Improving prediction of Total Organic Carbon in prospective Australian basins by employing machine learning

Irina Emelyanova*
CSIRO Energy
26 Dick Perry Avenue
Kensington WA 6151
Irina.Emelyanova@csiro.au

Marina Pervukhina
CSIRO Energy
26 Dick Perry Avenue
Kensington WA 6151
Marina.Pervukhina@csiro.au

M. Ben Clennell
CSIRO Energy
26 Dick Perry Avenue
Kensington WA 6151
Ben.Clennell@csiro.au

David N. Dewhurst
CSIRO Energy
26 Dick Perry Avenue
Kensington WA 6151
David.Dewhurst@csiro.au

**presenting author*

SUMMARY

Total organic carbon (TOC) is directly associated with total porosity and gas content and is a critical factor in assessing the potential of unconventional reservoirs. TOC content is only known at the depths where the laboratory measurements on recovered core samples are performed. However, reliable estimation of potential resources can only be based on information about vertical and lateral distribution of organic matter throughout the prospective gas shale reservoir. This information is commonly obtained from conventional wireline logs, such as gamma ray, density, transit time and resistivity. Due to the complexity of unconventional reservoirs, traditional methods based on distinct differences of resistivity, density and sonic velocity of organic matter from those of the inorganic matrix are not always successful. We investigate the best way to predict the TOC using gamma-ray, density, porosity, resistivity and sonic transit time log responses by applying machine learning methods such as Artificial Neural Network (ANN) and Support Vector Machine (SVM). The analysis is done on the data from seven wells drilled through onshore unconventional reservoirs in the McArthur Basin (Northern Territory) and Georgina Basin (Northern Territory and Queensland), Australia. The prediction quality of traditional, multiple liner regression (MLR) and machine learning methods was compared. The most accurate TOC estimates were generated by ANN- and SVM-based nonlinear predictors, followed by the MLR and traditional models. This indicates that geologic complexity affects the relationship between the log response and TOC in the area of interest.

Key words: TOC, unconventional gas reservoir, regression, ANN, SVM

INTRODUCTION

Total organic carbon (TOC) content is crucial in assessing the potential of unconventional reservoirs and is directly associated with total porosity and gas content. Methods routinely used for estimating TOC were developed for organic-rich shales from Northern American Basins, where organic matter of marine Type I and Type II is typically observed. In prospective Australian basins, organic matter ranges from marine Type I to Terrestrial Type III. This fact, along with different depositional history of Australian reservoirs, may lead to quite different logging tool responses. Thus, the relations between the amount of TOC and estimation from logs might be quite different. Traditional methods of TOC estimation are based on distinct differences of density, resistivity, and sonic velocity of organic matter from those of the inorganic matrix. Grain density of organic matter (~ 1.1–1.4 g/cc) is significantly lower than typical density of inorganic matter of organic-rich shales and thus can be used for TOC estimation (Schmoker, 1979). However, as the density log can be affected strongly by heavy minerals and matrix inhomogeneities, the related matrix grain-density variations may obscure the TOC effect. Resistivity is another parameter affected by the presence of organic matter in that major resistivity increases are caused by the decrease of the conductive fluid volume and the increase of gas/oil content. Separation between the resistivity log and sonic travel time was shown to be a good indicator of organic matter that is related linearly to TOC in the oil maturity window (Passey et al., 1990). However, in the case of abundant clay and low brine conductivity, the effect of the organic matter on resistivity may be insignificant. Pyrite also might increase formation conductivity considerably and lead to underestimation of TOC content. Moreover, some overmature organic-rich rocks can be highly conductive (Passey et al., 2010). The third log traditionally used for TOC estimation is sonic (e.g., Zhu, 2010; Dellenbach, 1983). Organic matter is characterized with very low elastic moduli K = 3.5–5 GPa and μ = 1.8–2.5 GPa (Yan and Han, 2013), which correspond to compressional and shear velocities of 2.1 to 2.7 km/s and 1.2 to 1.5 km/s, respectively. However, sonic velocities can be affected strongly by intervening layers of calcite or dolomite and other inhomogeneities that result in strong scatter in TOC-sonic correlations. The neutron porosity log generally is considered to be a poor indicator of TOC, but neutron porosity can be used as a gas indicator in clay-lean formations and for clay fraction estimation in clay-rich formations (Passey et al., 2010).

Due to the complexity of unconventional reservoirs, traditional methods are not always successful. Some research has been published recently on the application of advanced machine learning techniques, such as artificial neural network (ANN) and support vector machine (SVM) methods, to TOC prediction from log data (Tan et al., 2015; Liu et al., 2013), which implies that relationships between TOC and log data are intrinsically nonlinear. Both ANN and SVM models learn from data and establish relationships by generating nonlinear input-output mapping functions through the learning process. Due to the complexity of unconventional reservoirs, traditional methods are not always successful. Some research has been published recently on the application of advanced machine learning techniques, such as artificial neural network (ANN) and support vector

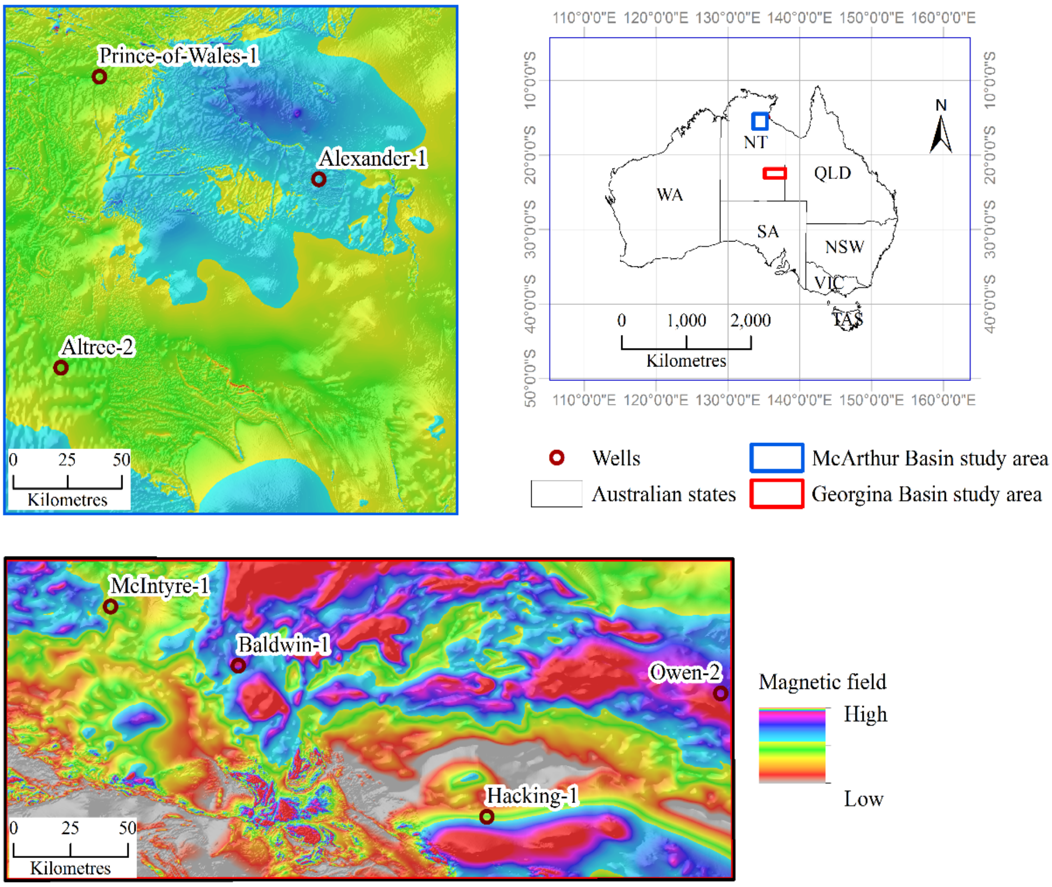

Figure 1: Study areas (color-coded with magnetic field values) and selected wells.

machine (SVM) methods, to TOC prediction from log data (Tan et al., 2015; Liu et al., 2013), which implies that relationships between TOC and log data are intrinsically nonlinear. Both ANN and SVM models learn from data and establish relationships by generating nonlinear input-output mapping functions (Bishop, 2003; Vapnik, 1998).

In this study, we apply traditional and advanced methods for establishing TOC profiles for seven wells drilled through onshore unconventional reservoirs in the Northern Territory, Australia (Figure 1). We use core TOC measurements to calibrate TOC models and evaluate accuracy. First, we apply traditional methods that use the density decrease and resistivity-transit time separation to identify organic-rich intervals and to predict TOC. Then, the multiple linear regression (MLR) statistical technique is used for modelling a linear relationship between the log responses and the TOC. Finally, nonlinear models are developed for predicting TOC using the ANN and SVM approaches.

METHOD AND RESULTS

The traditional methods we use include two petrophysics based methods, namely, the method that employs low density of organic matter to calculate TOC from bulk density and the ΔlogR method that uses separation between slowness (reciprocal of velocity) and logarithm of resistivity in organic rich intervals. Another traditional approach we apply is the Multiple Linear Regression (MLR), a statistical method modelling TOC content from five wireline logs such as gamma ray (GR), neutron porosity (NPHI), bulk density (RHOB), logarithm of resistivity (logLLD) and transit time (DT). For modelling nonlinear relationships, we apply a feedforward ANN, the Multilayer Perceptron (MLP) and an SVM-based function approximation method, Support Vector Regression (SVR). Both the MLP and SVR models are derived from the five wireline logs like the MLR.

The five methods mentioned above are applied for building and testing TOC models for each of the seven wells. The accuracy of the developed TOC-log data models is assessed by comparison of the predicted and core TOC data using the R^2, the squared correlation coefficient, and the mean square error (MSE). These statistics are applied to normalized predicted and core TOC values. Table 1 allows quantitative comparison of the described TOC models for all wells. The most accurate TOC estimates were generated by the nonlinear predictors followed by the MLR and traditional models. This indicates the geological complexity affecting the relationship between the log response and TOC in the area of interest.

Well	N	Density		ΔlogR		MLR		MLP		SVM	
		R^2	MSE	R^2	MSE	R^2	MSE	R^2	MSE	R^2	MSE
Alexander-1	49	0.40	0.05	0.53	0.05	0.70	0.04	0.91	0.01	0.93	0.01
Altree-2	161	0.64	0.05	0.67	0.07	0.76	0.04	0.86	0.01	0.84	0.01
Prince-of-Wales-1	29	0.12	0.16	0.32	0.08	0.35	0.08	0.56	0.05	0.82	0.02
McIntyre-1	62	2e-3	0.24	0.29	0.20	0.49	0.22	0.87	0.01	0.93	0.01
Owen-2	78	0.06	0.14	0.01	0.10	0.34	0.06	0.49	0.03	0.56	0.02
Baldwin-1	40	0.11	0.10	0.07	0.14	0.49	0.05	0.60	0.04	0.60	0.04
Hacking-1	30	0.13	0.10	0.08	0.15	0.33	0.10	0.52	0.05	0.86	0.01

Table 1. Models' performance for normalised TOC

The nonlinearity in the relationship between the log data and core TOC is graphically illustrated in Figure 2 in comparison with TOC predictions from the three most accurate models such as MLP, SVM and MLR (hereafter, top models). The TOC contours are shown in the RHOB-logLLD coordinate plane for Altree-2 and McIntyre-1. The core TOC contours demonstrate the most complicated patterns for both wells. The shapes and spatial distributions of the contours for predicted TOC are smoother and less complex. Visually, the MLP TOC contours are most similar to those for the core TOC from Altree-2. For McIntyre-1, the SVM TOC contours are almost identical to the core TOC. The MLR TOC contours appear to be much smoother, missing significant detail unlike the nonlinear models.

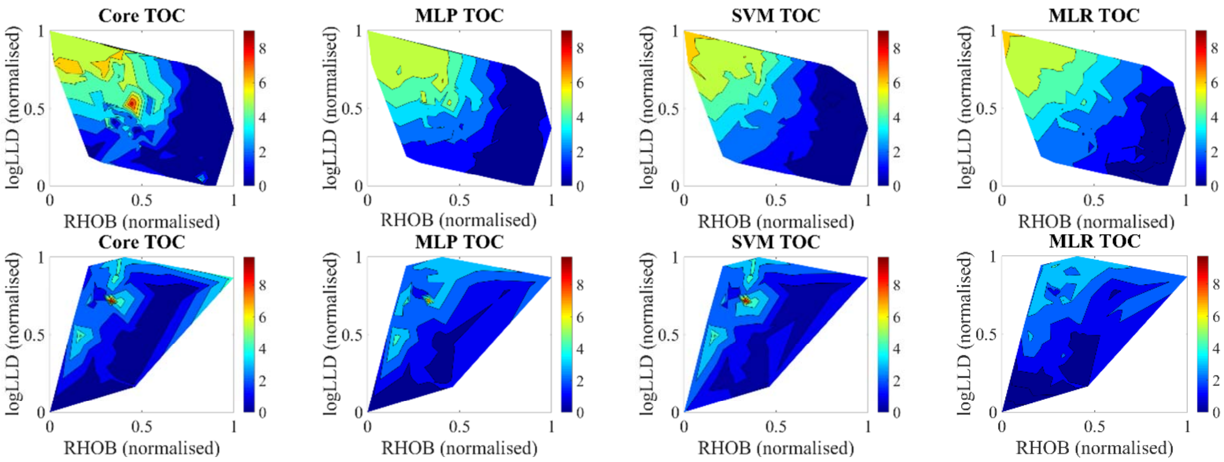

Figure 2: Core TOC contours for normalized RHOB and logLLD at Altree-2 (top row) and McIntyre-1 (bottom row) compared to TOC contours from MLP, SVM and MLR models. The Altree-2 MLP TOC and McIntyre-1 SVM TOC contours are most similar to those for the core TOC. The MLR TOC contours are much smoother, missing significant detail in comparison with the nonlinear models.

Furthermore, the three top models are used for generating continuous vertical TOC profiles. Figure 3 compares modelled TOC versus actual TOC for McIntyre-1. The accuracy of the top models varies dramatically (Table 1). While the MLP and SVM predictions are close to all core TOC values, the MLR model failed to accurately generate marginal values (greater than 4 wt% or near zero). The simulated TOC values outside the depth interval containing the core TOC values diverge for different models in comparison with the simulated TOC values within the interval. The SVM and MLP-generated TOC profiles show similar trends. In the absence of core TOC values, identification of the most accurate TOC simulations is problematic. The top three TOC models for all wells also are run on the other well data sets. Among all simulations, only the Altree-2 top three models produce TOC values that are correlated positively with core TOC values from Alexander-1. The TOC models show similar performance indicators with R^2 from 0.6 to 0.75 and MSE between 0.03 and 0.05.

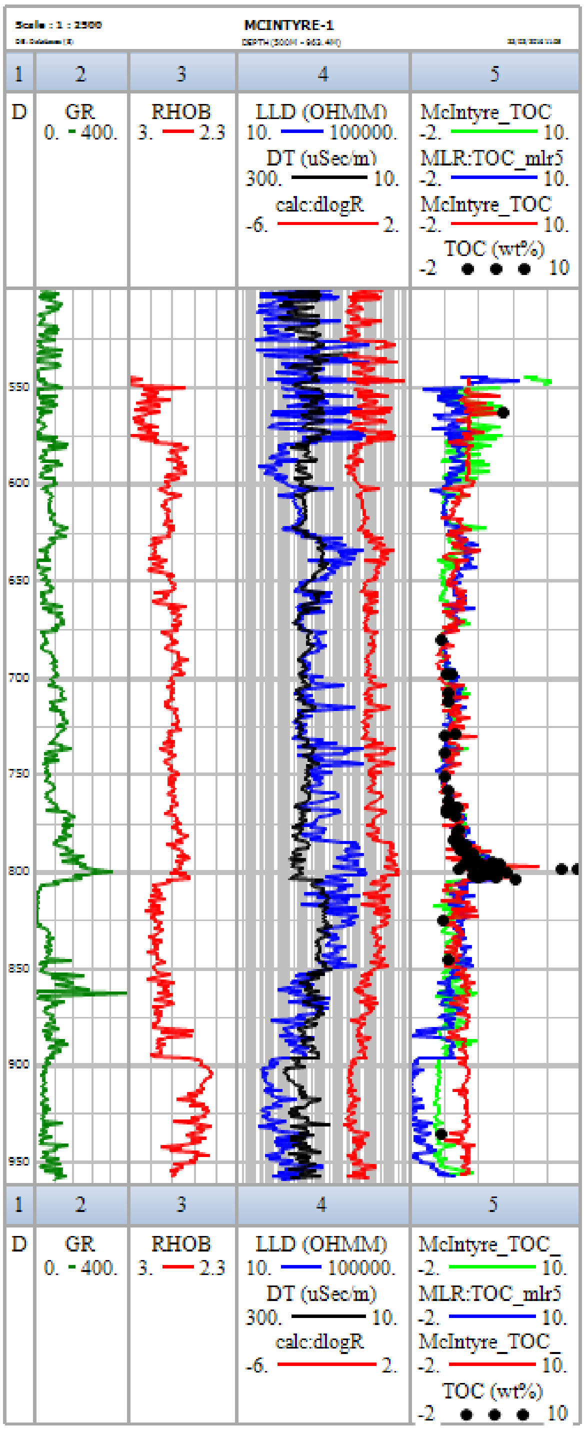

Figure 3: McIntyre-1 logs and TOC predictions: (1) depth, (2) GR, (3) RHOB, (4) logLLD and DT separation and ΔlogR, and (5) core TOC in comparison with MLR, MLP and SVM predictions.

CONCLUSIONS

In this study, wireline logs from seven wells in the McArthur Basin and Georgina Basin were used for TOC prediction. The prediction quality of traditional and machine learning methods was compared. The most accurate TOC estimates were generated by the nonlinear ANN- and SVM-based predictors, followed by the MLR and traditional models. Both nonlinear methods have an excellent capacity to generalise the basin-specific relationships which makes them applicable to cross-well TOC prediction for wells with similar geological structure. The accuracy of the cross-well prediction was substantially higher in the McArthur Basin, where the TOC model trained on one well was applied successfully to an adjacent well.

Further improvement of TOC modelling from log data is possible if input data are preprocessed to identify groups with similar geologic and geophysical characteristics. The data set of core TOC and log data then can be split accordingly and used for development of group-specific TOC prediction models. This will help to produce reliable TOC models with a high capacity for generalization on unknown data with similar characteristics. Such data preprocessing can be realized by applying a clustering algorithm such as self-organizing maps (Kohonen, 2001).

ACKNOWLEDGMENTS

The log data collection used in this study was kindly provided by the Northern Territory Geological Survey (http://www.nt.gov.au/d/Minerals_Energy/).

REFERENCES

Bishop, M., 2003, Neural networks for pattern recognition: University Press, Oxford.

Dellenbach, J., Espitalie, J. and Lebreton, F., 1983, Source rock logging: Transactions of the Eighth European Symposium, SPWLA, paper D.

Kohonen, T., 2001, Self-Organizing Maps: Springer: Berlin.

Liu, Y., Chen, Z., Hu, K. and Liu, C., 2013, Quantifying total organic carbon from well logs using support vector regression. GeoConvention, Canadian Society of Petroleum Geologists.

Passey, Q. R., Bohacs, K., Esch, W. L., Klimentidis, R. and Sinha, S., 2010, From oil-prone source rock to gas-producing shale reservoir - geologic and petrophysical characterization of unconventional shale gas reservoirs: International Oil and Gas Conference and Exhibition, SPE.

Schmoker, J. W., 1979, Determination of organic content of Appalachian Devonian shales from gamma-ray logs: AAPG Bulletin, 63(9), 1504–1537.

Passey, Q. R., Creaney, S., Kulls, J. B., Moretti, F. J. and Stroud, J. D., 1990, A practical model for organic richness from porosity and resistivity logs: AAPG Bulletin, 74(12), 1777–1794.

Tan, M., Song, X., Yang, X. and Q. Wu, Q., 2015, Support-vector-regression machine technology for total organic carbon content prediction from wireline logs in organic shale: A comparative study: Journal of Natural Gas Science and Engineering, 26, 792–802.

Vapnik, V. [1998] Statistical learning theory: Wiley.

Yan, F. and Han, D.-H., 2013, Measurement of elastic properties of kerogen: 83rd Annual International Meeting, SEG, Expanded Abstracts, 2778–2782.

Zhu, Y., Martinez, A., Liu, E., Harris, C., Xu, S., Payne, M. A. and Terrell M., 2010, Understanding geophysical responses of shale gas plays: Shale Workshop, EAGE.