Application of multiple artificial neural networks for estimation of total organic carbon content from petrophysical data.

Wafaa El-Shahat Afify

Department of geology, faculty of science, Benha univ., Egypt.

تطبيق تقنية الشبكات العصبيه الاصطناعيه لتقدير المحتوي الكربونى العضوى الكلى باستخدام البيانات البتروفيزيائيه.

الخلاصه.يؤثر المحتوى الكربونى العضوى الموجود بصخور المصدر على العديد من التسجيلات الكهربائيه, حيث تتميز صخور المصدر بقراءات مساميه وزمن مرور الموجات الصوتيه واشعاعات جاما عاليه يقابلها قراءات كثافه منخفضه بالاضافه الى مقاومه نوعيه اعلى من الصخور الاخرى. ويتضمن هذا البحث محاوله لتأسيس طريقه كميه للربط بين نتائج تسجيلات الآبار والمحتوى الكربونى العضوى باستخدام برامج الذكاء الاصطناعى حيث وظفت تقنية الشبكات العصبيه الاصطناعيه لاستنتاج المحتوى الكربونى العضوى الكلى من تسجيلات الآبار مع استخدام مثال من صخور خزان الكريتاوى العلوى فى الجزء الشرقى من شمال الصحراء الغربيه بمصر.

Abstract. *Total organic carbon content (TOC) present in the potential source rocks significantly affects the response of several types of well logs. They are characterized by higher porosity, higher sonic transit time, lower density, higher gamma-ray, and higher resistivity than other rocks. This paper attempts to establish a quantitative correlation between standard well logs (sonic, density, neutron, gamma-ray and resistivity) and total organic carbon by means of intelligent systems with an example from the Upper Cretaceous reservoirs, in the eastern part of the North Western Desert of Egypt. This dissertation utilizes the ability of neural networks to discover patterns in the data important for the required decision, which may be imperceptible to human brain or standard statistical methods. Thus the idea is not to eliminate the interpretation from an experienced petrophysicist but to make the task simpler and faster for future work.*

1. Introduction

Source rocks are commonly shales and lime mudstones that contain significant amounts of organic matter (Tissot and Welte, 1984), and has the capability to generate and expel enough hydrocarbons to form an accumulation of oil or gas. The most important factor controlling the generation of oil and gas is the hydrogen content of the organic matter (OM) (Hunt and Jaieson, 1956 and Hunt, 1996). The quantity of organic matter usually is expressed as total

organic carbon (TOC) and is measured with Rock-Eval technique. The response of various logging tools to the organic matter content of rocks is related to the distinctive physical properties of organic matter and to the tool's design. Organic matter disseminated throughout sediments has several distinct features: (1) it is often highly radioactive; (2) it is generally of low bulk density (1.03-1.1 g/cc); (Autric and Dumesnil, 1985); (3)
it has high hydrogen content; and (4) it is generally non-conductive or nearly non-conductive. As a result, natural gamma ray, porosity and resistivity logs frequently have a noticeable response to organic-rich rocks.

In this work a multiple networks systems developed by Bhatt and Helle, (2001) for predicting total organic carbon from wireline logs. The method is much more robust and accurate than a single network and the multiple linear regression method. The basic unit of a multiple artificial neural system is a multilayer perceptron network (MLP) whose optimum architecture and size of training data set has been discovered by trial and error, then the generated artificial neural network system has been successfully applied for predicting the TOC of the Upper Cretaceous reservoirs, in the study area, Fig.(1).
In summary, the main objectives of this study are:

• to investigate the ability of multiple artificial neural system , to predict TOC,

• to introduce unconventional solution to overcome the general absence of core samples from within source rock intervals, and the often incomplete, time-consuming and expensive geochemical laboratory analysis.

• to compare the performance of the proposed technique with that of conventional models.

• to apply the multiple artificial neural s technique to estimate the TOC of one of the studied wells (East Qarun Well), which has not been cored.

2. Total organic carbon-an overview.

Numerous studies have illustrated the potential value of wireline logs for source rock evaluation. Beers (1945), Swanson (1960), Fertle (1988), Schmoker (1981) and Hertzog et al. (1989) used gamma-ray spectral log for identifying organic-rich rocks. Schmoker and Hester (1983) proposed the use of the density log for estimating organic matter content. Dellenbach et al. (1983) and Hussain (1987) developed a method using the transit-time and gamma-ray curves to provide a parameter that relates linearly to organic richness. A method involving a combination of resistivity, density and sonic logs has been introduced by Meyer and Nederlof (1984). This method discriminates between source rocks and non-source rocks without attempting to quantify the organic richness from combination of logs. Mendelson and Toksoz (1985) applied multivariate analysis of log data to characterize source rocks. At last, Passey et al. (1990) invented a new technique called DLogR. This technique employs the overlaying of porosity logs (sonic, density and neutron) and resistivity log for identifying and calculating total organic carbon content.
Neural networks for quantitative analysis of source rocks from well logs have been demonstrated in several practical applications (e.g. Huang and Williamson, 1996), used artificial neural network (NN) modeling for source rock characterization, Kamali and Mirshady (2004) used DlogR and neuro-fuzzy (NF) techniques for determining TOC from well log data and Ali Kadkhodaie-Ilkhchi et al., (2008) integrates intelligent systems, (including genetic algorithms (Gas), fuzzy logic (FL), neural network NN, and neuro-fuzzy NF), and the concept of committee to develop an improved and more accurate model for TOC prediction in reservoir intervals.

3. Artificial Neural Networks.

The design and implementation
of intelligent systems with human capabilities is the starting point to design Artificial Neural Networks (ANNs). The original idea takes after neuroscience theory on how neurons in the human brain cooperate to learn from a set of input signals to produce an

answer. A neural network can be described as a massively parallel distributed processor made up of simple processing units called neurons, which has a natural tendency for storing experiential knowledge and making it available for use. Neurons are grouped into layers. In a multi-layer network there are usually an input layer, one or more hidden layers and an output layer. The number of neurons in the input layer corresponds to the number of parameters that are being presented to the network as input. The same is true for the output layer. The neurons in the hidden layer or layers are mainly responsible for feature extraction. Figure (2) is a schematic diagram of a fully connected three layered neural network.

Neural nets can be divided into two major categories based on the training methods, namely supervised and unsupervised neural networks.

Most of the neural network applications in the oil and gas industry are based on supervised training algorithms.

In a typical neural data processing procedure, the database is divided into three separate portions called training, calibration and verification sets (training, test and validation sets).

There are many types of ANN learning algorithms. One of the most popular is back-propagation and was developed by Rumel hart et al. (1986). An ANN that uses back-propagation (BP-ANN) always has an input layer, which contains several input nodes, an output layer, which contains one or several output nodes, and at least one hidden layer, which contains several nodes, (Fig.3). Nodes in the hidden and output layers are sometimes referred to as processing units. The processing units are nodes which are capable of manipulating data through the application of a certain transform function (e.g. the sigmoidal function, inverse hyberbolic tangent function, etc.). Nodes of adjacent layers are interconnected by weights which are initially randomized. In BP-ANNs, however, there is no interconnection amongst nodes of the same layer.

4. Data availability

In this study a multiple networks systems is developed to predict TOC from petrophysical data including sonic porosity (μs/ft), resistivity (ohm-m), neutron porosity, density (g/cm^3) and gamma-ray (API) logs, for 4 wells (N.B.Q-2X, N.B.Q-1X, Gindi Deep-1X and East Qarun) located at the eastern part of the North Western Desert of Egypt. For this purpose, 271 samples from the logged intervals of the Upper Cretaceous reservoirs, collected for Rock-Eval pyrolysis and measuring TOC. Core data are available for the first three wells. The dataset is divided into 218 data points used in the training set and 53 data points in the test set. 101 input data points were used in the validation set to predict the TOC of East Qarun well.

5. Physical relationships between TOC and input petrophysical data

There is a logical relationship between petrophysical data used and TOC content present in reservoir rocks. According to Fig. 4a–c, petrophysical data including GR, NPHI and DT show a direct relationship with TOC.

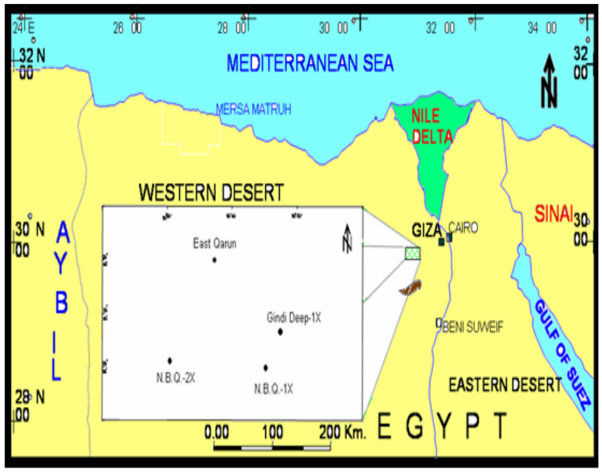

Fig.(1) Location map of the study area.

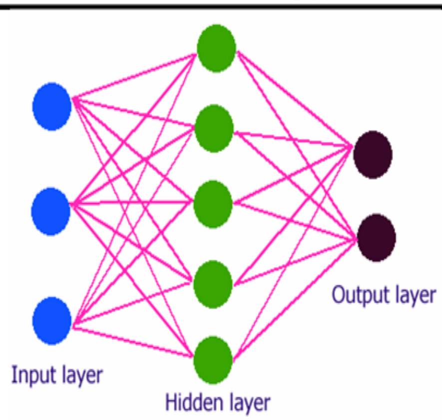

Figure (2): Schematic diagram of a three-layer neuron network.

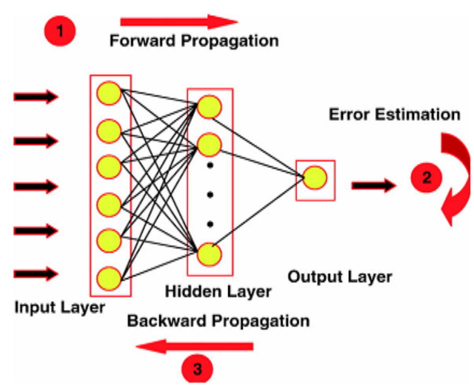

Fig. 3. Back-propagation multilayer ANN with one hidden layer.

This relationship is reversed for RHOB data. A gamma-ray tool measures the radioactivity of various formations. Generally, organic-rich rocks have high concentrations of radioactive elements including potassium, thorium and uranium and increase the γ-ray response. Neutron log reading is a response of hydrogen atoms concentration in rocks. The volume of organic matter in the formation has a direct relationship with hydrogen atoms content and porosity of the rock. Thus, neutron porosity increases in the organic-rich intervals. The sonic transit time (DT) is the reciprocal of the velocity of the compressional wave and is a function of formation lithology, porosity, type and distribution models of fluids (water, gas, oil, kerogen, etc.). With apparent DT value increase TOC content tends to elevate (Kamali and Mirshady, 2004). Density log measures the bulk density of the formation, a response of fluids and matrix constituent minerals density. Organic matters have a low density (about 1 g/cm3) and their concentration tends to reduce the bulk density of the rock. The resistivity log indirectly measures rock resistivity through variations in fluid saturation, because fluid content is a major control on the rock resistivity. Generally, organic-matter-bearing layers have higher resistivity than the other rocks. It is true, especially when kerogen becomes mature and generates hydrocarbon filling pore spaces.

6. Methodology.

The applied technique in the present work consists of a group of artificial neural networks, each one employs multi-layer perceptrons trained with the backpropagation algorithm. The input layer is comprised of six processing elements (PEs) representing the following input parameters: depth (ft.), sonic porosity (µs/ft), resistivity (ohm-m), neutron porosity (v/v), density (g/cc) and gamma-ray (API) logs, while the output layer is represented by a single TOC (wt.%). Table (1) summarizes the ranges of all input and output variables considered.

The design philosophy of the ANNs used in this study may be summarized by the following steps:

- Determination of model inputs
- Data handling and pre-processing
- Data division
- Network architecture design
- Weight optimization (training)
- Stopping criteria
- Model validation.

Table 1
The ranges of all input and output variables.

Parameter	Minimum	Maximum
Gamma ray (GAPI)	0.81	99.9
Neutron	0.0009	0.67
Bulk density (g/cm^3)	1.56	2.88
DT (US/F)	51.4	148
Resistivity (Ohm-m)	0.016	1626
Depth (Ft)	6030	13400
TOC(wt.%)	0.26	2.86

A number of simulation trials were conducted. The common key features of these trials may be summarized, as follows:

1. The use of three data subsets, including training, testing and validation sets. Cross-validation was used as the stopping criterion to avoid over-fitting. The available data were split so that they were statistically consistent. To check consistency, statistical measures such as average, median and standard deviation were calculated for all subsets. Table (2) shows a sample for some of the input and output variables. Furthermore, to comply with the activation function's range, all input and output variables are normalized using the following equation:

$$\chi_{nor} = \frac{\chi - \chi_{\min}}{\chi_{\max} - \chi_{\min}}$$

Where χ_{nor} normalized value, χ_{min} original lower limit, χ_{max} original upper limit and χ original value

2. The use of a network with a single hidden layer. The optimal number of hidden nodes was determined using a constructive approach (starting with 4 nodes and testing for a larger number of nodes). A maximum of 25 nodes was selected.

3. The use of the back-propagation algorithm for adjusting and optimizing connection weights and biases. The following tuning was deployed: learning rate=0.9, momentum term=0.6.

4. Calculation of mean absolute error (MAE), root mean square error (RMSE) and correlation coefficient (R^2) for all three data sets (train, test and validation). The mathematical expressions of R^2, MAE, and RMSE are defined in Sahoo and Ray (2006). In brief, the ANN predictions are optimum if R^2, MAE, and RMSE are found to be close to 1, 0, and 0, respectively. Error measurements to the testing set were used to assess network performance and to choose the optimal network. Table (3) summarizes some of the results obtained.

5. In the multiple artificial neural network approach 30 networks were trained on the same training dataset with random initial conditions of weights and bias with the goal that different networks will converge differently on the error surface. Out of those 30 networks 15 networks were selected, which gave minimum bias and variance on the validation set. The networks outputs are combined by ensemble averaging (Naftaly et al., 1997) and optimum linear combination (OLC) method (Hashem, 1997). The main aim of using multiple artificial neural networks is to obtain a better TOC prediction by a combination of networks instead of finding a single network by a trial and error approach, the architecture of a committee is shown in Figure (5).

In order to assess the validity of the multiple artificial neural technique, a comparison of the predicted TOC (ANNs) with the measured TOC (core data) was made using the ANN validation set, moreover, TOC is determined using the DlogR technique (Passey et al., 1990), and compared with the predicted and actual TOC.

The DlogR method proposed by Passey et al. (1990) combines the model-driven and data-driven approaches. The DlogR method overlays an appropriately scaled porosity log (sonic, density or neutron porosity) on a resistivity log (preferably from a deep-reading tool). At a certain scale, intervals where porosity log and resistivity log parallel or directly overlay each other are water saturated non-source rocks. In hydrocarbon reservoir rocks or source rocks, a separation between the two logs occurs. When the two logs are base lined, this separation (termed DlogR) can be quantified and TOC can be calculated with a non-linear relationship among TOC, DlogR and maturity.

7. Case study.

The study area is located at the eastern part of the North Western Desert of Egypt, it lies in the northern unstable shelf of the Northern Western Desert, and is characterized by deep and rugged basement rocks, thick sedimentary succession of highly-printed complex structural effects represented by asymmetric linear folding, faulting and

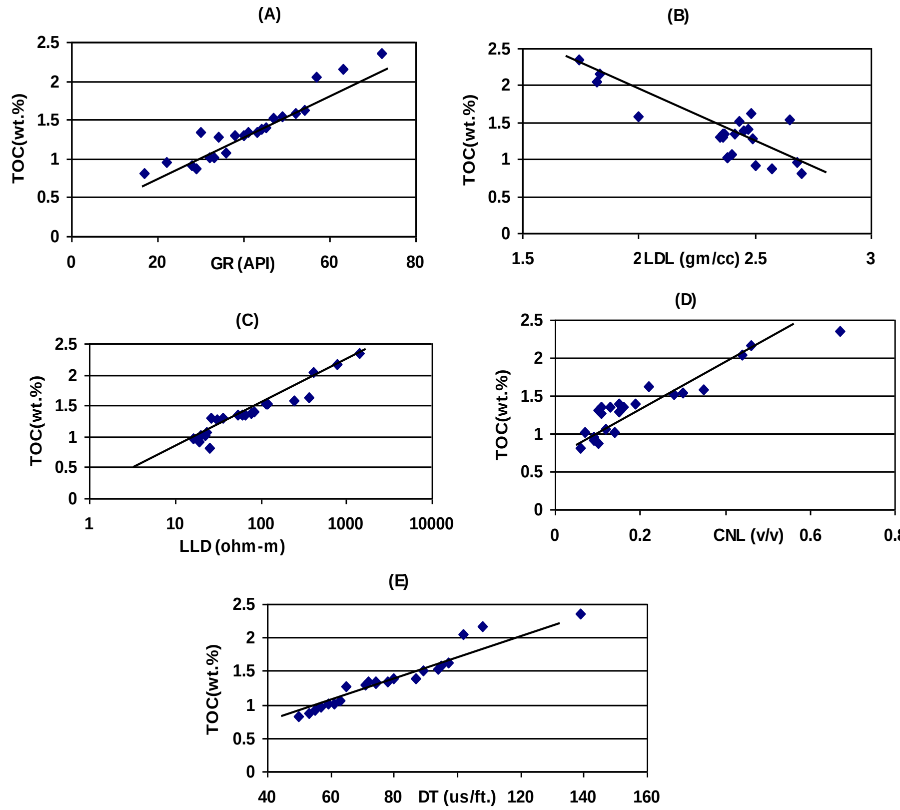

Fig. (4): Physical relationships between TOC and input petrophysical data

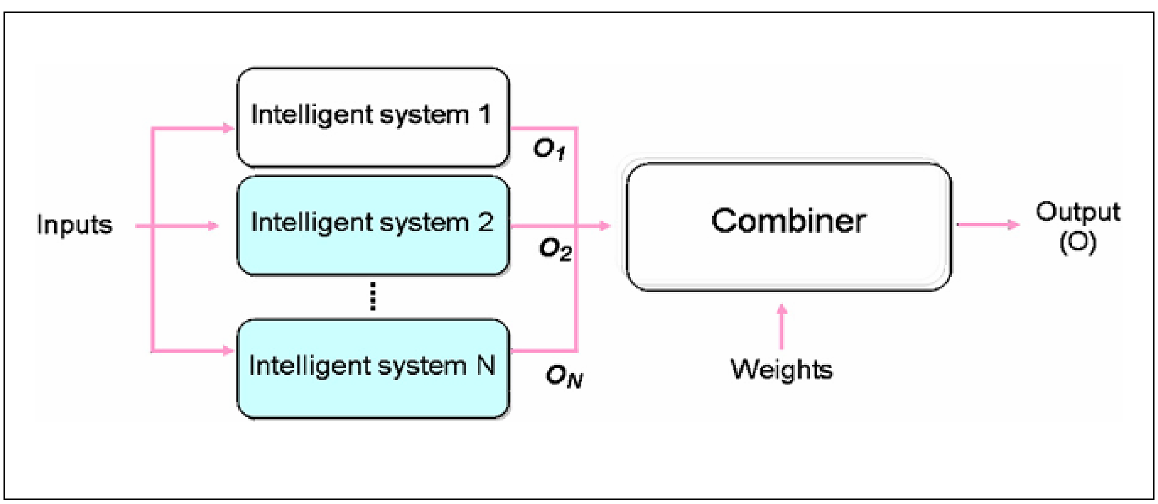

Fig. (5) A schematic diagram of CMIS

Table 2
Data distribution table showing three statistically consistent subsets (training, testing and validation data sets).

Parameters input/output	Training set	Test set	Validation set
DT			
Average	77	78	75
Median	75	79	76
Standard deviation	17	16	18
CNL			
Average	0.21	0.24	0.21
Median	0.15	0.13	0.16
Standard deviation	0.15	0.13	0.16
LLD			
Average	112	115	110
Median	7	5	5
Standard deviation	9	8	8
LDL			
Average	2.51	2.43	2.51
Median	2.57	2.48	2.59
Standard deviation	0.23	0.23	0.22
GR			
Median	62	60	58
Standard Average	57	61	59
Deviation	16	8	18
TOC			
Average	1.43	1.58	1.6
Median	1.61	1.57	1.58
Standard deviation	0.52	0.48	0.54

Table 3
A representative sample of artificial neural network models investigated for predicting TOC (wt.%).

Network No.	Hidden layer	Performance measurements: R2, RMSE and MAE					
		R2		RMSE		MAE	
	No. Neurons.	Train	Test	Train	Test	Train	Test
1	6	0.69	0.880	0.35	0.149	0.003	0.003
2	8	0.72	0.938	0.35	0.107	0.006	0.002
3	9	0.75	0.932	0.34	0.112	0.003	0.007
4	11	0.77	0.987	0.33	0.048	0.007	0.002
5	12	0.77	0.989	0.32	0.044	0.008	0.001
6	14	0.81	0.986	0.30	0.05 0	0.012	0.006
7	16	0.84	0.979	0.27	0.061	0.014	0.001
8	17	0.84	0.990	0.27	0.043	0.013	0.002
9	19	0.92	0.990	0.32	0.042	0.014	0.002
10	20	0.91	0.991	0.30	0.04 1	0.055	0.001
11	21	0.92	0.974	0.18	0.071	0.004	0.01
12	22	0.93	0.992	0.25	0.037	0.015	0.0006
13	23	0.93	0.991	0.21	0.040	0.018	0.0009
14	24	0.92	0.987	0.19	0.049	0.070	0.006
15	25	0.89	0.991	0.20	0.041	0.025	0.0004

diaprism resulted from considerable and varying tectonic disturbances, Abd El-Aziz et al. (1998). The Upper Cretaceous reservoir rocks (Bahariya and Abu Roash formations) are heterogeneous in composition. Bahariya Formation consists of sandstone with shale, limestone and dolomite. Abu Roash Formation, on the other hand, is subdivided into seven members designated from bottom to top as G, F, E, D, C, B and A. The rocks forming these members are mainly composed of calcareous rocks with argillaceous intercalations in the A, B, D and F members and argillaceous rocks are dominant with calcareous and arenaceous interbeds in the members C, E and G. The thickness of Bahariya Formation in the study area ranges from 686 ft. at N.B.Q-1X well to 1291 ft at N.B.Q-2X well, while the thickness of Abu Roash Formation varies from 1975 ft. at Gindi deep-1X well to 3795 ft at East Qarun well.

8. Results and discussion.

In this study, depth, sonic, resistivity, neutron, density and gamma-ray log responses are used as inputs to the neural networks and TOC values are the outputs. A set of data was chosen randomly (using a random number generator) and put aside. The data were not shown to the network during the training period. This data, called the test set, is used to check the integrity and robustness of the network after it has reached some stable state. Using this data, neural networks try to discover possible pattern that might exist between inputs and the corresponding output. After learning the data and recognizing the possible noises that might exist in the data, the network will eventually converge. It should be noted that one may encounter many different problems throughout this process. But a good understanding of the fundamentals of the problem, the tool, and their interaction can be of great help in overcoming these difficulties. In order to assess the validity of the ANN model developed, a comparison of TOC values determined by DlogR technique (Passey et al., 1990) was made with the ANN predicted TOC and the actual TOC values, Table (4) and Figure (6).

In comparing results obtained from DlogR technique with those obtained from the multiple artificial neural system, indications are that the latter outperforms the first.

Results obtained from all trials are presented in tables (3). Moreover, Figures (7), (8) and (9) illustrate the actual and the predicted TOC versus depth for three of the studied wells.

The technique introduced in this study is used to estimate TOC from well log data for East Qarun well, which have not been cored and its TOC's are not measured. Figure (10) represents the predicted TOC of East Qarun well using the preserved weights of the artificial neural network in the present work.

Table 4

Comparison between ANNs predicted TOC and DlogR TOC.

Error measurements	ANN predicted TOC	DlogR method
R2	0.95	0.16
RMSE	0.17	1.64
MAE	0.005	0.76

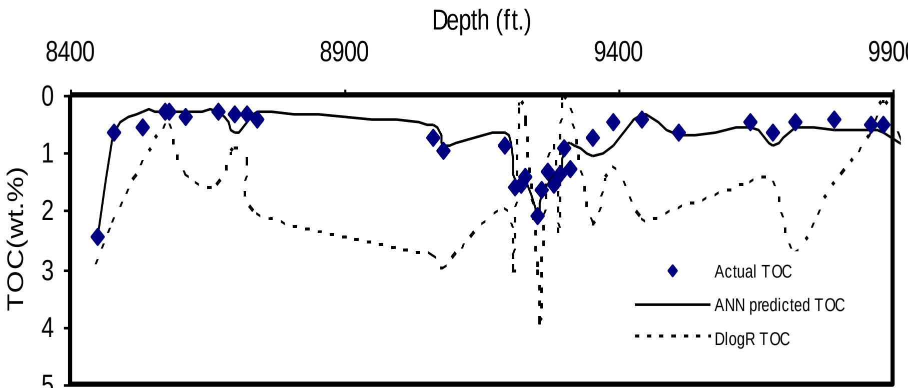

Fig.(6). Comparison between actual TOC, DlogR TOC and ANN predicted TOC of N.B.Q.-2X Well.

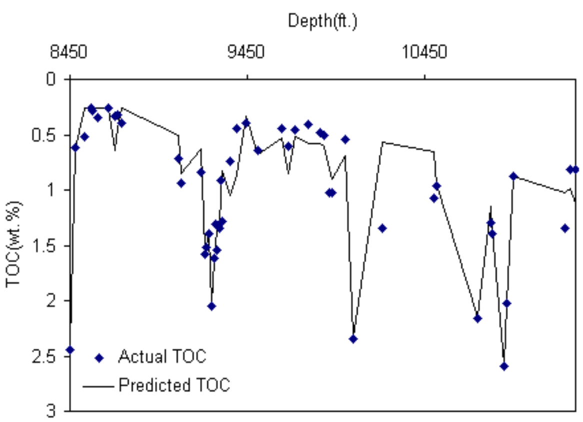

Fig.(7). Predicted TOC (wt.%) of N.B.Q.-2X Well.

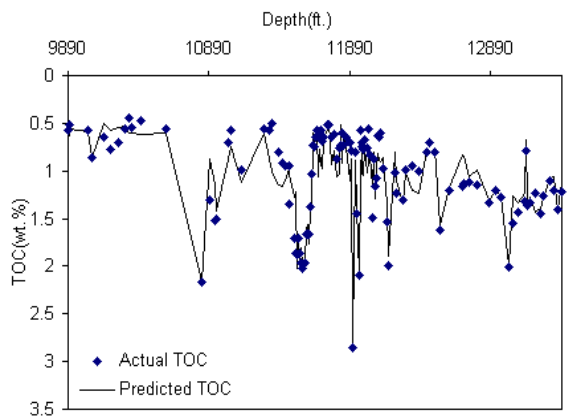

Fig.(8).Predicted TOC(wt.%) of N.B.Q.-1X Well.

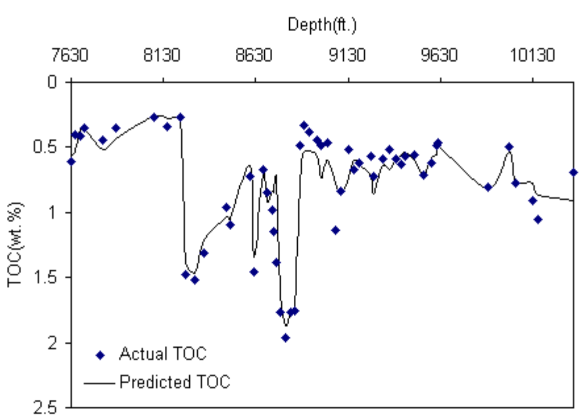

Fig.(9). Predicted TOC (wt.%) of Gindi Deep-1X Well.

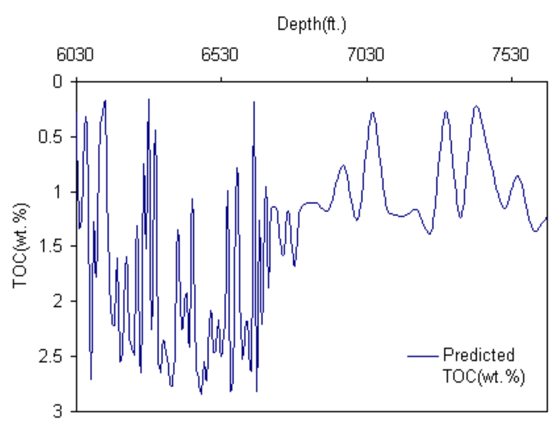

Fig.(10). Predicted TOC(wt.%) of East Qarun Well.

Conclusions

Multiple artificial neural network technique was used for the estimation of TOC from petrophysical data in eastern part of the North Western Desert of Egypt .Regarding the results developed in this research, the following points are concluded:

(a) ANNs has been successful for making a quantitative correlation between TOC and petrophysical data. The minimum RMSE for estimation of TOC in the test data is 0.037, which correspond to the R^2 value of 0.992.

(b) In this study, the number of measured TOC data was limited. So, there were not sufficient data for training. This problem associated with rock heterogeneities such as changes in mineralogy, fluid content and saturation could lead to unusual responses of the ANN at extreme value (over-estimation or under-estimation).

(e) Due to high costs of Rock-Eval pyrolysis, a limited number of samples were used in this study. However, ANN predictions for TOC were satisfying. So, it could be concluded that when there is a logical relationship between input and output data (such as those mentioned for TOC and petrophysical data), intelligent systems could recognize the patterns even with limited data.

(f) The multiple artificial neural network system introduced in this study is able to estimate TOC from well log data for other wells in the studied area, which have not been cored or their TOC's are not measured.

(g) Artificial neural network has a simple and easy structure and when there are multiple ways to solve a problem, it could provide smaller errors than the average of all experts by combining the outputs of each method.

References.

Abd El-Aziz, M., Moustafa, A. R. and Said, S., 1998. Impact of basin inversion on hydrocarbon habitat in the Qarun concession, Western Desert, Egypt; 14th EGPC Expl. & Prod. Conf.; V.1, P. 139-155.

Kadkhodaie Ilkhchi, A., Hossain Rahimpour-Bonab, and Rezaee, M.R., 2008. A committee machine with intelligent systems for estimation of total organic carbon content from petrophysical data: An example from Kangan and Dalan reservoirs in South Pars Gas Field, Iran Journal of Computers and Geosciences, doi:10.1016/j.cageo.2007.12.007

Autric, A. and Dusmesnil, P. (1985). Resistivity, radioactivity and sonic transit time logs to evaluate the organic content of low permeability rocks Log Anal. 26, 36-45

Beers, R.F., 1945. Radioactivity and organic content of some Paleozoic shales. AAPG Bulletin 26, 1 – 22.

Dellenbach, J., Espitalie, J., Lebreton, F., 1983. Source Rock Logging: Transactions of 8th European SPWLA Symposium, paper D.

Helle H.B. and Bhatt A. 2001. Fluid saturation from well logs using committee neural networks. Petroleum Geoscience 8 May 2002. (Expanded abstract presented in EAGE 64th conference and technical exhibitions-Florence, Italy, 2002).

Fertle, H., 1988. Total organic carbon content determined from well logs. SPE Formation Evaluation 15612, 407– 419.

Hashem S. 1997. Optimal Linear Combination of Neural Networks. Neural Networks 10, 599-614.
Herron, S.L., 1988. Source rock evaluation using geochemical information from wireline logs and cores (abs). AAPG Bulletin 72, 1007.
Hertzog, R., Colson, L., Seeman, B., O'Brian, M., Scott, H., et al., 1989. Geochemical logging with spectrometry tools. SPE Formation Evaluation 4, 153– 162.
Huang, Z., Williamson, M.A., 1996. Artificial neural network modeling as an aid to source rock characterization. Marine and Petroleum Geology 13, 227–290.
Hunt J.M., Jaieson, G.W., 1956. Oil and organic matter in source rock of petroleum. AAPG Bulletin 40, 477– 488.
Hunt, J.M., 1996. Petroleum Geochemistry and Geology, 2nd ed. W.H. Freeman and Company, New York.
Hussain, F.A., 1987. Source rock identification in the state of Kuwait using wireline logs. SPE 15747, 477– 488.
Langford, F.F., 1990. Interpreting Rock-Eval pyrolysis data using graphs of pyrolizable hydrocarbon vs. total organic carbon. AAPG Bulletin 74, 799– 804.
Kamali, M.R., Mirshady, A.A., 2004. Total organic carbon content determined from well logs using Dlog R and neuro fuzzy techniques. Journal of Petroleum Science and Engineering 45, 141–148.
Mendelson, J.D., Toksoz, M.N., 1985. Source rock characterization using multivariate analysis of log data. Transactions of the Twenty-Sixth SPWLA Annual Logging Symposium, paper UU.
Meyer, B.L., Nederlof, M.H., 1984. Identification of source rocks on wireline logs by density/resistivity and sonic transit time/resistivity cross plots. AAPG Bulletin 68, 121–129.
Naftaly, U., Intrator, N., Horn, D., 1997. Optimal ensemble averaging of neural networks. Computation in Neural Systems 8, 283–296.
Passey, O.R., Moretti, F.U., Stroud, J.D., 1990. A practical modal for organic richness from porosity and resistivity logs. AAPGBulletin 74, 1777–1794.
Rumelhart, D. E., Hinton, G. E. and Williams, R. J. (1986) Learning internal representations by error propagation. In: Parallel Distributed Processing, Vol. 1: Foundations (Eds D.E. Rumelhart and J.L. McClelland), MIT Press, Cambridge, Massachusetts, 318-362.
Sahoo, G.B., Ray, C., 2006. Flow forecasting for a Hawaiian stream using rating curves and neural networks. J. Hydrol. 317, 63–80.
Saxena, A., Saad, A., 2006. Evolving an artificial neural network classifier for condition monitoring of rotating mechanical systems J. Appl. Soft Comput. 1568–4946.
Schmoker, J.W., 1981. Determination of organic-matter content of Appalachian Devonian shales from gamma-ray logs. AAPG Bulletin 65, 2165–2174.
Schmoker, J.W., Hester, T.C., 1983. Organic carbon in Bakken Formation, United States portion of Williston Basin. AAPG Bulletin 67, 2165–2174.
Swanson, B.F., 1980. Rationalizing the influence of crude wetting on reservoir fluid flow with electrical resistivity behaviour. J. Pet. Technol., 1459–1464.
Tissot, B.P., Welte, D.H., 1984. Petroleum Formation and Occurrence. Springer-Verlag, New York. 966 pp.