PERGAMON

Computers & Geosciences 26 (2000) 907–913

COMPUTERS & GEOSCIENCES

Multiple permeability predictions using an observational learning algorithm

Patrick M. Wong[a,*], Min Jang[b], Sungzoon Cho[c], Tom D. Gedeon[d]

[a]*School of Petroleum Engineering, University of New South Wales, Sydney, 2052, Australia*
[b]*Department of Computer Science and Engineering, Pohang University of Science and Technology, Pohang, 790-784, South Korea*
[c]*Department of Industrial Engineering, Seoul National University, Seoul, 151-742, South Korea*
[d]*School of Information Technology, Murdoch University, Murdoch, WA 6150, Australia*

Received 15 March 1999; accepted 9 September 1999

Abstract

Reservoir permeability is a critical parameter for the evaluation of hydrocarbon reservoirs. Well log data are frequently available to infer this parameter along drilled wells. Many fundamental problems remain unsolved by most predictive models. This paper introduces the use of an improved neural network trained by an observational learning algorithm to provide solutions for two particular problems: the generation of additional or "virtual" samples when the number of training data is insufficient; and the generation of multiple permeability values at the same reservoir depth for reliability analyses. The methodology is illustrated by a case study in western Australia. Four drilled wells with well logs and core permeability are used in this study. The data from the first two wells are used for training, while the others are used as unseen data to test the performance of the model. The results show that the proposed method gives smaller error compared to multiple linear regression and other neural networks (simple committee networks and bootstrap aggregating). It also provides valuable information on the reliability of the permeability predictions which is consistent with the geological studies.

Keywords: Permeability; Well logs; Neural networks; Observational learning; Bootstrapping; Virtual samples

1. Introduction

Reservoir permeability is a critical parameter for the evaluation of hydrocarbon reservoirs. The most accurate method to date for measuring such an important property is core analysis. It is well known that closely-spaced core permeability values are often not available because of unfavourable wellbore conditions and high cost of coring. Well log data, however, are abundant and are frequently used to infer permeability along the drilled wells. Although no well log is currently capable of measuring permeability directly, correlating well logs with core permeability at the cored well has become a common practice in the industry. The subsequent correlation model can be used to predict permeability at the uncored intervals and wells, providing appropriate well logs are available.

Many empirical equations are available to transform well log data to permeability (Coates and Dumanoir, 1973; Johnson, 1994). These models often require a

* Corresponding author.
E-mail address: pm.wong@unsw.edu.au (P.M. Wong).

0098-3004/00/$ - see front matter
PII: S0098-3004(00)00026-1

labour-intensive exercise to adjust constants or exponents or to introduce compensations. Despite these observations, theoretical relations between permeability and porosity have been sought. For example, Kozeny-Carmen theory relates permeability to porosity and the specific area of a porous rock with pores treated as an idealised bundle of capillary tubes. This theory treats the highly complex porous medium in a very simple manner and ignores the influence of conical flow in the constrictions and expansions of flow channels.

The statistical approach is comparatively a more versatile approach to the problem of permeability prediction. It makes use of the available core permeability (the dependent variable) and develops functional relationships with the well log data (the independent variables). It, however, requires the assumption and satisfaction of multi-normal behaviour and linearity, and hence it must be applied with caution.

Besides statistical methods, neural networks, or more specifically, multilayer perceptrons (MLP), have become increasingly popular in well log analysis. This intelligent technique is non-linear and non-parametric, and has been applied to permeability prediction from well logs. Recent comparison studies have shown that MLP gave better performance compared to empirical equations (Balan et al., 1995; Malki et al., 1996) and statistical methods (Huang et al., 1996; Mohaghegh et al., 1995; Rogers et al., 1995; Wong et al., 1995, 1998). While MLP have been shown to have great promise in permeability prediction, none of the previous studies report any information about the reliability of the estimates. The other important problem in applying such methods in practice is how we can obtain a sufficiently good estimator trained by only a small data set. These are in fact critical issues in geosciences.

The objective of this work is to provide an improved MLP method trained by an observational learning algorithm, via the use of multiple networks and the generation of virtual samples (see later sections), to estimate permeability from well logs in petroleum reservoirs. The proposed methodology makes use of appropriate well log and core permeability to provide multiple permeability predictions at every reservoir depth. The approach is significant in that it is simple, easy to implement and permits evaluation of the reliability of the permeability predictions.

2. Multilayer perceptrons

2.1. A typical model

Multilayer perceptrons (MLP) are popular for developing highly non-linear relationships between known inputs and outputs. They are particularly useful for solving pattern recognition problems. A typical model is shown below:

$$f(x) = g\left(b_0 + \sum_{j=1}^{\mathrm{m}} b_j \cdot g\left(a_{0j} + \sum_{i=1}^{n} a_{ij} \cdot x_i\right)\right) \tag{1}$$

and

$$g(z) = (1 + e^{-z})^{-1} \tag{2}$$

where $x = x_1 \ldots x_n)$ represents n types of well log values at every sampled depth, a, b are coefficients (or "weights") and $f(\cdot)$ is the permeability estimator. $g(\cdot)$ is a transfer function which increases the non-linearity and complexity of the relations. Eq. (1) is also a universal function approximator which can fit any continuous function with a desired precision (Funahashi, 1989). Note that if $g(z) = z$, Eq. (1) becomes the standard multiple linear regression estimator.

With the use of a set of known data pairs or training patterns, $\{(x_k, y_k) \mid k = 1 \ldots n_{\mathrm{p}}\}$, where n_{p} is the number of patterns and y is the target variable, which can be obtained from the cored wells after proper depth-matching exercise, the connection weights a, b can be approximated using a variety of gradient-based optimisation algorithms (Bishop, 1995), or other random search routines such as genetic algorithms (Huang et al., 1998). This paper uses the Levenberg-Marquardt algorithm (Bishop, 1995), a kind of quasi-Newton method, which is faster than the conventional backpropagation algorithm (Bishop, 1995).

Like most optimisation algorithms, the Levenberg-Marquardt algorithm requires a set of initial weights (randomised values) as the starting point. The weights are iteratively updated in such a way that the objective function is optimised. A typical objective function is the mean square error (MSE):

$$\mathrm{MSE} = \frac{1}{n_{\mathrm{p}}} \sum_{k=1}^{n_{\mathrm{p}}} (f(x_k) - y_k)^2. \tag{3}$$

The network can be optimised by lowering the MSE with respect to a, b. The estimator can then be used to predict permeability at the uncored intervals or wells where the same well logs are also available.

2.2. Multiple predictions

Most neural network studies reported in geosciences provide only a single realisation of the predictions. It is, however, important to infer the reliability of each and every prediction along the wells because predictions involve uncertainty. Two reasons are given here:

1. Most well log data and core permeability values are corrupted by "natural noise" (such as uncertain depth-matching, core testing conditions, thin beddings);

2. Core permeability and log-derived permeability are not of compatible measurement scales as they measure different volumes of rock.

In order to account for the reliability of the predictions, it is necessary to have a distribution of predictions at the same reservoir depth along the drilled wells. Many neural network methods are available for generating a distribution of outputs but are still not yet popular in petrophysical evaluation. Examples are the use of multiple (or a committee of) networks or the more recent Bayesian networks (MacKay, 1992). With the use of an input vector, the former provides multiple predictions, while the latter provides the mean and the variance of the predictions.

In this study, we use a multiple network because of its simplicity and its flexibility to provide any number of predictions. Once multiple predictions are produced, it is straightforward to obtain the corresponding statistical information (e.g. minimum, average and maximum values).

A simple committee network (Sharkey, 1996) is an example of multiple networks. This committee of networks is composed of a set of networks with different initial weights while holding all the other parameters of the network constant. With different initial weights, each network is trained by the gradient descent algorithm and is able to produce different outputs. It has been shown to be more effective than a single network in terms of generalisation performance (Parmanto et al., 1996). The final output can be computed as a simple average of the outputs produced by the individual networks, or we can treat each individual output as a realisation of a random process. We can also construct a local probability distribution function at each estimation point and perform stochastic simulation, such as the use of Monte Carlo sampling to simulate a final output.

The other popular way to generate multiple predictions is bootstrap aggregating or "bagging" (Brieman, 1994). Bootstrapping replicates a multitude of training data sets by sampling with replacement from the original data set. These bootstrapped data sets are then used to train a committee of networks and produce multiple predictions as in the simple committee networks. The major difference between a simple committee network and bagging is that one uses different initial weights while the other uses different training data sets. Use of different initial weights in bagging is also possible.

3. Observational learning algorithm

MLP are examples of universal function approximators. In practice, however, there are several factors which make it difficult to fit a continuous function accurately. The major assumption is that a sufficient number of well-represented training samples are available. In many real world problems, training patterns are expensive to obtain. Since the problem complexity dictates the size of the network, we may end up with too few training patterns for a network of a certain size. This results in over-fitting and hence poor generalisation.

In Cho et al. (1997), a technique known as "virtual sampling" was proposed to improve generalisation by artificially generating "virtual samples" from a committee of networks and using them to train the network in addition to the original training patterns. This technique has been applied to permeability and porosity prediction from well logs and the results were improved by 10% and 35% respectively (Cho et al., 1998; Jang et al., 1998). Observational learning algorithm is the improved virtual sampling technique with the incorporation of bootstrapping.

Observational learning is found among children (Bandura, 1971) who learn not only by doing but also by observing other children. This approach can be implemented by treating each trained network as a child. A virtual sample is then generated by picking a nearby point of a real sample for its input. The output part of the virtual sample is then determined by observing what the networks in the committee produce as output when the newly picked input is given. A single output could be obtained by averaging the output values, or all the output values could be used.

The learning algorithm for a multidimensional data set can be described as follows. Consider L neural networks, f_i, $i = 1 \ldots L$, and original data set, $D = \{(x_k, y_k) \mid k = 1 \ldots n_p\}$, where $x_k = (x_k^1 \ldots x_k^n)$ and $y_k = (y_k^1 .. y_k^q)$ are n and q dimensional vectors, respectively. First, in order to construct different training data sets for individual networks, the bootstrapping method is employed as follows. Let s_i denote the index of the data element which was selected at the ith time. The initial training data set for network f_i is then defined as $D_i = \{(x_{s}, y_{s}) \ldots (x_{s_{np}}, y_{s_{np}})\}$.. Since the elements are selected with replacement, duplicates among the elements in D_i exist. By repeating this process for L networks, we construct initial data sets for individual networks.

After creating the initial training data sets, the training step (T-step) and observation step (O-step) are repeated for G epochs. Value G is empirically set. In the T-step, the weights of a neural network are updated using both its initial data and virtual data sets. At epoch t, weights are updated as follows:

$$w_i^{t+1} = w_i^t + \Delta w_i^t \tag{4}$$

where w_i^t is a set of weights of the network f_i and Δw_i^t is the weight change. For normal backpropagation training, the weight change is given by $\Delta w_i^t = -\eta_i \nabla E(w_i)$, where η_i is the learning rate and $E(w_i)$ is the standard mean square error function.

After updating the weights of the individual networks in the T-step, the virtual data for each network are generated from the other networks in the O-step. The virtual data of network f_i^t are generated as follows: At first, an input vector, x_{s_j}, is selected from the initial data set, D_i, and a virtual input vector v_{s_j} is created by adding a Gaussian noise with a zero mean and a covariance of Σ. Then, the virtual output vector corresponding to the input is computed using the $-i$ ensemble, i.e., an ensemble output without network f_i, which is denoted as $\vec{f}_{-i}$. A variety of methods for constructing a consensus output exists. For this study, we employed an averaging method. The virtual data generating process is repeated for all data in D_i. We now generated a virtual data set which is of the same size as the initial data set. In order to construct the virtual data sets for all individual networks, this process is again repeated L times. Now after this O-step is finished, each member is incrementally trained with both the newly generated virtual data set and the initial data set.

After the T- and O-steps are repeated for G epochs, the final output of the ensemble is computed by the simple averaging method as:

$$\vec{f}^G = \frac{1}{L}\sum_{i=1}^{L} f_i^G. \tag{5}$$

The observational learning algorithm can be summarised as follows:

1. Let the number of networks in the committee be L.
2. Let D_i, $i = 1 \ldots L$ be the bootstrapped data sets from the original data set D.
3. Repeat the following two steps during G epochs: T-step. For each network f_i, weights are updated as follows:

$$E_i^t(w_i) = \sum_{(x,\,y)\in D_i \cup V_{-i}} (f_i^t(x) - y)^2 \tag{6}$$

$$w_i^{t+1} = w_i^t - \eta_i \nabla E_i^t(w_i) \tag{7}$$

O-step. For each network f_i, $i = 1 \ldots L$, the observational data set V_{-i} is generated as follows:

$$V_{-i} = \{(v_{s_j}, \vec{f}_{-i}(v_{s_j}) \mid v_{s_j} \sim N(x_{s_j}, \Sigma), x_{s_j} \in D_i, \; j = 1 \ldots n_p\} \tag{8}$$

where

$$\vec{f}_{-i} = \frac{1}{L-1}\sum_{j=1,\,j\neq i}^{L} f_j^t. \tag{9}$$

4. Obtain the outputs by simple averaging:

$$\bar{f} = \frac{1}{L}\sum_{i=1}^{L} f_i. \tag{10}$$

In this study, we empirically set the covariance by a diagonal matrix with diagonal elements set to 0.01. In practice, the observational learning algorithm is robust to the parameter.

4. Case study

4.1. Objective

In this paper, we used a data set from the North West Shelf (offshore western Australia) which has well logs and core permeability from four oil and gas wells. The formation contains many thin beds composed of eleven dominant lithofacies. The major lithofacies are sandstone, mudstone, carbonated-cemented facies, conglomerate and glauconite. The well logs used for the analyses are gamma ray, deep resistivity, sonic travel time, bulk density and neutron porosity. Two wells (Well 1 and Well 2) were used for training and network optimisation. The other two wells (Well 3 and Well 4) were used to blind-test the model.

In this study, we compared the performance of four different estimators: multiple linear regression (MLR), simple committee networks (SCN), bagging (BAG) and the observational learning algorithm (OLA). Once we obtain the results, we will evaluate the reliability of the models by comparing the predictions with the core permeability. The indicator used for performance comparison was the MSE as shown in Eq. (3). We will achieve this task by evaluating if the average prediction is close to the core data at the test wells.

4.2. Model setup

In this study, we used seven input neurons (six for well logs and one for lithofacies indicator) and one output neuron (permeability). All the data were normalised in the range of (0,1). The number of data pairs available in each of the four wells is 152, 156, 115 and 140, respectively. The first two wells were used for training and the last two wells were used for testing.

Five-fold cross-validation was used to determine the optimal number of hidden neurons. A range of 2–10 hidden neurons was tried and the minimum average validation error was found when there were 5 hidden neurons. Ten ($L = 10$) networks were used in the com-

Table 1
Mean square errors ($\times 10^{-2}$) for different estimators. MLR is multiple linear regression, SCN is simple committee network, BAG is bagging and OLA is observational learning algorithm

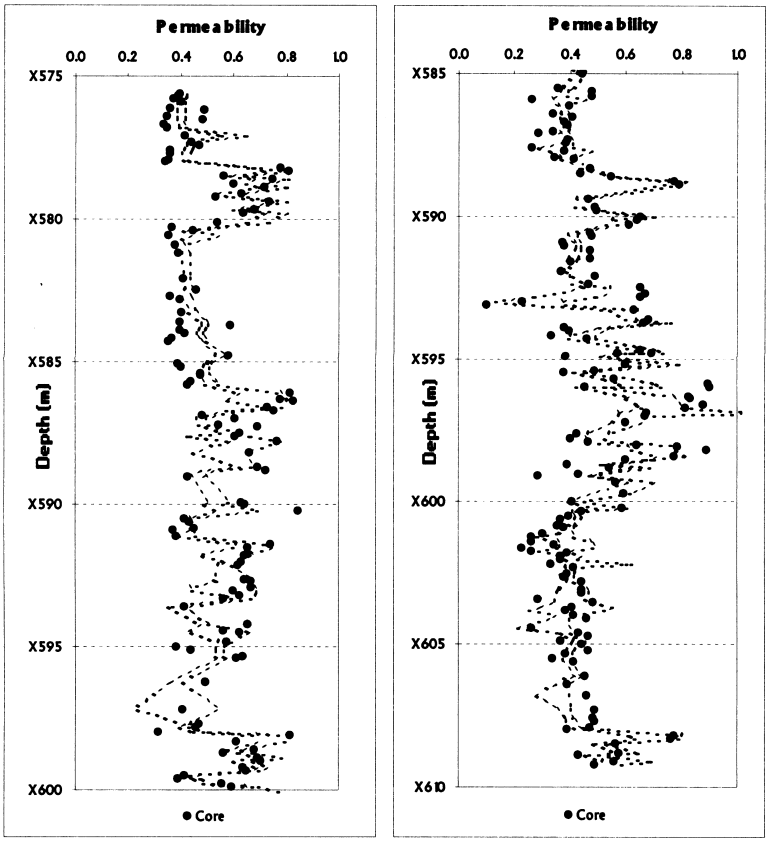
	Well 3			Well 4		
	Min	Mean	Max	Min	Mean	Max
MLR		1.210			1.461	
SCN	0.709	1.108	2.091	0.714	0.939	1.717
BAG	0.773	2.703	7.907	0.747	2.212	7.154
OLA	0.665	0.741	0.841	0.758	0.838	0.999

mittee and each network was trained by the Levenberg-Marquardt algorithm. In this study, the same number of virtual samples as the original training samples (308) were generated (i.e. $G = 308$). This was used as the termination criterion for the observational learning. Hence, a total of 716 patterns was used for observational learning.

4.3. Results

After training, the networks were applied to the last two wells for prediction. Since $L = 10$, we obtained 10 predictions for each reservoir depth. The results are summarised in Table 1. The minimum, average and maximum MSEs were computed for each test well. Note that only one prediction at each depth was obtained from MLR. Different MLR models were developed for different lithofacies for improved performance (Wong et al., 1995). Three of the eleven facies, however, did not have enough data pairs for their individual MLR models. Hence they were lumped into one cluster. This resulted in a total of 9 MLR models. The MSEs of the test wells were 0.0121 and 0.0146 respectively. Since MLR is a linear model, therefore its MSE could be set as the maximum acceptable error for the non-linear neural networks. The MLR's MSEs were also used as the basis for comparison.

From Table 1, the average results from the SCN



a) Well 3. b) Well 4.

Fig. 1. Permeability profiles (min, max) derived by observational learning algorithm at test wells. Note that permeability values are normalised.

(1.108) were better than those obtained from the MLR in both of the test wells with a much greater improvement in Well 4 (36%). Hence a committee of networks using different initial weights provided more robust results than MLR.

The BAG results were unsatisfactory. The MSEs were 123% and 51% larger than MLR's in Well 3 and Well 4 respectively. This was due to its ability to generate a wider range of predictions resulting from different training sets. This problem is worse if there is only a small number of training patterns as each of the bootstrapped data sets has a higher probability to contain the same training pattern. Generally speaking, the BAG predictor is expected to perform better if we use a larger number of networks, say $L > 50$, which is computationally more expensive.

The other practical solution for improving BAG performance is to increase the number of training patterns which is the basic objective of virtual sampling. As shown in Table 1, the average results of OLA were best with 39% improvement in Well 3 and 43% improvement in Well 4 compared to MLR.

For visualisation purposes, we plotted the results showing the minimum and maximum predictions (obtained from the 10 predictions) at each depth. Fig. 1 shows the OLA-derived permeability profiles for the test wells together with the core permeability. The minimum and maximum predictions are also shown. In order to assess the reliability of the predictions, one may look at the range of predictions (R) as the uncertainty indicator:

$$R_k = \max\{f_i(x_k)\}_1^L - \min\{f_i(x_k)\}_1^L. \quad (11)$$

Using the above indicator, the predictions become more unreliable when R is large. Fig. 2 shows the OLA-derived R profiles for both of the test wells. The profiles were consistent with the geological studies as R is large when there is a rapid facies transition. This is mainly due to the physical limitation of well logs as the log measurements are strongly affected in the presence of thin beds (the "shoulder" effect), and hence log-derived permeability values from any methods are

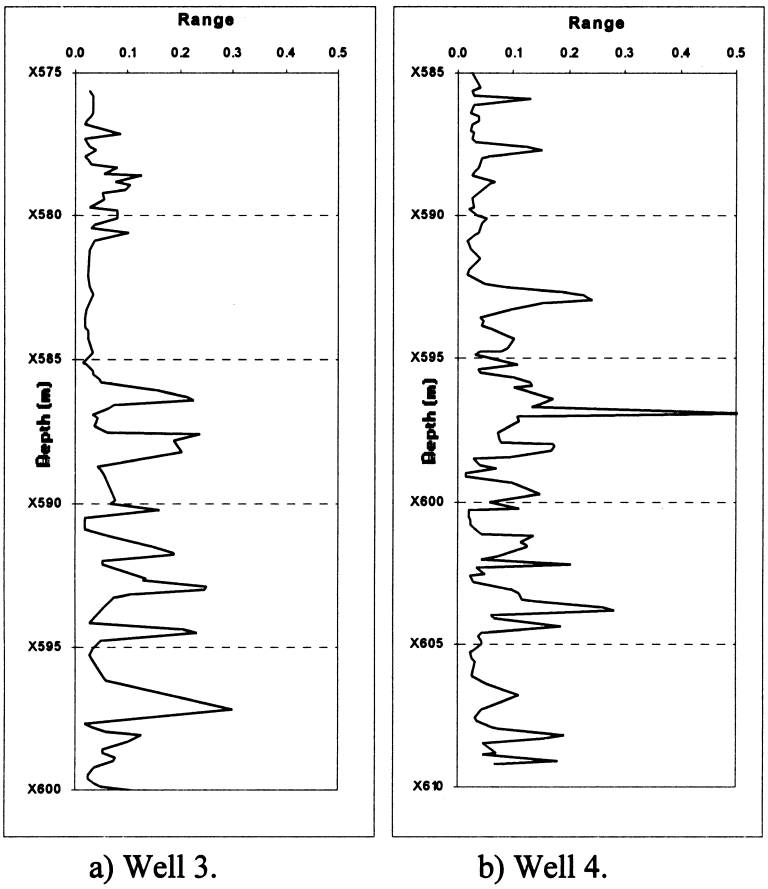


Fig. 2. Range (R) profiles derived by observational learning algorithm at test wells.

less reliable. This information is valuable for making subsequent modelling decisions.

5. Conclusions

This work introduces the use of an observational learning algorithm to predict permeability from well logs based on a given set of training patterns. The technique is implemented in multilayer perceptrons (neural networks) and is particularly useful when there are insufficient training patterns. It also generates multiple predictions at the same estimation point. The technique is demonstrated via a field example in offshore western Australia. Four drilled wells with well logs and core permeability are used in this study. The data from the first two wells are used for training, while the others are used as unseen data to test the performance of the model. The results show that the proposed method gives smaller error compared to multiple linear regression and other neural networks (simple committee networks and bootstrap aggregating) and produces valuable information on the reliability of the permeability predictions.

Acknowledgements

This work was supported by the Australian Research Council's Small Grant in 1998 for the project, entitled An Uncertainty Analysis of Permeability Predictions from Well Log Data using Neural Networks with Virtual Samples. This work was funded by the Brain Science & Engineering Research Program (the Ministry of Science and Technology) and the BK-21 Program (the Ministry of Education) in Korea.

References

Balan, B., Mohaghegh, S., Ameri, S., 1995. State-of-the-art in permeability determination from well log data: part 1 — a comparative study, model development. SPE 30978. In: Proceedings of the SPE Eastern Regional Conference and Exhibition, pp. 33–42.

Bandura, A., 1971. Social Learning Theory. General Learning Press, New York, 46 pp.

Bishop, C., 1995. Neural Networks for Pattern Recognition. Oxford University Press, New York, 482 pp.

Brieman, L., 1994. Bagging predictors. Machine Learning 24, 123–140.

Cho, S., Jang, M., Chang, S., 1997. Virtual sample generation using a population of networks. Neural Processing Letters 5 (2), 83–89.

Cho, S., Wong, P.M., Kwon, Y., Gedeon, T.D., 1998. Estimation of rock permeability in petroleum reservoir. In: Proceedings of 4th International Conference on Engineering Applications of Neural Networks (EANN), Gibraltar, 10–12 June, pp. 161–164.

Coates, G.R., Dumanoir, J.L., 1973. A new approach to improved log-derived permeability. In: Proceedings of SPWLA 14th Annual Logging Symposium, Paper R.

Funahashi, K., 1989. On the approximate realization of continuous mappings by neural networks. Neural Networks 2, 189–195.

Huang, Z., Shimeld, J., Williamson, M., Katsube, J., 1996. Permeability prediction with artificial neural network modelling in the Venture Gas Field, offshore Eastern Canada. Geophysics 61 (2), 422–436.

Huang, Y., Wong, P.M., Gedeon, T.D., 1998. Prediction of reservoir permeability using genetic algorithms. AI Applications 12 (13), 67–75.

Jang, M., Cho, S., Wong, P.M., Fung, C.C., Wong, K.W., 1998. Rock porosity prediction using multilayer perceptrons. In: Proceedings of International Conference on Neural Information Processing (ICONIP), Kitakyushu, 21–23 October, Vol. 2, pp. 1016–1019.

Johnson, W.W., 1994. Permeability determination from well logs and core data. SPE 27646. In: Proceedings of the SPE Permian Basin Oil and Gas Recovery Conference, Midland, Texas, 16–18 March.

MacKay, D., 1992. A practical Bayesian framework for backpropagation networks. Neural Computation 4 (3), 448–472.

Malki, H.A., Baldwin, J.L., Kwari, M.A., 1996. Estimating permeability by the use of neural networks in thinly bedded shaly gas sands. SPE Computer Applications April, pp. 58–62.

Mohaghegh, S., Balan, B., Ameri, S., 1995. State-of-the-art in permeability determination from well log data: part 2 — verification, accurate permeability predictions, the touchstone of all models. SPE 30979. In: Proceedings of the SPE Eastern Regional Conference and Exhibition, pp. 43–47.

Parmanto, B., Munro, P.W., Doyle, H.R., 1996. Reducing variance of committee prediction with resampling techniques. Connection Science 8, 405–426.

Rogers, S.J., Chen, H.C., Kopaska-Merkel, D.C., Fang, J.H., 1995. Predicting permeability from porosity using artificial neural networks. American Association of Petroleum Geologists Bulletin 79 (12), 1786–1797.

Sharkey, A.J.C., 1996. On combining artificial neural networks. Connection Science 8, 299–314.

Wong, P.M., Henderson, D.J., Brooks, L.J., 1998. Permeability determination using neural networks in the Ravva Field, offshore India. SPE Reservoir Evaluation and Engineering 1 (2), 99–104.

Wong, P.M., Taggart, I.J., Jian, F.X., 1995. A critical comparison of neural networks and discriminant analysis in lithofacies, porosity and permeability predictions. Journal of Petroleum Geology 18 (2), 191–206.