Petroleum and Coal

Article **Open Access**

Evaluating Medium Decision Tree Model, Support Vector Machine Rational Quadratic Gaussian Process Regression to Estimate the Total Organic Carbon of Shale Gas Reservoirs

Sayed Gomaa[1,2], Mohamed Mongy[3,], Ramadan Emara[1], Ashraf Fahmy[1,2], Attia Attia[3]*

[1] *Mining and Petroleum Engineering Department, Faculty of Engineering, Al-Azhar University, Nasr City, Cairo, Egypt.*
[2] *Department of Petroleum Engineering, Faculty of Engineering and Technology, Future University in Egypt.*
[3] *Department of Petroleum Engineering and Gas Technology, Faculty of Energy and Environmental Engineering, British University in Egypt (BUE), Elshorouk city, Cairo, Egypt*

Received December 22, 2022; Accepted May 26, 2023

Abstract

As a result of an energy crisis due to the Russian-Ukrainian war, the eyes of great countries such as America and others began to turn strongly towards exploiting unconventional resources to increase the oil and gas production. The first step in exploiting unconventional sources is to estimate the Total Organic Carbon (TOC). TOC measurements are expensive as well as time consuming, as samples of cuttings or core samples must be present to do the required lab tests. This issue encouraged the researchers to develop mathematical correlation to estimate the TOC. The paper aims at evaluating three of machine learning models namely medium decision tree model (MDT), support vector machine (SVM) and rational quadratic Gaussian process regression (GPR) learned based on well logs data for estimating the TOC. To reach this target, 334 datasets of TOC a function of gamma ray, formation resistivity and sonic transit time. The results showed that rational quadratic Gaussian process regression (GPR) has higher accuracy than other models in estimating TOC. GPR achieved correlation coefficient of 0.91 with root mean square error (RMSE) of 1.01% and mean average error (MAE) of 0.74%.

Keywords: *ANN; SVM; Medium decision tree; GPR; TOC; Well logs.*

1. Introduction

Organic matter in soil and sediment is located in almost all terrestrial and aquatic settings, and its miles broadly dispersed all through the surface [1]. Organic compounds in soil and sediments vary from easy monosaccharaides and carbohydrates to greater complicated proteins, lipids, waxes, and natural acids. The following properties are essential characteristics of organic matter; with the presence of metal ions as well as hydrous oxides, it forms both "water-soluble" as well as "water-insoluble" complexes, it binds soil particles together by interacting with clay minerals, it can desorb and adsorb both naturally present and chemically introduced organic substances and it absorbs as well as releases nutrients from the plants. Total organic carbon (a degree that represent the presence of one of the chemical factors in the organic matter, which is most of the times considered as an index of soil or sediment development) is a main part of formation characterization due to those characteristics. Its presence or absence can cause reactions which are triggered with the presence of chemical compounds within the soil or sediment. As a part of an ecological hazard evaluation statistics package, figuring out overall natural carbon (TOC) for soils and sediments is often required for pollutant analysis. The contents of the TOC may be used to research the character of the

pattern place (for example, in a sedimentary region) or to normalize a part of the analytical chemistry dataset in a qualitative way [2].

2. Types of carbon in soil and sediments

Three main types of carbon can be identified in soil and sediments [3]. These forms are inorganic, elemental, and organic carbons. The organic matter quality in the sediment is important for the distribution and bioavailability of sediment-associated contaminants. Elemental forms of carbon consist of graphite, charcoal, soot, and coal. The main sources of elemental carbon in soils and sediments are the products of incomplete combustion of organic matter (i.e. charcoal, graphite and soot), from geological sources (i.e., conversion or burn these materials. Inorganic types of carbon are derived from geological sources or natural soil materials. Carbonate is the most common inorganic type of carbon found in the formations which are made from soil and sediments. The two major carbonate minerals present in soils and sediments are dolomite ($CaMg(CO_3)$ 2) and calcite ($CaCO_3$), while other forms (e.g. siderite, $FeCO_3$) may be present depending on the area where the soil was created or the source of the sediment. It's worth noting that agricultural inputs can cause calcite and, to a lesser extent, dolomite to appear in soils and sediments (i.e. liming operations). Organic carbon is produced in a natural way by the decomposition process of plants and animals. There are many distinct types of organic carbon in soils and sediments, ranging from newly deposited litter (e.g. leaves, branches, twigs) to fully decomposed forms like humus.

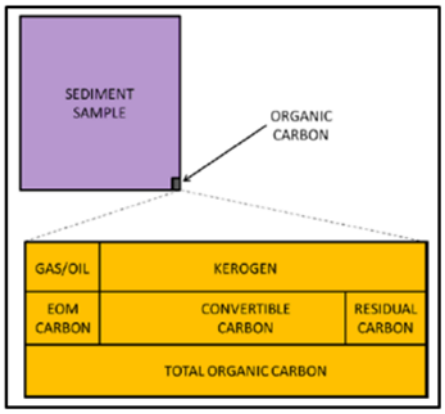

Fig. 1. Carbon distribution in a sediment sample (after [6]).

Soil and sediments are the most common habitats. The sum of inorganic and organic carbon is known as total carbon (TC). TOC content can be calculated directly or by subtracting total and inorganic carbon content. Total carbon becomes the value of organic carbon in the absence of inorganic carbon types in soils and sediments. The first step in determining hydrocarbon generation potential is to conduct a TOC analysis. Organic carbon only makes up a small fraction of a sediment sample volume, as seen in Fig. 1. (typically, shale). A TOC score of 1 wt% means that only 1 gram of organic carbon occurs in 100 grams of silt [4]. Coal, on the other hand, typically comprises between 50 and 70 percent by volume of TOC [5].

This is frequently the case; thus, total organic carbon measurements should produce similar results. However, inorganic types of carbon can be detected in the samples that are found in areas where the parent material/geology is dolomite, limestone, or another carbonate-bearing mineral. In arid areas, inorganic carbonates in soils and sediments may contain more carbon than organic carbon sources. Although there are a few non-destructive technologies being researched, the fundamental idea for estimating TOC depends on destroying the organic matter in the soil or sediment. Biological materials can be destroyed by chemicals or high-temperature heat. All carbon structures present in the sample are transformed to carbon dioxide, this is due to the existence of inorganic carbonates, which is then quantified either in direct or in-direct way and transfigured to total organic carbon or total carbon content.

These approaches can be either quantitative or semi-quantitative, and choosing the approach depends on the process of eliminating the organic matter and the ways of quantifying or detecting the carbon present. There are qualitative methods that can correctly detect the type of carbon components (e.g., sugars, carbohydrates, lignin) in the sample but are not yet developed enough to quantify the carbon content.

3. The traditional mathematical TOC estimation methods

The Devonian shale formation was defined as a matrix of four component with interstitial pores, organic material, and pyrite, and on that basis the smoker correlation for Devonian shale formation was created. The formation total density was calculated with respect to density as well as the partial volume of the previously mentioned component. The density of pyrite was 5.0 g/cm^3, organic matter density was set to be 1 g/cm^3 and the matrix density was set to 2.69 g/cm^3.

The previously mentioned assumptions made Sckmoker uses Eq.1 to determine the organic amount in volume percentage. Converting volume to weight percent will give the value of TOC in weight percent as there is well reliable link in Devonian shale between weight percentage and volume percentage [7]. According to the author a huge percentage of Appalachian basins (around 135000 square kilometer) will take advantage from this method. Schmoker and Hester mentioned that TOC may be calculated as a function of bulk density ρ_B) and the density of organic material (ρ) as in Eq. 1 [8].

$$TOC(vol\%) = \left(\frac{\rho_B - \rho}{1.378}\right) \quad (1)$$

In the Bakken shale formation, Eq. 2 was utilized to refine the Schmoker model. The pyrite-organic matter volume link, first discovered in Devonian shale, was supposed to apply to the Bakken formation as well. By assuming fixed pore fluid density and porosity profiles a total organic carbon correlation (Eq.2) for the Bakken formation (lower and upper member) is obtained. Values of 1.01, 2.68 g/cm^3 for organic matter density and matrix density respectively and a ratio of 1.3 between the organic matter to carbon content are used in simplifying the previously mentioned relationship. The results from 59 laboratory tests which were taken from 39 Bakken shale wells were compared by Schmoker and Hester and they found out that the organic content varies from 6 to 20 weight percentages to those of Eq.3. They concluded that Eq.3 predicts a good estimate of TOC with an average absolute deviation (AAD) of 1.1 % [8].

$$TOC(wt\%) = \frac{[(100\rho_o)(\rho_B - 0.9922\rho_{mi} - 0.039)]}{[(R\rho)(\rho_o - 1.135\rho_{mi} - 0.675)]} \quad (2)$$

$$TOC\ (wt\%) = \left(154.497 * \frac{1}{\rho_B}\right) - 57.261 \quad (3)$$

where ρo stands for the density of organic material (g/m^3); ρ_B is the bulk density; R is the weight percentage for organic matter to weight percentage of organic carbon ratio; $\rho_{mi}$is the volume weighted average density of grain and pore fluid in grams per cubic meter.

Schmoker approach assumes that the formation bulk density as well as porosity are constant values, and because the occurrence or the missing of low-density organic kerogen there are variations in bulk density.

Passey *et al.* devised a workable approach that uses the appropriate sonic log scale and deep resistivity logs. The logs should match the organic lean and the water saturated intervals, and the match should be used as the baseline. The presence of organic stuff will lead to a difference between the two curves. The separation can be calculated using the formula below [9]:

$$\Delta logR = log_{10}\left(\frac{R}{R_b}\right) + 0.02 * (\Delta t - \Delta t_b) \quad (4)$$

Then, the TOC is calculated by the following equation

$$TOC = \Delta \log R * 10^{(2.297 - 0.1688 * LOM)} \quad (5)$$

where, R: rock deep resistivity in Ω·m; Rb : baseline deep resistivity in Ω·m; and △t, △tb is the transit time of rock and baseline respectively; μs/m and LOM is the level of organic maturity.

There are two major problems concerning the ΔlogR methodology. The first problem is that it assumes a constant composition for the rock, as well as constant texture and compaction. This can lead to erroneous target shale calculations. The organic rich shale texture and composition vary significantly depending on the source [10]. The method accuracy for resistivity and acoustic transit time is low, this is because the assumption of a (1:50) linear relationship (constant slope of (0.02), Eq.4) between the logarithmic resistivity logs and the

porosity. Another weakness in using ΔlogR approach is the use of LOM, which is rarely used in measuring the organic matter thermal changes [4].

Using the data from four wells drilled with water, as well as oil-based mud and different diversity of formations, Schmoker and ΔlogR models were investigated by Charsky and Herron. The models showed poor accuracy as AAD for the TOC values from the core is 1.6 weights percent for the Schmoker technique and 1.7 weights percentage for the log R approach [11].

Wang *et al.* used sonic, density, gamma ray (GR) and resistivity logs to modify the sonic/density-based logR models for TOC assessment. To eliminate the presumed linear approximation, The Log R was redefined with estimated slopes associated to hit on shale in their models. They also proposed that instead of LOM, they utilized more popular thermal sensors (Tmax or vitrinite reflectance (Ro)), because LOM involves a conversion between (Tmax or Ro percent) and LOM, that can cause problems in practice [12].

According to a literature review, present TOC models either cannot accurately forecast TOC or require time-consuming of extensive lab work to find the fitting parameters. The goal of this research paper is to create a novel robust empirical correlation that can be used to forecast TOC with greater accuracy using artificial neural networks and conventional well logs.

Developing different optimization methods is one of the main aims of any researcher especially if this research is having a direct influence on a high impact industry like the petroleum sector. Since the start of the petroleum industry in 1940 different optimization techniques have been developed and those techniques helped in solving problems like asphaltene precipitation, well placement and wettability. Fig. 2 below shows different optimization techniques used in developing petroleum industry.

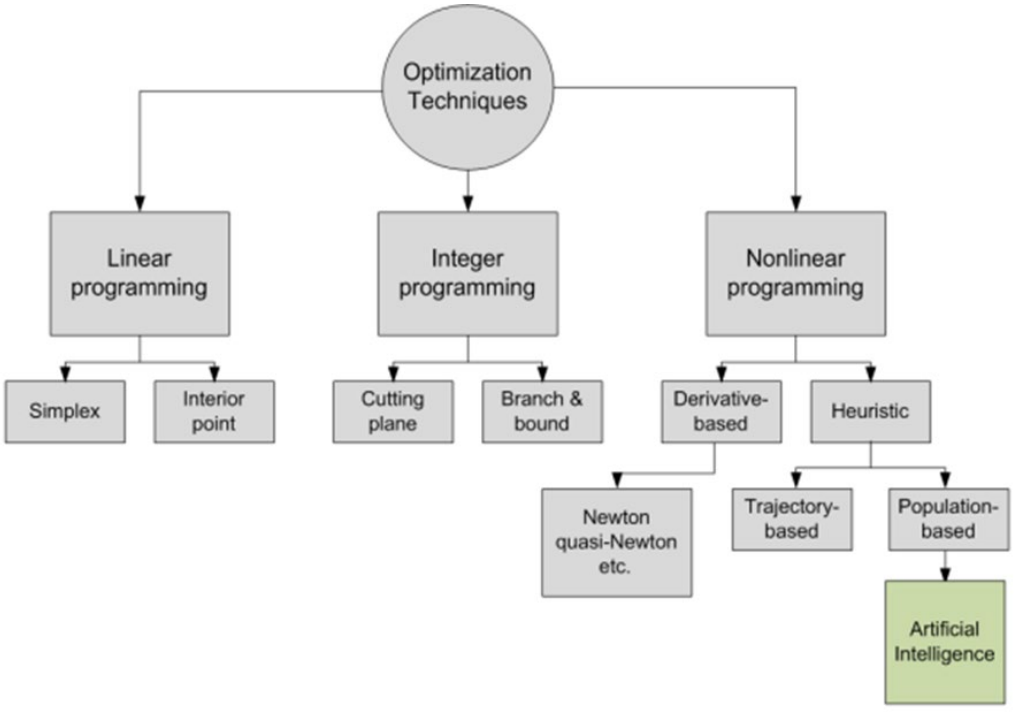

Fig. 2. Group of optimization techniques (modified after [13]).

Artificial intelligence is defined as the ability of intelligent agents to perform continuous learning in the corresponding environment and perceiving certain activities [14-15].

Artificial Neural Networks (ANNs) consist of an input and output layers and layer(s) between them (hidden layers). Each hidden layer consists of number of nodes. Every node has certain weight which is adjusted during the learning process to produce the expected output [16]. Recently artificial neural technique was applied extensively in oil and gas industry [17-18]. The

researcher succeeded to use this technique to develop models to predict the formation and fluid properties [19-21], reservoir production performance, water flood performance [22-23], drilling parameters [24] and others [25-29].

TOC is one of the essential parameters that improve the assessment process in reservoir evaluation as well as source rock characterization. High TOC value is an indicator that we have a high potential of hydrocarbon present. TOC also aid in determination of the geophysical characteristics of shale gas resources. One of the traditional approaches in determining the hydrocarbon potential in reservoirs is Rock pyrolysis assessment. TOC levels in the reservoir were determined previously using the conventional interpretation of gamma ray, resistivity, and acoustic recordings. The previously mentioned techniques can give an acceptable estimate to TOC levels in the reservoirs however they are time consuming, expensive, and hard to implement. Using the artificial intelligence approach is the new trend used by the engineers to predict the TOC from well log data.

A radial base function (RBF) model was built by Tan *et al.* to estimate the TOC values of a Chinese gas reservoir. A lab measured TOC data was used to create the model. Neutron log, density log, deep induction log, acoustic logs and gamma ray logs data were utilized to develop the model. The mean square error of the developed model was as low as 0.3 and the regression were high as 0.86 in TOC predictions. Tan *et al.* claimed that the created RPF model can give a better estimate for TOC values than traditional ANN methods. According to Tan et al. the main advantage of using RBF that there is only one hidden layer and the other number of hidden layers does not have to be defined in advance [30].

A two-step approach is developed by Sfidari et al. to give an estimate value for TOC using well logs. For the first step approach the well log data was grouped with respect to electro-facies (EF). They used statistical approaches like "K-means clustering" and "hierarchical analysis" against intelligent data clustering approaches (self –organizing maps) to find the EF. Cluster validity tests were also done in order to determine the optimum strategy for grouping the data to a specific number of EF, and they developed a well-defined ANN for each EF to estimate total organic carbon numbers from the well logs data. In their second step they employed a similar Ann model but over the full interval to estimate the total organic carbon. They made a comparison between the models developed from the two approaches, as well as they compared the results with traditional Log R technique, from the comparison they find out that grouping the data into different EF using self-organizing maps will lead to more accurate predictions than making a single model for the whole interval. They also mentioned that intelligent models are more reliable than the traditional methods in general. The mean square error values achieved by them were as low as 0.0073 [31].

An extreme machine learning (EML) algorithm utilized by Shi *et al.* to estimate the TOC values using the well-log data in a study published in 2016. They made a comparison between a single-layer feed forward model and a multilayer Levenberg-Marquardt model. From the comparison they found out that EML is much faster with a lower computing cost in addition to its accuracy compared to ANN model. There RMSE was about 0.30 and the regression coefficient is 0.93 [32].

A 442 data point from Barnett shale was utilized by Mahmoud et.al to construct an ANN model to predict the TOC value of the reservoir. Logs like bulk density, sonic transit, gamma ray were used as input data for the model. With an average absolute deviation percentage of 0.91 and a regression coefficient of 0.93 the model was able to predict the TOC for Barnett shale [33].

A new approach was developed by Asante-Okyere et al to predict the performance of a conventianl neural network (CNN), as the author used the mineral composition of the shale rock to enhance TOC predicting accuracy. The model was based on the input data provided from well logs as gamma ray and density logs beside the mineral components of fieldspar, calcite, dolomite, pyrite and quartz. The results of the experimental studies showed that adding the mineralogy parameter to the input data improves the error rate, the influence of each mineral content on the model was quantified by variable significance analysis. It was found that filedspar and pyrite have the higher contribution [34]. Some authors did not publish their ANN models in their papers, while the others used a narrow range of data.

This paper aims at evaluating three of machine learning models namely medium decision tree, support vector machine (SVM) and rational quadratic Gaussian process regression (GPR) learned based on well logs data for estimating the TOC.

4. Data collection and preparation

334 datasets were collected from the literature to perform this study [9,12,35-38]. These datasets are characterized by wide range of sonic transit time, density, gamma ray, true resistivity, and TOC values. Table 1 show the statistical analysis of the data. The ranges of the data are as follows; gamma ray range is 15.66-338.67, sonic transit time range is 53-140, true resistivity range is 0.49-183.2, density range is 2.07-2.79 and TOC range is 0.06-9.88. The previously mentioned datasets groups are divided randomly to three subgroups. 70% of the data are allocated for the first subgroup which is allocated for training the model, other 15 % are allocated to validate the model and the last 15 % are used to test the accuracy and generalize the model capabilities.

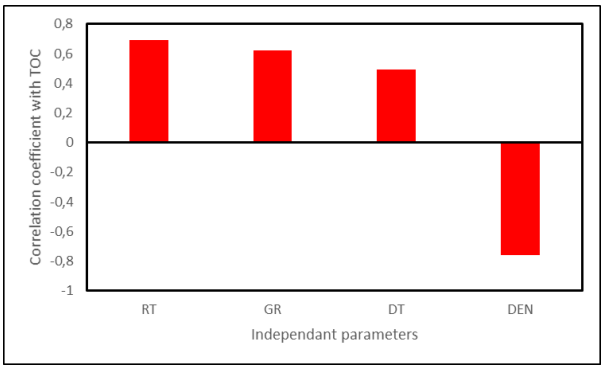

Fig. 3. Correlation coefficients between the input parameters and TOC.

From Fig. 3 it is clear that TOC value is in direct proportional relationship with true resistivity, sonic, as well as gamma ray. However, it also shows that it is inversely proportional to density values. The electrical conductivity of the formation is determined from resistivity logs. In shale reservoirs the main parameters that affect the overall conductivity of the rock are the formation water, conductive clay as well as mineral components like pyrite. On the other side, hydrocarbon fluids are non-conductive as well as they can displace formation water which increase the overall value of the resistivity.

The radioactivity of the rocks is determined by the Gamma ray logs. Gamma ray logs function perfectly in marine shale reservoirs as it gives true indicators about the amount of hydrocarbon present; this is due to the uranium enrichment in hydrocarbon. For lacustrine shale reservoirs, Gamma ray logs are detecting over all clay content better than the amount of hydrocarbon present. The capability of the formation to transmit sound waves or seismic waves is determined by the sonic logs. Sonic logs also provide the interval transit time of rock formations. The presence of hydrocarbon in the shale formation is translated from the sonic logs by having high sonic values. Rock density is measured by density logs, as hydrocarbon is present in rocks the density values will decrease; this is because the density of hydrocarbon is low compared to formation minerals [39].

Table 1. Statistical analysis of the data used in this work.

	RT	DEN	AC	GR	TOC
Mean	14.51	2.48	92.43	104.99	3.41
Standard error	1.16	0.01	0.91	2.19	0.13
Median	6.47	2.5	88.58	96.28	2.69
Mode	17.6	2.6	86	119.6	1.9
SD	21.24	0.13	16.62	40.02	2.37
Sample variance	451.06	0.02	276.21	1601.27	5.62
Kurtosis	20.32	0.6	0.35	2.59	0.27
Skewness	3.89	-0.54	0.59	0.98	1.04
Range	182.71	0.72	87	323.01	9.82
Minimum	0.49	2.07	53	15.66	0.06
Maximum	183.2	2.79	140	338.67	9.88

5. Results and discussion

For the cross validation, the 334 data sets are partitioned into five folds. For each fold, the MDT, VSM, and GPR techniques were used to develop a model is for predicting the total organic content using the out-of-fold observations, then the model performance is assessed using in-fold data and finally the average test error over all folds is calculated. This method gives a good estimate of the predictive accuracy of the final model trained using the full data set. The optimizable model parameters are presented in Tables (2-3).

Table 2. Optimizable hyperparameters of GPR model.

Bases function	Linear
Kernel function	Nonisotrobic rational quadratic
Kernel scale	14.6
Sigma	21.3246
Optimizer	Bayesian optimization

Table 3. Optimum design parameters of SVM model.

Epsilon	0.1779
Kernel function	Gausian
Kernel scale	2
Sigma	21.3246
Optimizer	Bayesian optimization

Fig. 4 shows the regression results of GPR model. After training the GPR model, the response plot displays the predicted TOC versus record number. As we use cross-validation, then these predictions are the predictions on the held-out observations. In other words, each prediction is obtained using a model that was trained without using the corresponding observation. The predicted values trace the true values as shown in Fig. 4. The residuals plot shown in Fig. 5 is used to check model performance. The residuals plot displays the difference between the predicted and true responses.

Table 4. Evaluation of SVM, MDT, and GPR models.

Evaluation parameters	SVM	Medium decision tree	Rational quadratic GPR
R	0.87	0.83	0.91
RMSE	1.1693	1.3294	1.0123
MAE	0.81991	0.94251	0.74852

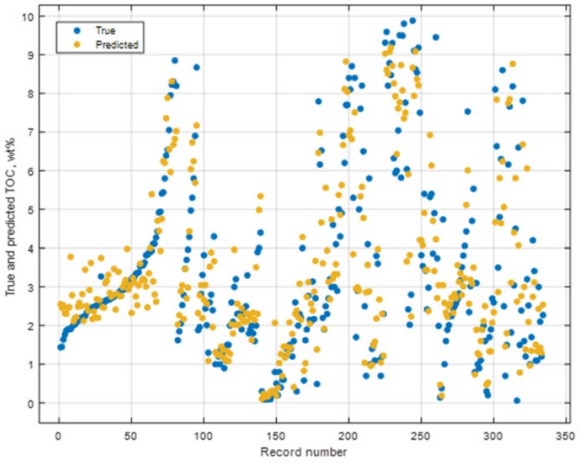

Figure 4. True and estimated values of TOC versus record number.

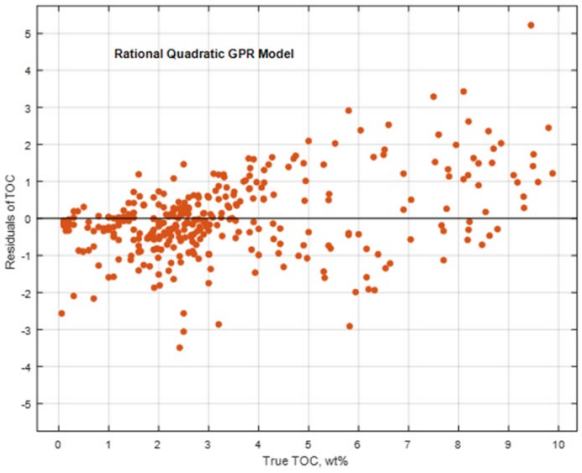

Figure 5. Evaluating the GPR model using residual plot.

As we see from this figure the differences between the true values and the values calculated by the GPR model are located around the zero and this indicates to the high accuracy of the

proposed GPR model. Fig. 6 depicts the cross plot of predicted values by GPR model versus the actual values which gives higher correlation coefficient (R = 0.91) than that of SVM (R = 0.87) and MDT (R = 0.83 shown in Figs. (7-8). The GPR model was optimized after 15 iterations to give minimum square error (MSE) of 1.01% and this shows the high performance of the proposed GPR model as depicted in Fig. 9. Table 4 summarizes the comparison among the three models in terms of correlation coefficient, root mean square error, and mean average error where the GPR model show the highest performance.

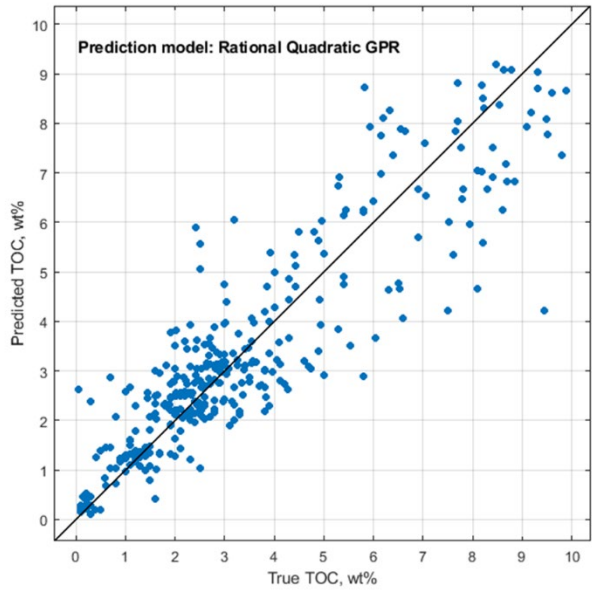

Figure 6. Evaluating the GPR model using cross plot.

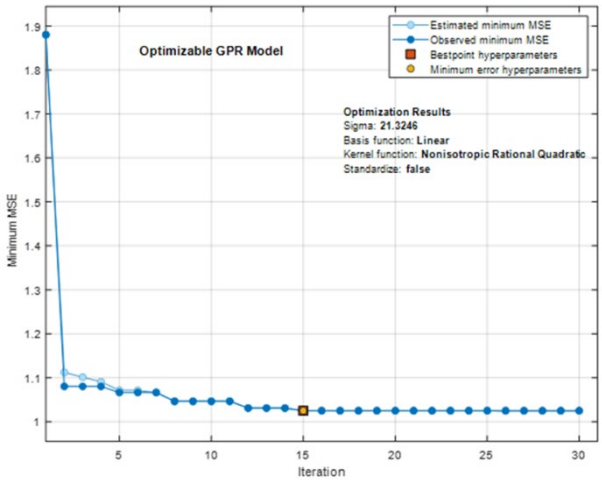

Figure 7. Minimum MSE of the GPR model.

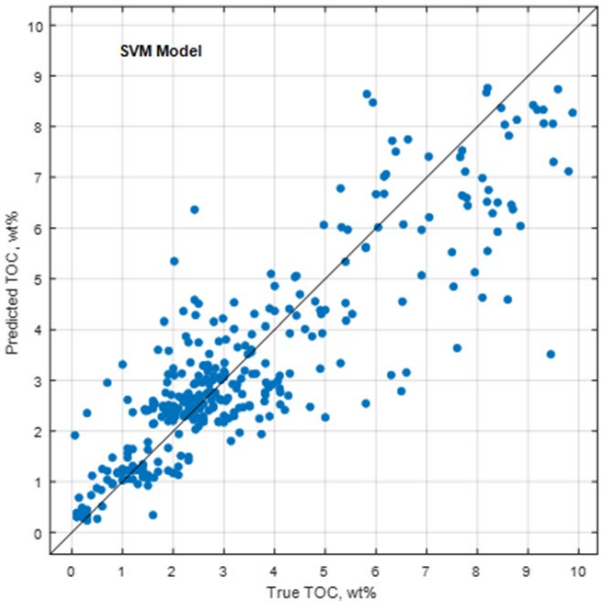

Figure 8. Evaluating the SVM model using cross plot.

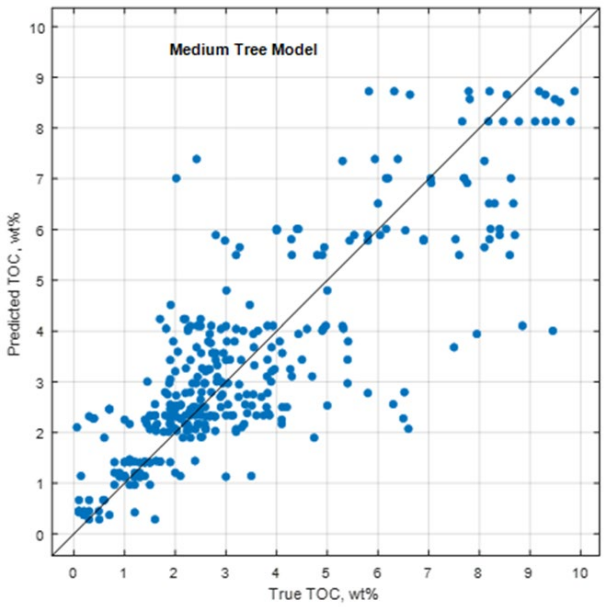

Figure 9. Evaluating the MDT model using cross plot.

6. Conclusion

Three of machine learning models namely medium decision tree model (MDT), support vector machine (SVM) and rational quadratic Gaussian process regression (GPR) are examined

to estimate the TOC. 334 datasets of TOC a function of gamma ray, formation resistivity and sonic transit time are used to achieve this target. The three models can estimate the TOC accurately. The rational quadratic Gaussian process regression (GPR) gives the highest accuracy with correlation coefficient of 0.91 with RMSE OF 1.01% and MAE of 0.74%.

Conflict of interest

The authors declare that there is no conflict of interest.

References

[1] Schnitzer M. Chapter 1-Humic substances: chemistry and reactions. Developments in soil science. 1978, Volume 8, 1-64.

[2] Schumacher BA. Methods for the determination of total organic carbon (TOC) in soils and sediments. Method NCEA-C-1282. EPA.

[3] Talley JW, Ghosh U, Tucker SG, Furey JS, Luthy RG. Particle-scale understanding of the bioavailability of PAHs in sediment. Environ Sci Technol., 2002; 36(3): 477-483.

[4] Jarvie D. Total Organic Carbon (TOC) Analysis: Chapter 11: Geochemical Methods and Exploration. (1991).

[5] Rice DW, Seltenrich CP, Spies RB, Keller ML. Seasonal and annual distribution of organic contaminants in marine sediments from Elkhorn slough, moss landing harbor and nearshore Monterey Bay, California. Environ Pollut., 1993; 82(1): 79-91.

[6] Berch H, Nunn J. Predicting potential unconventional production in the Tuscaloosa marine shale play using thermal modeling and log overlay analysis. https://doi.org/10.31390/gradschool_theses.3676

[7] Schmoker JW. Organic content of Devonian shale in western Appalachian Basin. AAPG Bulletin, 1980; 64(12): 2156-2165.

[8] Schmoker JW, Hester TC. Organic Carbon on Bakken Formation, United States Portion of Williston Basin. AAPG Bulletin, 1983; 67(12): 2165-2174.

[9] Passey QR, Creaney S. and Kulla JB. A Practical Model for Organic Richness from Porosity and Resistivity Logs. AAPG Bulletin, 1990; 4: 1777-1794.

[10] Rokosh D, Lyster S, Berhane H, & Snowdon L. Mineralogy and grain density of Alberta shale. AAPG Search and Discovery Article, 2010; 90122, 2011.

[11] Charsky A, Herron S. Accurate, direct total organic carbon (TOC) log from a new advanced geochemical spectroscopy tool: Comparison with conventional approaches for TOC estimation. Search and Discovery. 2013; 41162(1):1-7.

[12] Wang P, Chen Z, Pang X, Hu K, Sun M, Chen X. Revised models for determining TOC in shale play: Example from Devonian Duvernay shale, Western Canada sedimentary basin. Marine and Petroleum Geology. 2016; 70: 304-319.

[13] Rahaman MS, Islam J, Watada J, Vasant P, Alhitmi HK, Hossain TM. Artificial Intelligence Approach to Total Organic Carbon Content Prediction in Shale Gas Reservoir Using Well Logs: A Review. Artificial Intelligence, 2011; 7(1):1-0.

[14] Cohen PR, Feigenbaum EA, editors. The handbook of artificial intelligence,2014; Volume 3. Butterworth-Heinemann.

[15] Berente N, Gu B, Recker J, Santhanam R. Managing artificial intelligence. MIS quarterly. 2021; 1; 45(3).

[16] Wang SC. Interdisciplinary computing in Java programming. Springer Science & Business Media; 2003 Aug 31.

[17] Choubey S, Karmakar GP. Artificial intelligence techniques and their application in oil and gas industry. Artificial Intelligence Review. 2021; 54(5): 3665-3683.

[18] Alkinani HH, Al-Hameedi AT, Dunn-Norman S. Data-driven recurrent neural network model to predict the rate of penetration: Upstream Oil and Gas Technology. Upstream Oil and Gas Technology; 2021; 7: 100047.

[19] Gouda A, Gomaa S, Attia A, Emara R, Desouky SM, El-hoshoudy AN. Development of an artificial neural network model for predicting the dew point pressure of retrograde gas condensate. Journal of Petroleum Science and Engineering; 2022; 208: 109284.

[20] Gomaa S, Soliman AA, Mohamed A, Emara R, Attia AM. New correlation for calculating water saturation based on permeability, porosity, and resistivity index in carbonate reservoirs. ACS Omega. 2022; 7(4): 3549-3556..

[21] El-hoshoudy AN, Ahmed A, Gomaa S, Abdelhady A. An artificial neural network model for predicting the hydrate formation temperature. Arabian Journal for Science and Engineering. 2022: 47:11599-11608.
[22] Gomaa S, Emara R, Mahmoud O, El-Hoshoudy AN. New correlations to calculate vertical sweep efficiency in oil reservoirs using nonlinear multiple regression and artificial neural network. Journal of King Saud University-Engineering Sciences. 2022; 34(7): 368-375.
[23] Gomaa S, Soliman AA, Nasr K, Emara R, El-Hoshoudy AN, Attia AM. Development of artificial neural network models to calculate the areal sweep efficiency for direct line, staggered line drive, five-spot, and nine-spot injection patterns. Fuel, 2022; 317: 123564.
[24] Gouda A, Khaled S, Gomaa S, Attia AM. Prediction of the rheological properties of invert emulsion mud using an artificial neural network. ACS Omega. 2021; 6(48): 32948-32959.
[25] Abbasi T, Lim KH, Yam KS. Predictive maintenance of oil and gas equipment using recurrent neural network. In Iop conference series: Materials science and engineering, 2019; 495: 012067.
[26] Werbos PJ. Maximizing long-term gas industry profits in two minutes in Lotus using neural network methods. IEEE Transactions on Systems, Man, and Cybernetics; 1989; 19(2): 315-333.
[27] Ahmad Z, Zhang J, Kashiwao T, Bahadori A. Prediction of absorption and stripping factors in natural gas processing industries using feedforward artificial neural network. Petroleum Science and Technology, 2016; 34(2): 105-113.
[28] Venkatasubramanian V, Chan K. A neural network methodology for process fault diagnosis. AIChE Journal;1989; 35(12): 1993-2002.
[29] Ahmed DF, Khalaf AH. Artificial neural networks controller for crude oil distillation column of Baiji refinery. J Chem Eng Process Technol., 2016; 7: 272.
[30] Tan M, Liu Q, Zhang S. A dynamic adaptive radial basis function approach for total organic carbon content prediction in organic shale. Geophysics, 2013; 78(6): D445-459.
[31] Sfidari E, Kadkhodaie-Ilkhchi A, Najjari S. Comparison of intelligent and statistical clustering approaches to predicting total organic carbon using intelligent systems. Journal of Petroleum Science and Engineering, 2012; 86: 190-205.
[32] Shi X, Wang J, Liu G, Yang L, Ge X, Jiang S. Application of extreme learning machine and neural networks in total organic carbon content prediction in organic shale with wire line logs. Journal of Natural Gas Science and Engineering, 2016; 33(C): 687-702.
[33] Mahmoud AA, Elkatatny S, Mahmoud M, Abouelresh M, Abdulraheem A, Ali A. Determination of the total organic carbon (TOC) based on conventional well logs using artificial neural network. International Journal of Coal Geology, 2017; 179: 72-80.
[34] Asante-Okyere S, Ziggah YY, Marfo SA. Improved total organic carbon convolutional neural network model based on mineralogy and geophysical well log data. Unconventional Resources, 2021; 1: 1-8.
[35] Zheng D, Wu S, Hou M. Fully connected deep network: An improved method to predict TOC of shale reservoirs from well logs. Marine and Petroleum Geology, 2021; 132: 105205.
[36] Wang H, Wu W, Chen T, Dong X, Wang G. An improved neural network for TOC, S1 and S2 estimation based on conventional well logs. Journal of Petroleum Science and Engineering, 2019; 176: 664-78.
[37] Zhao P, Ma H, Rasouli V, Liu W, Cai J, Huang Z. An improved model for estimating the TOC in shale formations. Marine and Petroleum Geology, 2017; 83: 174-83.
[38] Zhu L, Zhang C, Zhang Z, Zhou X, Liu W. An improved method for evaluating the TOC content of a shale formation using the dual-difference ΔlogR method. Marine and Petroleum Geology, 2019; 102: 800-16..
[39] Bohacs KM. Contrasting expressions of depositional sequences in mudrocks from marine to non marine environs. In Shales and mudstones; 1988: Volume I, Basin studies, sedimentology, and paleontology,pp. 33-78.

To whom correspondence should be addressed: Mohamed Mongy, Department of Petroleum Engineering and Gas Technology, Faculty of Energy and Environmental Engineering, British University in Egypt (BUE), Elshorouk city, Cairo, Egypt, E-mail: mohamed.mongy@bue.edu.eg