Journal of Natural Gas Science and Engineering 27 (2015) 98–108

Contents lists available at ScienceDirect

Journal of Natural Gas Science and Engineering

journal homepage: www.elsevier.com/locate/jngse

# Investigating the effect of correlation-based feature selection on the performance of neural network in reservoir characterization

Kabiru O. Akande [a,*], Taoreed O. Owolabi [b,c,1], Sunday O. Olatunji [d]

[a] Electrical Engineering Department, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia
[b] Physics Department, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia
[c] Physics and Electronics Department, Adekunle Ajasin University, Akungba Akoko, Ondo State Nigeria
[d] Computer Science Department, University of Dammam, Dammam, Saudi Arabia

ARTICLE INFO

*Article history:*
Received 19 June 2015
Received in revised form
20 August 2015
Accepted 21 August 2015
Available online 28 August 2015

*Keywords:*
Feature selection
Artificial neural network
Reservoir characterization
Permeability prediction

ABSTRACT

Accurate prediction of permeability is very important in characterization of hydrocarbon reservoir and successful oil and gas exploration. In this work, generalization performance and predictive capability of artificial neural network (ANN) in prediction of permeability from petrophysical well logs have been improved by a correlation-based feature extraction technique. This technique is unique in that it improves the performance of ANN by employing fewer datasets thereby saving valuable processing time and computing resources. The effect of this technique is investigated using datasets obtained from five distinct wells in a Middle Eastern oil and gas field. It is found that the proposed extraction technique systematically reduces the required features to about half of the original size by selecting the best combination of well logs leading to performance improvement in virtually all the wells considered. The systematic approach to feature selection eliminates trial and error method and significantly reduces the time needed for model development. The result obtained is very encouraging and suggest a way to improve hydrocarbons exploration at reduced cost of production. Furthermore, performance of ANN and other computational intelligence techniques can be improved through this technique.



## 1. Introduction

Permeability is one of the most important properties of underground reservoir containing hydrocarbons. It is normally estimated from core analysis and well test data. During the life cycle of a well, it is paramount to gather information which accurately gives the reservoir description, characterizes it and can be used in making decisions regarding its capacity for commercial exploitation and production. Knowledge of permeability is one of the three most fundamental requirements required in successfully managing a reservoir along with saturation and porosity. Accurate information about permeability helps petroleum engineers answer the important question as to whether hydrocarbon fluid trapped underground can be successfully exploited (Olatunji et al., Mar. 2014). Well logs are detailed record of the physical, chemical, electrical and other properties of the geological formations penetrated by instruments bored into the earth crust. Though well logs are usually taken in many oil and gas exploration, they do not really include the most important physical properties that determine the capacity of the reservoir or its productive capability as regards the hydrocarbons it contains. Rather, such important information as permeability, which is the most difficult of the properties to estimate (Mohaghegh et al., 1997), are derived from well logs using varieties of techniques and methods.

There are basically three ways in which permeability is determined from well logs. These methods are empirical, statistical and virtual measurement corresponding to empirically determined models, use of multiple regression analysis and artificial neural network respectively (Mohaghegh et al., 1997). The basic principle of empirical models are based on employing correlation between the three most important properties of rock formation which are permeability, porosity and irreducible water saturation (Virginia

* Corresponding author.Electrical Engineering Department, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia.
*E-mail addresses:* koakande@gmail.com (K.O. Akande), owolabitaoreedolakunle@gmail.com (T.O. Owolabi), oluolatunji.aadam@gmail.com (S.O. Olatunji).
[1] (On study leave).

http://dx.doi.org/10.1016/j.jngse.2015.08.042
1875-5100/

and Engineers, 1995).

One of the early empirical models was that of Kozeny who formulated an equation relating permeability to other measurable rock properties (Carrier, Nov. 2003). Despite the later modification by Carman (Carman-Kozeny equation. W, 2008), the model is only valid for packs of uniformly shaped spheres and moreover, surface area of packs can only be determined by core analysis with use of special equipment (Virginia and Engineers, 1995). Archie's model provide a basis for quantitative log interpretation by using electrical resistivity log in determination of some reservoir properties including permeability (Archie, April 2013). However, embedded in this model is saturation component which is one of the sources contributing the highest uncertainty to permeability determination (Wyllie and Rose, Apr. 2013). Tixier developed a model for determining permeability from electric-log resistivity gradients using empirical relationships between many rock formation properties (Tixier, 1949). The model not only has physical limitation but also its permeability estimate is just the average of the area corresponding to the gradient resistivity. He later developed a more widely-used simpler model based on the work of Wyllie and Rose (Wyllie and Rose, Apr. 2013). Also worthy of mention is the improved empirical log-derived permeability technique developed by Coates & Dumanoir which is applicable to formation that are not at irreducible water saturation and is the first to satisfy condition of zero permeability (Coates and Dumanoir, 1974). Coates later developed another model for permeability prediction which also satisfies the zero permeability condition at zero porosity but the formation must be at irreducible water saturation (Limited, 1988). Other contributions include Pirson's permeability formula which is unsuitable for high gravity crudes and deep depth (Pirson, 1963), Sheffield's proposed model that is valid only for clean sands (Virginia and Engineers, 1995) and Timur's work which require existence of residual water saturation for validity (Timur, Jul. 1968). Although, a better insight and deeper understanding of variables determining permeability has been gained from these developments, they are ill-suited to predict accurately the complex relationship existing between permeability and other properties (Mohaghegh et al., 1997).

Regression analysis is a statistical method widely used in estimating relationship existing between variables and is employed for prediction and forecasting. Multiple regression analysis is an extension of regression analysis which incorporates the use of independent variables also known as predictors to predict or estimate the value of a dependent variable also known as target. The work of Wendt and Sakurai are among the early works to have developed a permeability prediction model based on multiple regression. They pointed out in their work that the distribution of the predicted permeability using multiple regression is narrower than that of the original data (Wendt and Sakurai, 1986). The reason for this observation was given by Kendall and Stuart in their work stating that multiple regression only gives the best estimate on the average. A fundamental problem in multiple regression is that it inherently assumes a linear relationship between the predictors and the target rather than test whether this is the case or not. Also, it assumes that the predictors are not related among themselves. This is rarely the case between permeability and other properties of rock formation, hence multiple linear regression falls short in generalizing well for the unseen data and this prevents it from accurate prediction of permeability.

A significant amount of works has been done on permeability prediction from well logs using computational intelligence techniques. This is as a result of the excellent performance of computational intelligence techniques in solving many real life problems (Akande et al., 2015), (Owolabi et al., Aug. 2015), (Owolabi et al., 2015). Permeability of heterogeneous well has been predicted successfully from core data and well testing data (well depth and porosity) using radial basis function (RBF) neural network architecture (Jamialahmadi and. Javadpour, May 2000). A mathematically non-parametric nonlinear smooth modeling tool has also been used for finding the best combination of well logs to estimate permeability from a carbonate aquifer. This approach saves time during the construction and training of ANN model (Iturrarán-Viveros and Parra, Aug. 2014). Also, fuzzy logic was used in Lim and Kim (2004) for selecting the best related well logs with core porosity and permeability data and Neural network was subsequently used to develop transformation between the selected well logs and core measurements. The developed model from this method shows higher accuracy and more reliable reservoir properties estimation compared with conventional computing methods. Similarly, both (Zargari et al., May 2013) and (Afshari et al., Apr. 2014) estimated permeability from well logs using ANN. Both studies confirmed the superiority of ANN model over multiple regression methods.

Our work focused mainly on improvement of the generalization performance and predictive capability of ANN model using a correlation-based feature extraction technique. This improvement was achieved by implementing a technique which automatically selects the best combination of well logs data from the pool of available ones based on their correlation coefficient with cored permeability. The automatic selection of features saves time during model development as it eliminates trial and error method approach to feature selection (Iturrarán-Viveros and Parra, Aug. 2014). The results show significant improvement in performance of the developed model and indicate the effectiveness of the proposed technique.

## 2. .Model description and development

### 2.1. *Artificial neural network (ANN)*

ANN is a powerful learning algorithm which has proven very successful in learning complex patterns existing between variables. ANN learns in a parallel and distributed manner and is able to recognize complex patterns existing among several variables. There are several techniques used in training ANN but the one most often used is back-propagation feed forward neural network (BPNN) developed by Rumelhart (Rumelhart et al., 1986). It has been established that a neural network with one hidden layer and sufficient number of neurons in this hidden layer can learn any complex pattern existing between a target (variable to be predicted) and its predictors and is sufficient for approximating any function to whatever degree of accuracy required (Olatunji et al., Sep. 2011), (Hornik et al., Jan. 1989). Neural network usually has three or more layers consisting of input, hidden (may be more than one) and output layer. A layer is a logically grouped processing units of neurons which made up neural network. Each neuron in a layer (apart from input layer) is connected to all the neurons in a preceding layer through a set of adjustable weights. These tunable weights are means through which the resultant output of the network and its performance can be optimized with adequate training. The output of each neuron is decided by an activation function and the commonly-used one is sigmoid activation function. Network layers are arranged in hierarchical order such that each layer receives weighted inputs from the one preceding it and feeds its own weighted output to the next layer (Rojas, Aug. 1996). This arrangement is referred to as feedforward neural network. A typical neural network architecture with a single hidden layer is shown in Fig. 1.

Using backpropagation algorithm, the network described above