PREDICTING RESERVOIR QUALITY IN SANDSTONES THROUGH NEURAL MODELING

SANDRO DA SILVA CAMARGO[1], PAULO MARTINS ENGEL[2]

ABSTRACT

Due to limited understanding of the details of many diagenetic processes, mathematical models become a very useful tool to predict reservoir quality prior to drilling. Porosity prediction is an important component in pre-drill and post-drill evaluation of reservoir quality. In this context, we have developed a mathematical model to predict porosity of sandstones reservoir systems. This model is based on artificial neural networks techniques. We propose a score to quantify their importance of each feature in prediction process. This score allows creating progressive enhancement neural models, which are simpler and more accurate than conventional neural network models and multiple regression. The main contribution of this paper is the building of a reduced model just with the most relevant features to porosity prediction. A dataset about Uerê formation sandstone reservoir was investigated. This formation is an important oil exploration target in Solimões Basin, western Brazilian Amazonia. Study results show that progressive enhancement neural network is able to predict porosity with accuracy near 90%, suggesting that this technique is a valuable tool for reservoir quality prediction.

KEY-WORDS: Progressive Enhancement Neural Model. Sandstones Reservoir Quality. Porosity Prediction.

1. INTRODUCTION

World economy is petroleum-dependent. Nowadays, more than 85% of world energy consumption comes from fossil fuels [1]. However, fossil fuels are a limited resource, and reservoirs are consumed more rapidly each year. More accessible reservoirs, whose exploration is low cost, are supposed to be almost vanished. Besides, remaining reservoirs are increasingly more technically difficult to extract and, therefore, more expensive. Later reservoirs will only be economically feasible to extract at extremely high costs. In this context, exploration costs are a very important variable in reservoir exploration decision. In order to know the costs of field exploration, and increase exploration success rates, oil companies must try to predict the quality of reservoirs. Accurate prediction of reservoir

[1] Instituto de Informática, Universidade Federal do Rio Grande do Sul, Doutorando em Ciência da Computação, scamargo@inf.ufrgs.br .
[2] Instituto de Informática, Universidade Federal do Rio Grande do Sul, Doutor em Engenharia, engel@inf.ufrgs.br .

quality is, and will continue to be, a key challenge for hydrocarbon exploration and development [2].

Due to limited understanding of the details of many diagenetic processes, there is a lack for new techniques and tools to support quality predictions. Despite its notable economic importance, relatively few papers illustrate the application of studies to reservoir quality prediction. The main difficulty to execute this task is that the creation of models to reservoir quality prediction is highly dependent of quality and availability of calibration datasets [2,3]. Biased datasets will generate poor models, furthermore, lack of observations combined to a high amount of features describing each observation can become more difficult, or even prohibitive, to fit a multivariate model to forecast reservoir quality. This problem is known as curse of dimensionality [4].

Regression analysis is the most commonly used technique to predict reservoir quality [1,5,6]. However, this technique has limitations and demands intense interaction with domain experts. Moreover, such models are sensitive to the limits imposed by the calibration dataset. Recently, soft computing techniques have been used in reservoir characterization and modeling [7]. Among these techniques, Artificial Neural Networks (ANNs) have been used to identify relationships between permeability, measured logs and core data [8].

In this paper, we propose a mathematical model to predict porosity in sandstones reservoir systems. Our approach is based on a Progressive Enhancement Neural Model (PENM) to predict reservoir quality. The main contribution of our approach is to generate a reduced model with more relevant features to porosity prediction. The results show that our approach generates more accurate predictions than commonly used techniques, multiple regression analysis, and conventional ANNs. Another advantage is a lower dependency of domain expert than regression analysis.

Two general approaches have been used for predicting reservoir quality in sandstones: empirical and process-oriented techniques. Empirical techniques use multiple regression analysis. Process-oriented techniques use chemical and mathematical models to understand diagenetic processes and their effects on the evolution of reservoir [2]. Despite the uncertainty associated with simulator-based forecasts, reservoir simulation continues to be the most reliable method for making performance predictions, particularly for reservoirs that do not have an extensive history [3].

Regression analysis is the most commonly used technique to predict reservoir quality [1,5,6]. However, this technique has limitations and demands intense interaction with domain expert. Moreover, such models are sensitive to the limits imposed by the calibration dataset. Differences in depositional controls, depositional and sequence stratigraphic settings, and sequence stratigraphic concepts between sandstones and carbonates imply creating different models for predictions. Usually, if the dataset embraces data with two or more of

these differences, the dataset must be divided in multiple subsets, which must encompass similar observations. This task is domain expert dependent. To achieve more accurate predictions, a model for each subset must be created [6]. Consequently, models are likely to be basin-specific, and may even be restricted to particular facies or stratigraphic horizons, thus inherently limited in their application.

Recently, soft computing techniques have been used in reservoir characterization and modeling [7]. Among these techniques, ANNs have been used to identify relationships between permeability, measured logs and core data [8], to predict water saturation from log data [9], to predict asphaltene precipitation in crude oil [10] and to predict reservoir volume [11]. ANNs also have been used combined with other techniques, making part of hybrid models, like described in Chao et al. [12] to predict borehole stability. Similarly to our work, ANNs and multiple regression were compared when predicting trap quality [13].

The remaining of the paper is organized as follows: Section 2 presents basic concepts of ANNs. Section 3 presents our approach. Section 4 presents experiment results and the analysis of the approach. Section 5 concludes the paper and gives some directions of future work.

2. MATERIAL AND METHODS

To model this problem, we have selected ANNs technique. The study of ANNs started as an attempt to build mathematical models which worked in the same way that brains do. This study had lead to an abstract computer model of the human brain. This model uses a group of mathematical algorithms which are considered to implement the fundamental functional source of intelligence [2]. Nowadays, ANNs are a solid technique and became a powerful language for using large flexible nonlinear models [14].

Analogously to the brain, an ANN is composed of processing nodes, also called artificial neurons, which are interconnected by weighted edges, called synapss. Each neuron receives one or more inputs, multiplied by their weights, and sums these products to generate an output, which is adjusted by an activation function and sent to one or more neurons. Formally, a neuron have inputs x_1, x_2, …, x_m. Each input x_i is multiplied by its corresponding weight w_i. That is, the neuron evaluates $net = x_1w_1 + x_2w_2 + \ldots + x_mw_m$. Finally, the neuron computes its output y as an activation function of net, i.e., $y = f(net)$.

Based on structure of the connections among neurons, two different classes of network architecture are identified: non-layered recurrent and layered feed-forward, which is the scope of this work. Regarding to layered networks, neurons are organized forming layers. Neurons in a layer get input from the previous layer and feed their output to the next layer. The first layer is called input layer. Neurons at this layer just transmit their inputs to next

layer, with no computation. The last layer is called output layer. Typically, for regression problems, there is only a single neuron in this layer. For classification problems, there is a neuron for each category of the target variable. Between input and output layers, it can be one or more layers, called hidden layers.

Based on the amount of layers, two different classes of layered feed-forward neural networks are identified: single layer and multilayer. In single layer there is only one computing layer, the output layer. If there are one or more hidden layers, the network is called multilayer or MLP (Multi Layer Perceptron). Single layer networks can only learn linearly separable patterns. Otherwise, MLP can learn non-linear patterns. The universal approximation theorem for ANNs states that MLP can approximate any continuous function [15]. FIGURE 1 shows a typical MLP.

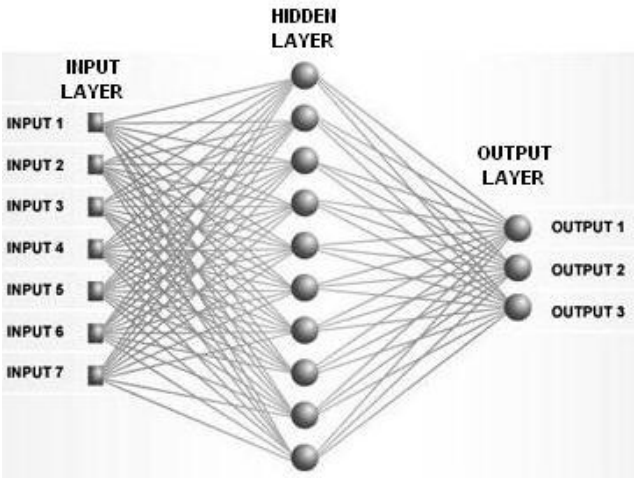


Figure 1 – A typical MLP containing 3-layers, 7 neurons in input layer, 10 neurons in hidden layer and 3 neurons in output layer is utilized.

In order to train an ANN to perform some task, we must adjust the weights of each synapse in such a way to create a model representing patterns expressed in data. This process is called learning. Although many learning algorithms for MLP have been proposed, backpropagation algorithm is the most widely used [16].

The main difficult in the use of MLP lies in design process. There are no rules to choose the best ANN configuration, in terms of layers and neurons in each layer, and the best training parameters. Therefore, designing an MLP demands some iterations in order to find the best MPL settings to a particular problem.

Although ANNs are a powerful technique they are affected by “curse of dimensionality” in two ways: 1) in high dimensional data the ANN can use almost all its resources to represent irrelevant portions of search space; 2) even if ANNs could focus on

important portions of search space, the higher the dimensionality of input space the more data may be needed to find out what is important and what is not. Moreover, according the Ockham's razor principle, the modeler should select the most simple model and grossest reservoir description that will allow the desired estimation of reservoir performance [3].

2.1 Proposed Approach

In order to overcome the "curse of dimensionality" problem, we propose a new approach called progressive enhancement neural model. Our approach includes creating regression models in which the choice of predictive features is carried out by an automatic procedure. To guide this automatic procedure, we propose a neural importance score, which is computed by:

$$s_i = \frac{\sum_{k=1}^{n} | w_{ik} |}{n}$$

where:

- s_i represents the score for *i*-th feature;
- n represents amount of neurons in the first hidden layer;
- w_{ik} represents the weight of synapse between *i*-th neuron of input layer and *k*-th neuron of first hidden layer;

The main objective of this score is to quantify the importance of each feature regarding target prediction. After MLP training, patterns expressed in data were learned and are represented in synaptic weights. So, the largest synaptic weights are supposed to be linking the most important input features to first hidden layer. If a synaptic weight tends to zero, its propagation to first hidden layer, and consequently to following layers, will tends to zero too, denoting its little importance regarding target prediction. While synapses among input and first hidden layer are supposed to encode patterns expressed in data, synapses among first hidden layer and output layer are supposed to decode patterns expressed in data to reconstruct this pattern. For this reason, our approach is concentrated in synaptic weights before first hidden layer, and ignores the others.

In order to provide data to compute the scores of features, a conventional MLP must contain knowledge about the dataset. This way, before computing the scores, MLP must be previously trained with original training dataset, in which each sample must be described by all available features. After this training, scores can be computed.

In order to achieve the progressive enhancement of initial neural model, an iterative process is performed. This process is performed as follows: After score computing, it is executed a forward selection, which involves starting with simplest model, containing just the

largest score feature. At each stage, the next feature available, which has the largest score among unselected features, is inserted in the model, and the model is evaluated. This process continues iteratively until the measure is locally maximized, or when the available improvement falls bellow some threshold. Prediction error is the metric used to evaluate models. Model evaluation is performed by some cross-validation technique.

The proposed approach is similar to Effroymson's algorithm, [17]. However, while Effroymson's approach is f-test and multiple regression-based, our approach is neural network-based.

The final model, generated by our approach, is supposed to be enhanced regarding the initial MLP. This enhancement is obtained through selection of a subset of available features. This subset must contain just the most important features regarding target prediction. Formally, the progressive enhancement process will choose a subset of M features from the original set of N features, where $M \leq N$. Furthermore, if the enhanced model contains just the most important features, it is expected to be more accurate than the original model, which contains the original features.

3. AVAILABLE DATA

For this study, we have used an existing dataset published by Lima and De Ros [18]. This database contains observations of petrographic and petrophysical parameters of Devonian sandstone reservoirs of the Uerê Formation, which is an important oil exploration target of the Solimões Basin. Further, exploration of the Uerê sandstones is complicated by the heterogeneous quality of these reservoirs, which range from highly porous to extremely tight. Moreover, although the Solimões Basin has been explored throughout the past three decades, little is known about the reservoir quality controls.

Another challenge is dataset dimensionality. This makes application of modeling techniques harder. The dataset contains 59 samples and 88 features, which describe petrographic and petrophysical parameters of sandstones.

3.1 Data Pre-Processing

Before model creation, the dataset must be prepared to improve modeling process. Pre-processing phase includes some tasks like cleaning and transformation.

During cleaning task, one sample was deleted. This sample was considered an outlier by domain expert. Features containing exactly the same value for all samples were deleted. A well known pattern involving prediction target was eliminated from input data. Macroporosity can be directly computed through a sum of intergranular, intragranular in

feldspar, intragranular in quartz grain, intragranular in mica, intragranular in heavy mineral, dissolution of pseudomatrix, dissolution of cement, mouldic, fracture and oversizes parameters. So, these features were deleted, in order to allow finding non-trivial relationships between macroporosity and other parameters. Some features holding the sum of other ones were deleted.

Regarding transformation task, input data were normalized by decimal scale technique [19]. Original data, which were in [0,100] range, were divided by 100. Subsequently, data were transformed to fall in [-1,1] range, in order to be used by the neural network.

Thus, after pre-processing, the dataset contained 58 samples and 60 features.

4. RESULTS AND DISCUSSION

The ANN was developed and tested on a Windows Based PC using MatLab software. Several MLP network structures were automatically tested and evaluated by cross-validation process. These structures had one or two hidden layers and from one to 60 neurons in each hidden layer. Different learning algorithms were tested too. After these several tests, the best performance ANN was adopted.

4.1 Adopted Neural Network

After testing and evaluating of many architectures and learning algorithms, it was selected an ANN with following characteristics:

Network Architecture:

- Multi Layer Perceptron;
- Activation strategy: Feedforward;
- Totally connected nodes;
- 3 layers;
- 60 neurons in input layer;
- 4 neurons in hidden layer;
- 1 neuron in output layer;

Information processing in nodes:

- Hidden layer(s) information processing: Sum is the input/propagation function. Hyperbolic tangent sigmoid activation function. BIAS node.

- Output layer information processing: Sum is the input/propagation function. Linear output function. BIAS node.

Learning algorithm:

- Form of learning algorithm: Gradient descent with momentum and adaptive learning backpropagation;
- Learning parameters: Learning rate: 0.01, ratio to increase learning rate: 1.05, ratio to decrease learning rate: 0.7, momentum constant: 0.0. Epochs:
- Dynamic strategy of learning parameter change: If the new error exceeds the old error by more than a predefined ratio (defined as 1.04), the new weights and biases are discarded. In addition, the learning rate is decreased (multiplied by 0.7). Otherwise, the new weights, etc., are kept. If the new error is less than the old error, the learning rate is increased (multiplied by 1.05).
- 20 initializations / retrainings;
- Synaptic weights were randomly generated and were normalized to a length of 1;
- Stopping method: minimum gradient reached or maximum epoch reached;
- Error measure: Mean Squared Error (MSE);

Model Selection

- MSE was the error measure used for evaluation;
- Selection of the best model was based upon lowest error on test set;

4.2 Model Evaluation

Unfortunately, prediction methods are susceptible to overfitting the learning examples at the cost of decreasing generalization accuracy over unseen examples. For small training sets this problem is severer. Due to the lack of samples to perform an early stopping approach, and prevent overfitting, models evaluation must be carefully planned. One of the most successful methods for evaluating performance accuracy is the leave-one-out cross-validation technique, in which the set of *m* training instances is repeatedly divided into a training set of size *m*-1 and test set of size 1, in all *m* possible ways. Model error is the sum of absolute errors for each one of *m* tests performed during leave-one-out cross-validation process.

4.3 Progressive Enhancement

After selecting the best ANN architecture and algorithms, and after its training, patterns found in data are expressed in ANN synaptic weights. FIGURE 2 shows the scores

from trained ANN. Few features, which are high scored, are identified as important and the most part of features are low scored. Low scored features are perceived as noise in conventional ANN. Higher scored features are supposed to be most important in target feature prediction.

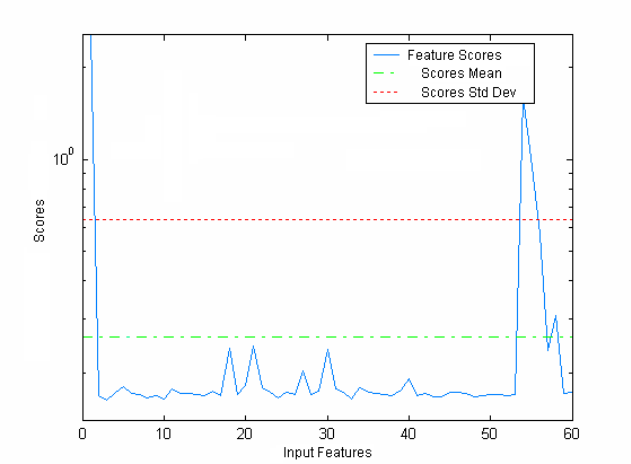


Figure 2 – Scores for each input feature

TABLE 1 shows a top 10 ranking containing the scores of each feature. These scores are the basis for PENM creation. Features, which their scores are greater than mean plus a standard deviation, were emphasized.

Table 1 – Top 10 feature scores for macroporosity prediction

Order	Feature	Score
1	Quartz Monocrystalline	2.5538
2	Intergranular Volume	1.5992
3	Cement Total	1.0115
4	Carbonate Total	0.5677
5	Grain Replacement Total	0.3078
6	Quartz Overgrowth	0.2446
7	Clay Ooid	0.2396
8	Silicified Secondary Matrix	0.2379
9	Silica Total	0.2375
10	Microquartz Rims	0.2030

Let *n* as the amount of features in original dataset, the iterative process performed during PENM creation can generate until *n* models. Table II shows evaluation results for each one of these models. Clearly, some models, which contain just a subset of all available

features, are more accurate than full model, which contains 60 features of original dataset. In this case, the model containing the 3 largest scored features is the enhanced neural model generated by our approach.

In order to evaluate the results obtained with proposed approach, other approaches were applied to this dataset: Multiple Regression, which is the most commonly used technique to predict reservoir quality [1,5,6], and conventional neural networks.

Table 2 – Macroporosity prediction errors

Amount of Input Features	Absolute Error	% Error
1	2.7173	26.17%
2	2.7110	26.11%
3	1.8637	17.97%
4	1.9578	18.86%
5	2.0533	19.78%
…	…	…
60	2.1403	20.61%

Figures 3, 4 and 5 show a plotting of predicted and measured macroporosity using Multiple Regression, conventional ANN, and proposed PENM, respectively. In each figure, correlation coefficients between measured macroporosity and predicted macroporosity are shown.

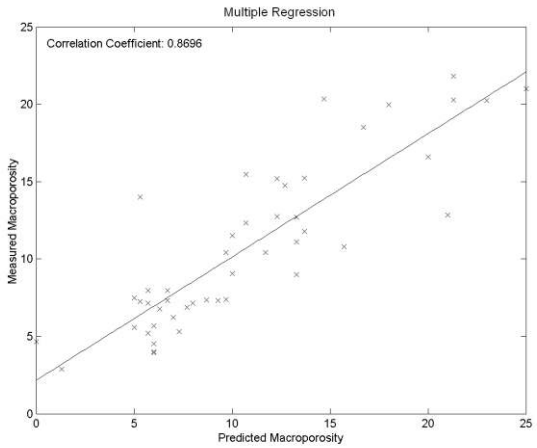


Figure 3 – Correlation Coefficient between Predicted and Measured Macroporosity using Multiple Regression.

In Uerê formation dataset [18], comparing correlation coefficients, it is possible conclude that macroporosity values predicted by PENM are the closest values of measured macroporosity.

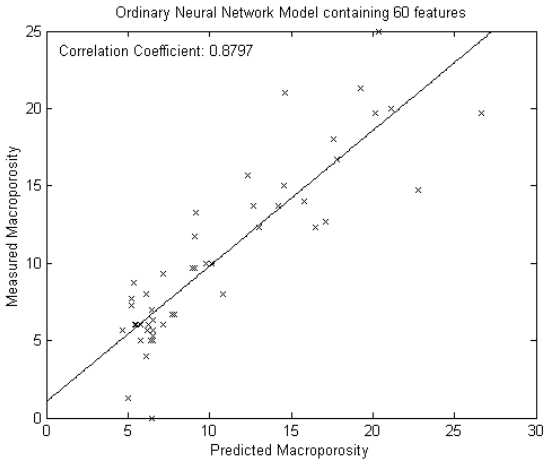


Figure 4 – Correlation Coefficient between Predicted and Measured Macroporosity using a Conventional ANN.

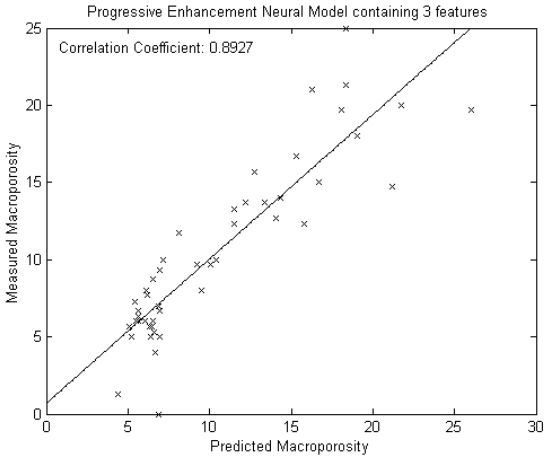


Figure 5 – Correlation Coefficient between Predicted and Measured Macroporosity using proposed PENM

Furthermore, figure 6 shows a comparison among residual errors obtained through application of different predictive approaches employed in this work. In order to improve results visualization, residual errors of each approach were sorted in ascendant order. This figure shows that PENM achieves less residual errors than the other used approaches.

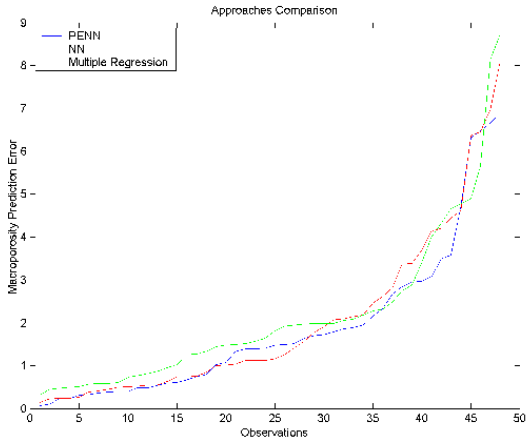


Figure 6 – Residual error comparison obtained through application of three different approaches

5. CONCLUSIONS

In this paper, we have proposed a Progressive Enhancement Neural Model to predict reservoir quality in sandstones. The main contributions of this paper include the following topics:

1) The Progressive Enhancement neural model successfully predicted macroporosity with a correlation coefficient of 0.8927 on fifty eight data points in a well in the Uerê formation. The conventional neural model predicted macroporosity with a correlation coefficient of 0.8789 in same dataset. The commonly used multiple regression predicted macroporosity with a correlation coefficient of 0.8696. Thus, the progressive enhancement neural model has proven to be a powerful approach to predict sandstones macroporosity.

2) Despite a small accuracy gain obtained with PENM, the main contribution of our approach is generating simpler models, which contain just a small subset of features, without accuracy loss. Obeying Ockham's razor principle, models generated by PENM approach must be preferred due to simplicity.

3) Our approach can generate more explainable models, because it ranks the importance of each feature regarding target feature prediction. One of the greatest criticisms to ANN use is the generation of unexplainable black box models. Hence, this study can indicate a way to open the black box.

We also applied our approach to create models to predict permeability and to classify observations according its petrofacies in this same dataset. In these experiments, small accuracy gains were repeated and simpler models were generated.

Experiments in datasets containing data from reservoirs with other diagenetic characteristics are being performed.

REFERENCES

[1] INTERNATIONAL ENERGY AGENCY. World energy outlook, 2007.

[2] KUPECZ, J.; GLUYAS J. ; BLOCK, S. Reservoir quality prediction in sandstones and carbonates: An Overview. In: KUPECZ J., GLUYAS J. and BLOCK S. (Eds). **Reservoir quality prediction in sandstones and carbonates**, AAPG Memoir 69, 1997.

[3] FANCHI, J. R. **Principles of Applied Reservoir Simulation**, 2nd ed., MA, USA, Elsevier Burlington, 2001.

[4] BELLMAN, R.; DREYFUSS, S. **Applied Dynamic Programming**, Princeton, USA, Princeton University Press, 1962.

[5] LOVE, K. M.; STROHMENGER, C.; WORONOW, A.; ROCKENBAUCH, K. Predicting Reservoir Quality Using Linear Regression Models and Neural Networks. In KUPECZ J., GLUYAS J. and BLOCK S. (Eds). **Reservoir quality prediction in sandstones and carbonates**, AAPG Memoir 69, p.47-60, 1997.

[6] BLOCH, S. Empirical predictions of porosity and permeability in sandstones, **AAPG Bulletin**, v. 75, p. 1145-1160, 1991.

[7] NIKRAVESH, M.; AMINZADEH, F.; ZADEH, L. A. **Soft computing and intelligent data analysis in oil exploration**, Ney York, USA, Elsevier, 2003.

[8] LIGTENBERT, J. H.; WANSINK, A. G. Neural Network Prediction of permeability in the El Garia Formation, Ashtart Oilfield, Offshore Tunisia. In: NIKRAVESH M., AMINZADEH F. and ZADEH, L.A (Eds). **Developments in Petroleum Science**, 51, p. 397-411, 2003.

[9] AL-BULUSHI, N.; KING, P. R.; BLUNT, M. J.; KRAAIJVELD, M. Development of artificial neural network models for predicting water saturation and fluid distribution. **Journal of Petroleum Science and Engineering,** v.68, p.197-208, 2009.

[10] ZAHEDI, G.; FAZLALI, A. R.; HOSSEINI, S. M.; PAZUKI, G. R.; SHEIKHATTAR, L. Prediction of asphaltene precipitation in crude oil, **Journal of Petroleum Science and Engineering**. v.68, p.218-222, 2009.

[11] AKIN, S.; ROSS, C. M.; KOVSCEK, A. R. Combination of well log and pore-scale data to predict pretrophysical properties of diatomite, **Journal of Petroleum Science and Engineering**. v.60, p.133-149, 2008.

[12] CHAO, W.; MIAN, C.; YAN, J. A prediction method of borehole stability based on seismic attribute technology, **Journal of Petroleum Science and Engineering**. v.65, p.208-216, 2009.

[13] SHI, G.; ZHOU, X.; ZHANG, G.; SHI, X.; LI. H. The use of artificial neural next term network analysis and multiple regression for trap quality evaluation: a case study of the Northern Kuqa Depression of Tarim Basin in western China. **Marine and Petroleum Geology**. v.21 i.3, p.411-420, 2004.

[14] GERSHEINFELD, N. **The nature of mathematical modeling**. Cambridge, UK, Cambridge University Press, 1998.

[15] BISHOP, C. M. **Neural networks for pattern recognition**, Oxford, UK, Oxford Press, 1996.
[16] MUNAKATA, T. Fundamentals of the New Artificial Intelligence: Neural, Evolutionary, Fuzzy and More, 2nd ed., London, UK, Springer, 2008.

[17] EFFROYMSON, M. A. Multiple regression analysis, In: A. Ralston, and H. S. Wilf (Eds), **Mathematical Methods for Digital Computers**. New York, USA, Wiley Sons, 1960, p191-203,

[18] LIMA R. D.; DE ROS, L. F. The role of depositional setting and diagenesis on the reservoir quality of Devonian sandstones from the Solimões Basin, **Brazilian Amazonia, Marine and Petroleum Geology**, v.19, p.1047-1071, 2002.

[19] HAN, J.; KAMBER, M. **Data Mining Concepts and Techniques**, San Diego, USA, Morgan Kaufman, 2001.