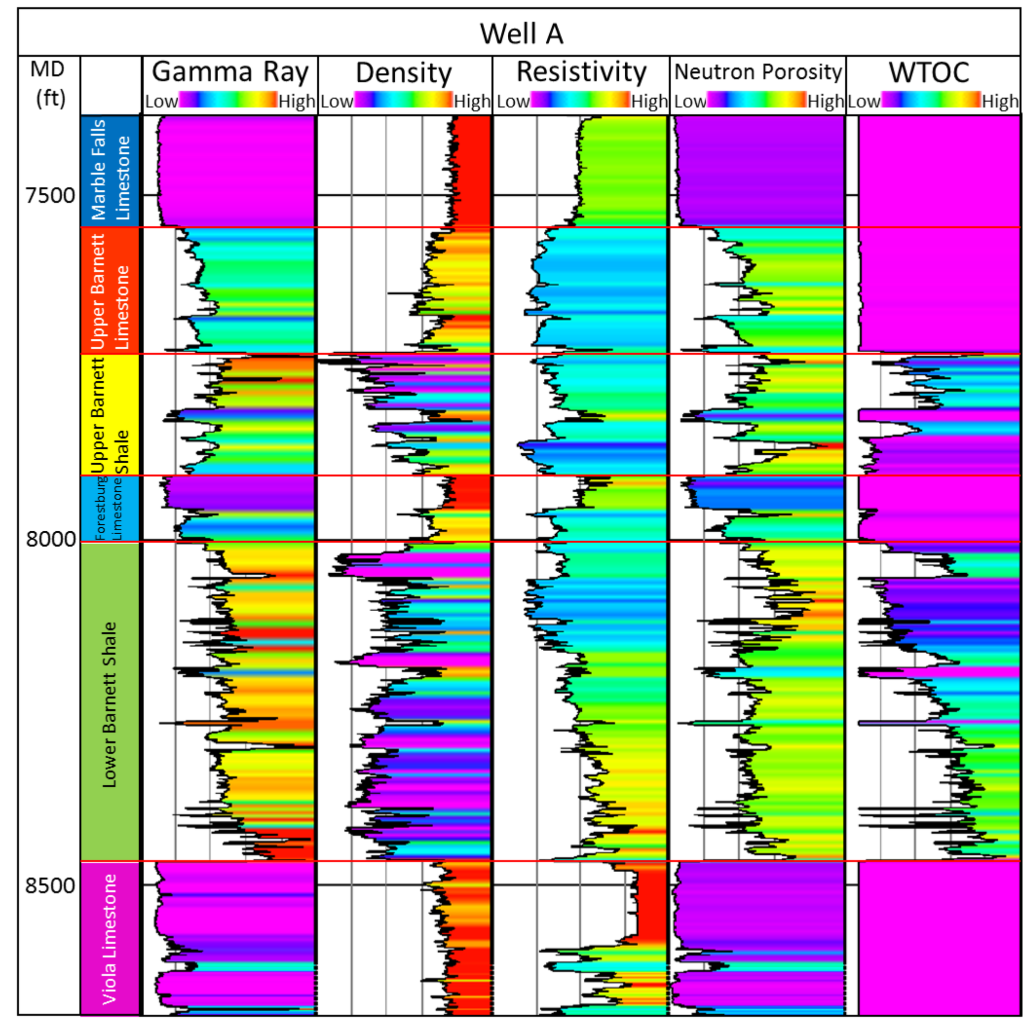


Figure 7. Log displays on testing well B over the interval of Lower Barnett Shale. Bulk density, deep induction resistivity, thermal neutron porosity logs are used as input logs for SVM. Gamma ray log is displayed as a lithology indicator, which is not used as input for SVM. Discretized WTOC is a WTOC log discretized into 20 discrete numbers, 1 being the lowest WTOC and 20 being the highest. Estimated WTOC is from 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.

Figure 8. 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.

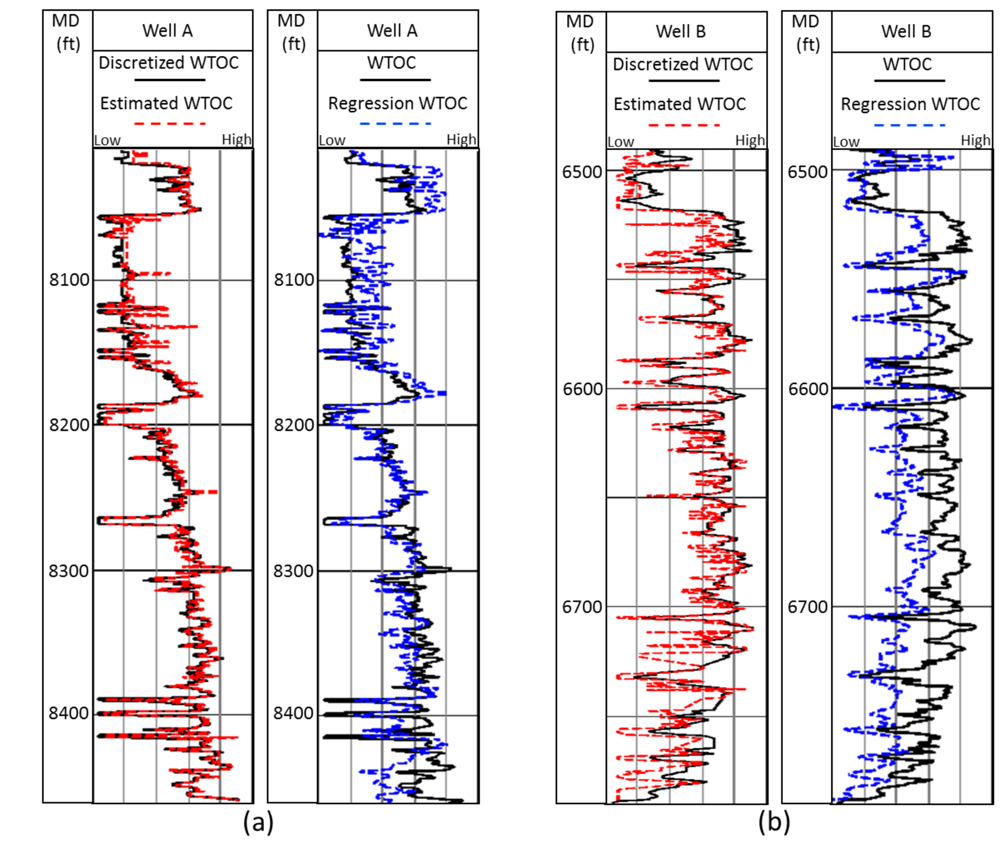


Figure 9. Comparison between SVM estimation and multi-linear regression on **(a)** training well A and **(b)** testing well B. The multi-linear regression coefficients are listed in Table 1. Note the better correlation of SVM estimated WTOC comparing to WTOC from regression on both wells. Also note the drift of the WTOC from multi-linear regression on testing well B.

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 SVM algorithms was provided by the industry sponsors of the Attribute-Assisted Seismic Processing and Interpretation (AASPI) consortium at 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.

References

Aydemir, A., 2011, Comparison of Mississippian Barnett Shale, Northern-Central Texas, USA and Silurian Dadas Formation in Southeast Turkey: Journal of Petroleum Science and Engineering, **80**, 81 - 93.
Bishop, C. M., 2006, Pattern recognition and machine learning: Springer, New York, United States.
Herron, S., M. Herron, I. Pirie, P. Saldungaray, P. Craddock, A. Charsky, M. Polyakov, F. Shray, and T. Li, 2014, Application and quality control of core data for the development and validation of elemental spectroscopy log interpretation: Petrophysics, **55**, 392-414.
Liu, Y., Z. Chen, K. Hu, and C. Liu, 2013, Quantifying total organic carbon (TOC) from well logs using support vector regression: CSPG/CSEG/CWLS GeoConvention 2013, AAPG Search and Discovery Article #90187.
Loucks, G. R, and C. S. Ruppel, 2007, Mississippian Barnett Shale: Lithofacies and depositional setting of a deep-water shale-gas succession in the Fort Worth Basin, Texas: AAPG Bulletin, **91**, 579-601.
Passey, Q. R., S. Creaney, J. B. Kulla, F. J. Moretti, and J. D. Stroud, 1990, A practical model for organic richness from porosity and resistivity logs: AAPG Bulletin, **74**, 1777-1794.

For communication: tao-zhao@ou.edu