

# Synthesis of missing openhole well log data through artificial neural networks

Amna M. Handhal

Department of Geology, College of Science, University of Basra, Basra, Iraq

Email: [amhgeo@gmail.com](mailto:amhgeo@gmail.com)

## Abstract:

A methodology is presented for deducing missing intervals of well logs data through applying artificial neural networks (ANNs) models. Three ANNs were performed for synthesizing sonic, neutron, and density logs. An example from Mishrif Formation of Nasryia oil field in southern Iraq was used to reveal the capability of ANNs model to synthesis missing intervals for these logs. Basically, ANNs models developed in this study were based on commonly multilayer perceptron and trained with backpropagation algorithm. Two statistical errors, namely, root mean squared error and correlation of determination were employed to assess the accuracy of the ANN models. Results indicated the capability of ANNs model to recreation of missing well interval with high accuracy.

**Keywords:** artificial neural networks, backpropagation algorithm, well logs, Iraq.

## تصنيع المجسات المفقودة للآبار المفتوحة باستخدام تقنية الشبكات العصبية الصناعية

آمنة مال الله حنظل الكنعاني

قسم علم الارض، كلية العلوم، جامعة البصرة، البصرة، العراق

### الخلاصة:

يتطرق البحث الحالي الى استخدام تقنية الشبكات العصبية الصناعية لتوليد قيم منحنيات مجسات الآبار المفقودة لثلاثة مجسات شائعة هي المجس الصوتي ومجس الكثافة ومجس النيوترون. اختيرت معمارية الشبكة العصبية من النوع المتعدد الطبقات واختبارها باستخدام طريقة انسياب الخطأ الى الخلف. وتقييم ادائية الشبكات العصبية المصممة باستخدام معيارين احصائيين هما مربع معدل الخطأ ومعامل الارتباط. اختيرت الطريقة على احدى آبار حقل الناصرية في جنوب العراق فبينت النتائج الامكانية الكبيرة للشبكات العصبية لتوليد قيم المجسات المفترضة. وخرجت الدراسة بنتيجة مفادها امكانية استخدام مثل هذه التقنيات في مجال جيولوجيا النفط في العراق بسبب مرونتها العالية وسهولة تصميمها واختبارها.

**الكلمات المفتاحية:** الشبكات العصبية الصناعية، خوارزمية الانسياب للخلف، المجسات البئرية، العراق.

## 1. Introduction

Reliable evaluation of petroleum reservoir requires reasonable knowledge of certain fundamental reservoir properties. Although the coring process is the main

approach for acquiring data, information extracted by well log are mostly common used for determining the physical properties of the rock and pore fluid. They are convenient and economic if they are

compared with the very expensive coring process. Unfortunately, it is common for several wells in an oil field to have incomplete suits of well logs because a full set of logs was not carried out at the time when the well was logged, problems encountered in repeat logging, and inappropriate or bed conditions during the logging process (Saggaf and Nebrija, 2003). Therefore, generating of synthesis well logs for those well intervals that are absent of incomplete is crucial task.

Few authors tried to deduced missing well logs using data mining techniques, for example artificial neural networks and fuzzy logic models. Saggaf and Nebrija (2003) used regularized back-propagation neural networks to estimate the missing log values in wells with limited log suits. They concluded this method produces an accurate estimate that are significantly close to the measured values and that method could be an effective tool of enhancing the limited suits of well logs. Rolon et al. (2005) used ANN technique to re-creation of missing logs from data basically acquired from conventional well logs. Their main conclusion revealed the high accuracy of developed ANN models to generate missing well logs. On the other hand, Constantin (2007) proposed a methodology using ANN to synthesis missing sonic logs by studying the relationship between transient time (DT) and data of deep resistivity ( $R_t$ ) and gamma ray (GR) logs. His study revealed the effectiveness of using ANN to generate DT log for those well having only the input curves of gamma ray and deep resistivity logs. Razaee et al. (2007) used ANN and fuzzy logic models to synthesis porosity logs (sonic, density, and neutron) and deep resistivity log. Their results revealed the high ability of these techniques to deduce missing intervals. They also build a simple

average committee machine model through sum the output from single model and after assign appropriate weights for each individual model and the result demonstrated a significant improvement in the model accuracy.

The main goal of this contribution is to use ANNs black box model to synthesize missing well logs and to apply this technique in the oil industry of Iraq in order to reduce the costs of exploration and management of oil filed across the whole country with fast and very cheap constructed models.

## 2. Artificial Neural Networks (ANNs)

ANN is a computing system able to acquire, represent, and compute a mapping from one multivariate space of information to another, given a set of data representing that mapping (Garrett, 1994). The computing system is made up of a number of simple highly interconnected processing elements called node, that passes information by their dynamic state as a response to appropriate inputs. Because of this character, ANN can be used to approximate non-linear relationships giving sufficient data input and output with acceptable accuracy. Hence, ANN could be regarded as a universal approximators and this lead to use it in different fields of science and engineering with promising and very successful outcomes. Basically, many different types of ANN model were developed during the last few decades. The most commonly used was the Multilayer Perceptron (MLP) network (Lippman, 1987). It is a feed-forward network (information passing in one direction) and typically trained using Backpropagation (BR) algorithm. Theoretically, this network is capable to approximate any function with a reasonable accuracy

(Hornik et al., 1989). Topologically, the network consists of layers of parallel processing called neurons that fully connected to each other by interconnect strength or weights,  $w$  (Fig. 1). The neurons in the first layer  $i$  (input layer) process and passes information after weighed them to the next layer through links. The connection between  $i^{\text{th}}$  and  $j^{\text{th}}$  neurons is defined by the weight coefficient  $w_{ij}$  and the  $i^{\text{th}}$  neuron via the threshold coefficient  $\mathcal{G}_i$ . The final output is calculated according to Haykin (1998):

$$x_i = f(\xi_i) \quad (1)$$

$$\text{With } \xi_i = \mathcal{G}_i + \sum_{j \in \Gamma_i^{-1}} w_{ij} x_j \quad (2)$$

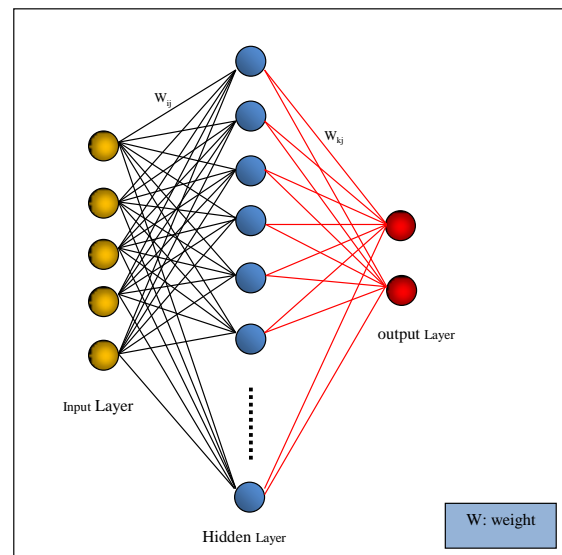
in which,  $f(\xi_i)$  is the activation function.

The BR algorithm update weights in the connections between layers by computing the error between the output of the network and target variable and propagating this backward through the network as:

$$\Delta w_{ij}(t) = -\eta \frac{\partial E}{\partial w_{ij}} + \mu \Delta w_{ij}(t-1) \quad (3)$$

where,  $\eta$  and  $\mu$  are the learning and momentum rates, respectively.

A more detail description about BR is found in Fausett (1994) and Haykin (1998).



**Fig.1:** A simple Multilayered perceptron network.

### 3. The data used

The data used in this study are collected from archive of South Oil Company (SOC) at Basra Governorate – Iraq in Mishrif Formation in Nasyria Oil Field. The main reservoir in this field is Mishrif Formation. Bellen et al. (1957) described this formation as a heterogeneous formation consists mainly of detrital limestones with beds of algal, rudist and coral-reef limestones. The average porosity and permeability of this formation are 22% and 23 md, respectively. For the purpose of this study, five exploration wells remarked as A, B, C, D and E were used. The available well logs for these five wells are Gamma ray (GR), sonic log (DT), density log (RHOB), neutron log (NPHI) and deep resistivity log ( $R_t$ ). Logs of NPHI, DT, and RHOB were choose here for testing the methodology and similar method could be used for generation other logs accordingly. The A, B, C and D wells were used for training ANN models, while well E was used for testing phase. Table (1) shows the basic

statistic for the used logs for both training and testing phases.

**Table 1:** The basic statistic for the training and testing phases

Statistic of training data set (A, B, C, and D wells)			
Logs	Min.	Max.	Mean
RHOB (gm/cm <sup>3</sup> )	1.81	2.77	2.36
NPHI (%)	0.02	0.59	0.25
DT (μs/ft)	52.00	150.00	76.48
Statistic of testing data set (E well)			
Logs	Min.	Max.	Mean
RHOB (gm/cm <sup>3</sup> )	2.05	2.57	2.34
NPHI (%)	0.13	0.47	0.24
DT (μs/ft)	57.10	130.00	79.57

#### 4. Methodology

Although the well logs for five wells used in this study are available, we assume that logs data of (E) well are not completed available and try to use the ANNs model on the existing data of other logs and available well logs of (E) for predicting the assumed missing log of (E). The essential methodology for developing neural network models for synthesizing missing well logs is to train constructed models on the available borehole data. If the well logs data covers a reasonable part of the area and the data obtained are sufficient, the trained networks would have enough information to recognize and predict interpolate missing logs at locations close to the drilling area. Table 2 demonstrates the appropriate inputs for synthesizing assumed missing well logs in well (E).

**Table 2:** Inputs for synthesizing assumed missing well logs

Predicted logs	Inputs
DT	NPHI & RHOB
NPHI	DT & RHOB
RHOB	DT & NPHI

To implement ANN models, a neural network toolbox in MATLAB R2015a was used in this study. To synthesis sonic, density, and neutron logs, three different ANN models were generated and tested. The raw data (well logs) were preprocessing and scaled into range [-1, 1] using Mapminmax function. After that, the scaled data were partitioned into three sets with 60/20/20 ratio; 60% of data for training, 20% for validation, and 20% for testing. The training data is primary utilized to extract information about the problem being solved, while validation set is used to ensure that there is no overfitting problem. The overfitting problem means that the error on the training set is seem to be very small but when new (unseen data) is presented to the network the error becomes large (Matlab use guide, 2008). An early stopping procedure was used here to avoid overfitting and improve the developed ANN models. The last data, ie test data was used to provide an independent measure to test how well the training ANN model perform on data that never used to build it. The developed ANNs model in this study were trained using Levenberg-Marquardt technique as this technique is powerful and faster than the conventional gradient descent method (Hagan and Menhaj, 1994). The prediction accuracy of the ANN models was evaluated by using two statistical errors, namely, root mean squared error (RMSE) and correlation of determination ( $R^2$ ). The RMSE measures the residual variance and thus quantify the model error. The optimal

value of this statistics is 0, and calculated as:

$$RMSE = \sqrt{\frac{1}{2} \sum_{i=1}^n (x_i - \hat{x})^2} \tag{4}$$

where,  $x$  and  $\hat{x}$  are the observed and predicted values, respectively,  $n$  is the total number of observations.

In contrast,  $R^2$  measures the goodness of fit of the model; the optimal value is 1. It is defined as:

$$R^2 = 1 - \frac{SSE}{SSy} \tag{5}$$

where

$$SSE = \sum_{i=1}^n (x_i - \hat{x})^2 \tag{5}$$

$$SSy = \sum_{i=1}^n (x_i - \bar{x})^2 \tag{6}$$

$\bar{x}$  = Mean of the measured values

To improve and optimize the architecture and performance of the ANNs models, the number of hidden nodes and learning rate were changed many times (trial and error method) to minimize error between measured and predicted values depending on the error statistic measures RMSE and  $R^2$ , Table 3 and Fig.3.

**Table 3:** The best architecture and performance of ANNs designed for synthesizing well logs

Predicted logs	No. of hidden nodes	Learning rate $\mu$
DT	45	0.001
NPHI	35	0.005
RHOB	40	0.001

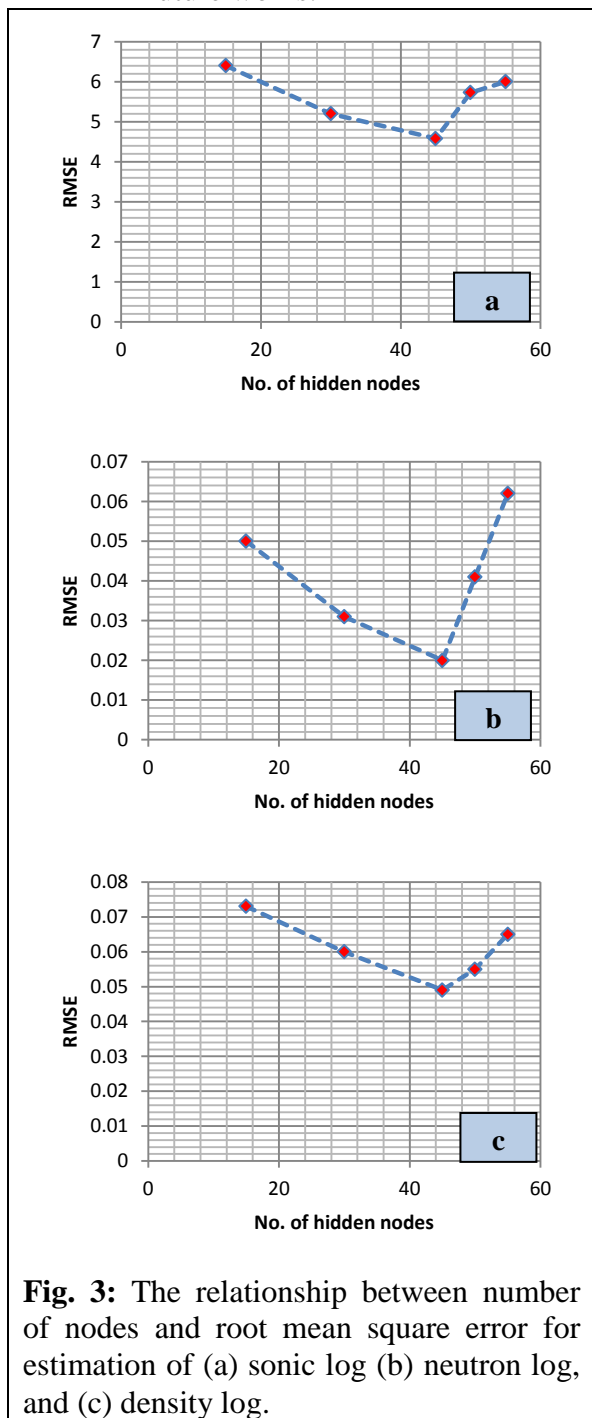
### 5. Results and Discussion

Fig. 4 (a-c) show the correlation between measured and predicted values for sonic, neutron, and density logs of testing well E (Ns-5), respectively. The comparison between measured and predicted sonic, neutron and density curves for testing well too are shown in Fig. 5 (a-c). Table 4 demonstrates the error statistics for the experiments and from figures 4 and 5 and table 4 it can be seen that the ANNs technique were capable for generating missing wells logs successfully. The error statistic strongly supports this statement. There is overestimation of some values of sonic log, Fig. 5 in the upper part of the curve; this may be, because only a small number of training patterns of low values are available. The other curves (RHOB and NPHI) exhibit underestimation and overestimation too. This may be attributed to the several factors such as reservoir complexity, changes in mineral composition, changes in fluid content and inherent errors in model or in test data (Razeea et al., 2007). Generally, the predicted values for all cases have the same trend of measured one in low and high values.

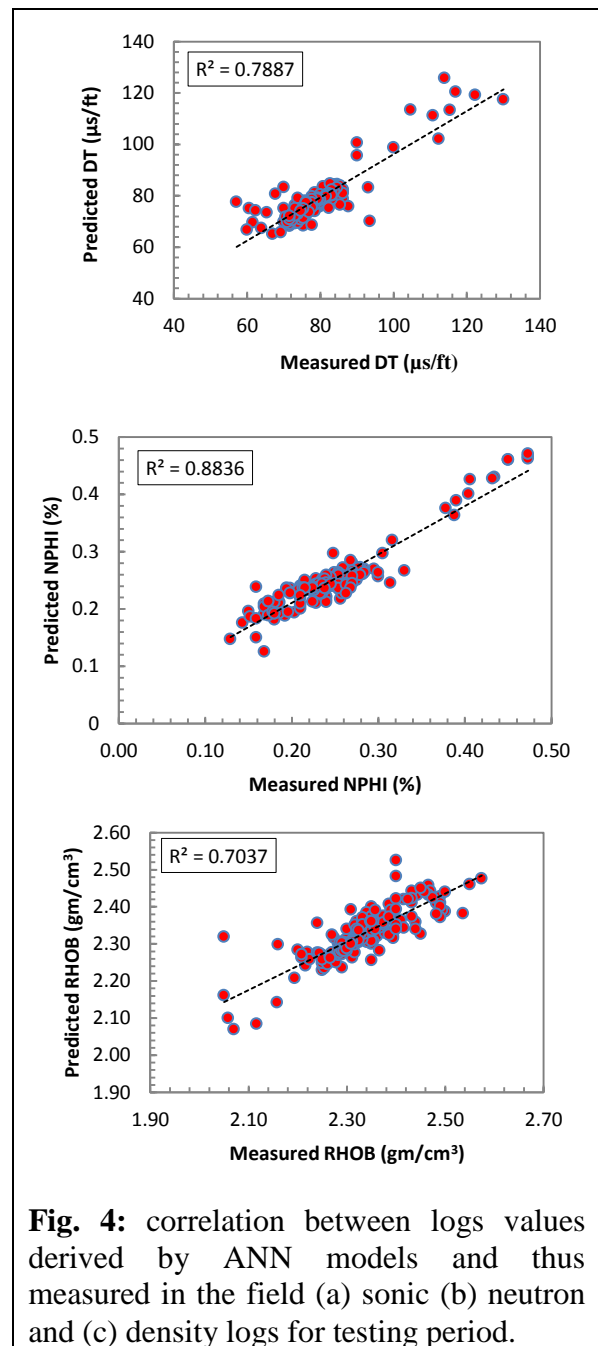
### 6. Conclusions and recommendation

1. Multilayer Perceptron artificial neural networks are a very efficient technology for generating missing well logs data. It is robust and inexpensive, and requires only sufficient data for training and memorizes the pattern.
2. Using of such technique in oil industry of Iraq could reduce the costs of management of oil field and enhance the opportunity of exploration and interpretation.
3. Other artificial intelligence techniques such as fuzzy inference system and genetic

programming may offer a similar flexible means for solving problems faced in oil industry and recommended to use in future works.



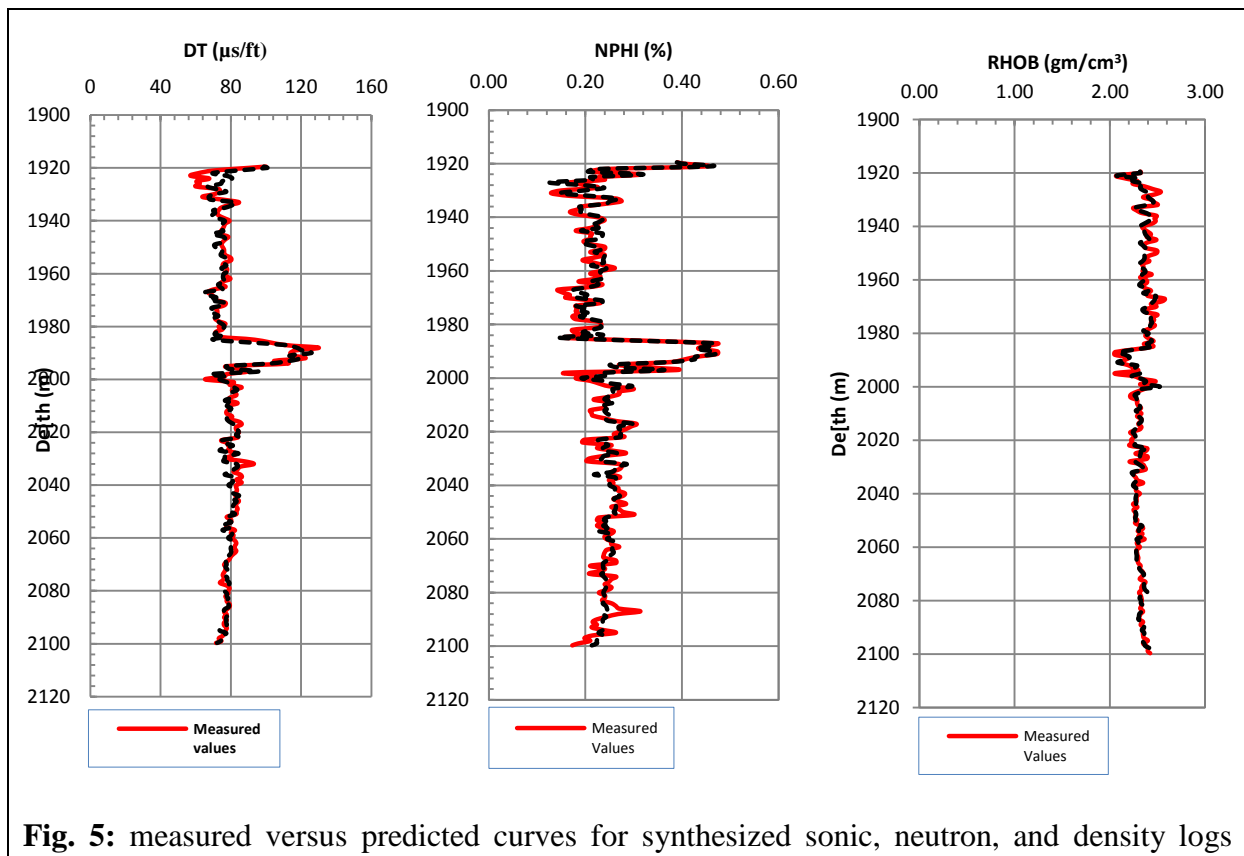
**Fig. 3:** The relationship between number of nodes and root mean square error for estimation of (a) sonic log (b) neutron log, and (c) density log.



**Fig. 4:** correlation between logs values derived by ANN models and thus measured in the field (a) sonic (b) neutron and (c) density logs for testing period.

**Table 4:** Evaluation of experiments results using error statistics (testing data set)

Predicted log	Error statistics	
	RMSE	R <sup>2</sup>
Sonic (μs/ft)	4.580	0.788
Neutron (%)	0.022	0.883
Density (gm/cm <sup>3</sup> )	0.049	0.703



**Fig. 5:** measured versus predicted curves for synthesized sonic, neutron, and density logs

#### 4. Reference

- [1] Bellen, R. C. Van, Dunnigton, H. V., Wetzel, R. and Morton, D., 1959. Lexique Stratigraphique International Asie. Iraq. Intern. Geol. Conger. Comm. Stratigr, 3, Fasc. 10a, 333p.
- [2] Constantin, C. (2007) Using artificial neural networks to predict the presence of overpressured zones in the Andarko basin, Oklahoma. Pure Appl. Geophys., DOI 10.10671 s00024-007-0257-9.
- [3] Fausett L. (1994) Fundamentals of Neural Network, Architectures, Algorithms, and Applications. A Simon and Schuster Company, New Jersey, USA, 461p.
- [4] Garrett, J. (1994) Where and why artificial neural networks are applicable in civil engineering. Journal of Computing in Civil Engineering, 8, pp. 129–130.
- [5] Hagan M. T., Monhaj, M. B. (1994) Training feedforward networks with the Marquardt algorithm. IEEE Transactions on Neural Network 6. 861-867.
- [6] Haykin, S. (1994) Neural Networks a Comprehensive Foundation, 1st edition, Macmillan College Publishing Company, Inc., New York.
- [7] Hornik, K., Stinchcombe, M., and White, H. (1989) Multi-layer feedforward networks are universal approximators. Neural Networks, 2, p. 359-366.
- [8] Lippmann, R. P. (1987) An introduction to computing with neural nets: IEEE ASSP Magazine, p. 4-22.
- [9] Matlab user's guide, 2008 Fuzzy logic & Neural Networks Toolboxes Matlab CD-Room, by the math works, Inc.
- [10] Rezaee, R. M., Kadkhaodaie-Ilkhchi, A., & Alizadeh M. P. (2007) Intelligent approaches for the

synthesis of petrophysical logs. J. Geophys. Eng. 4, p. 1-15.

- [11] Rolon, L. F., Chevron, Mohaghegh, S.D., Ameri, S., Gaskari, R. (2005) Developing Synthetic Well Logs for the Upper Devonian Units in Southern Pennsylvania. SPE 98013. SPE Eastern Regional Meeting held in Morgantown, W.V.
- [12] Saggaf, M. M. & Nebrija, Ed. L. (2003) A fuzzy logic approach for the estimation of lithologies and depositional facies from wire-line logs. AAPG Bulletin, 87(7), p.1223-1240.