Confidence Bounds of Petrophysical Predictions From Conventional Neural Networks

Patrick M. Wong, Alexander G. Bruce, and Tamás (Tom) D. Gedeon

***Abstract*—Neural networks are powerful tools for solving the complex regression problems which abound in geosciences. Estimation of prediction confidence from neural networks is an important area. Many procedures are available to date, but it is often tedious for practitioners to implement such procedures without significant modification of the existing learning algorithms. In many cases, the procedures are also computationally intensive. This paper presents a practical solution using conventional backpropagation networks with simple data pre-processing and post-processing algorithms. The methodology involves conversion of the target outputs into linguistic variables (classes) prior to learning. When the classification network converges, minimum and maximum predictions are derived from the output activations using a simple averaging algorithm. Two examples from petroleum reservoirs are used to demonstrate the proposed methodology. The results show that the confidence bounds of the petrophysical predictions are realistic in both cases. The proposed methodology is generally useful, and can be implemented in simple spreadsheets without altering any existing neural network code.**

***Index Terms*—Confidence, fuzzy logic, neural networks, petroleum reservoirs, petrophysics.**

I. Introduction

Petrophysical prediction problems are highly nonlinear. Recent use of artificial neural networks has outperformed many conventional techniques. Their abilities to learn, adapt and generalize from data offer a number of advantages in many geoscience applications [1]. Many advanced models are now available for significant reduction of prediction error and perform extremely well. One of the major challenges remaining in predictive model design is to generate accurate and precise estimates when actual data are not available for performance comparison. Therefore, reliable and realistic estimation of confidence bounds (i.e., estimated range of predictions) for the neural network outputs is an important research area.

A number of confidence estimation methods have been presented. Generally speaking, we can divide the methods into three categories:

1) manipulation of training data [2], [3];
2) network ensembles [4], [5];
3) advanced learning algorithms [6]–[9].

Although many of the previous techniques provide good performance on both synthetic and real data sets, it is often too tedious for practitioners to implement the procedures in their normal working environment. In many cases, such techniques are computational intensive and the adoption requires additional coding and/or significant modification of the existing procedures.

This paper presents a practical procedure to estimate confidence bounds from conventional backpropagation networks. The procedure involves conversion of the target outputs into linguistic classes (natural language labels) prior to learning. When the classification network converges, minimum and maximum predictions are derived from the output activations through a simple averaging algorithm. The data pre-processing and post-processing procedure is general and can be implemented in simple spreadsheets without any additional program code. Two examples from petroleum reservoirs are used to demonstrate the proposed methodology. The first example is permeability from well logs and the second one is porosity mapping from seismic attributes.

Manuscript received April 6, 2001; revised April 16, 2002.

P. M. Wong and A. G. Bruce are with the School of Petroleum Engineering, University of New South Wales, Sydney NSW 2052, Australia.

T. D. Gedeon is with the School of Information Technology, Murdoch University, Murdoch WA 6150, Australia.

Publisher Item Identifier 10.1109/TGRS.2002.800278.

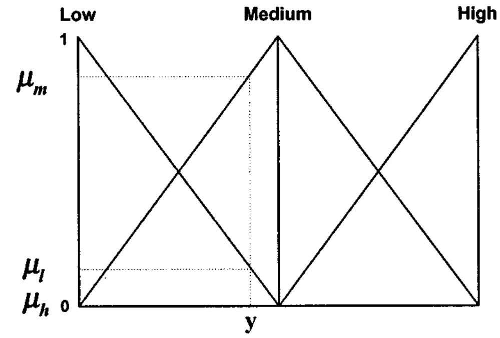

Fig. 1. Example of fuzzy representation of data value y.

II. Methodology

For simplicity' sake, we will present the proposed methodology for solving a multiple-inputs–single-output (MISO) regression problem. The methodology involves three steps.

A. Data Preprocessing

The first step starts with conversion of the target outputs of the training set into n linguistic variables (classes). This essentially transforms the MISO regression problem into a multiple outputs (MIMO) classification problem. The motivation of problem transformation also has some practical advantages. Human experts tend to work better in a classification environment and are able to reason well in terms of classification accuracy rather than mean square error. This is particularly true for geoscientists.

The target outputs can be divided into different classes according to their magnitude. Linguistic classes, such as "low," "medium," and "high" are popular. Each data value can be binned into the appropriate classes based on some pre-set cut-offs. The cut-offs can be obtained from histograms (percentiles or equal intervals) or the domain experts.

The linguistic classes can be "crisp" or "fuzzy." The crisp, i.e., Boolean, logic concept allows each data value to belong to one and only one class. Very often, the classification results are extremely sensitive to the cut-off values. The fuzzy logic concept, however, allows each data value to belong to more than one class with different degrees of membership (ranging from zero to one). In fact, fuzzy logic is a superset of the Boolean logic that has been extended to handle the concept of "partial truth," that is, truth values between "completely true" and "completely false" [10], [11].

Fig. 1 shows a typical fuzzy representation of data. In this example, the data value y is transformed into three linguistic variables ("low," "medium," "high") with membership values $(\mu_l,\ \mu_m,\ \mu_h)$ using triangular membership functions. Since y is between "low" and "medium," the value belongs to both "low" and medium" with different degrees of membership. Note that $\mu_h = 0$ in this case. It is easy to see that fuzzy classes become crisp when the membership values are only 0 or 1. For example, (1,0,0) is only "low," (0,1,0) is only "medium" and (0,0,1) is only "high." A fuzzy example is (0.2,0.8,0.0), which means the value belongs to low with a membership 0.2, medium 0.8, and high 0.0 (e.g., point y in Fig. 1). This value is "mostly medium."

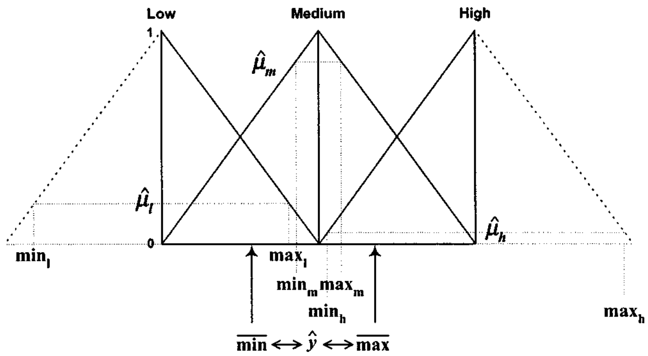

Fig. 2. Back-transformation of fuzzy membership values.

B. Learning From Data

Once the outputs of the training set are converted into linguistic classes, the input–output data can be presented to the MIMO neural networks for the optimization of connection weights (analogous to regression coefficients). The most popular learning algorithm is known as "backpropagation" [12], which is a type of gradient descent algorithm for the minimization of a given error function. A typical error function E is defined as

$$E(\mathbf{w}) = \sum_{k}^{K} \sum_{i}^{n} [\mu_{ik} - f(\mathbf{x}_k;\, \mathbf{w})]^2 \tag{1}$$

where $\mathbf{w}$ is the weight vector, $f(.)$ is the neural network estimator, $\mathbf{x}$ is a multidimensional input vector with n outputs. Once the network has converged for all the K data vectors (i.e., error is minimized), the least squares estimate of the true weight vector is obtained and the network can be used for prediction. The predicted outputs are

$$\hat{\boldsymbol{\mu}} = f(\mathbf{x};\, \mathbf{w}). \tag{2}$$

Further details of neural networks can be found in [13].

C. Confidence Bounds Estimation

After training, input data vectors with unknown outputs can be presented to the network for prediction. The multiple output activations, $\hat{\boldsymbol{\mu}} = (\hat{\mu}_1, \ldots \hat{\mu}_n)$, are the fuzzy membership values for each input vector. One way to quantify the confidence of the neural classifier is to calculate the entropy E from the membership values

$$E = -\sum_{i}^{n} \hat{\mu}_i \log \hat{\mu}_i. \tag{3}$$

The above measure gives a value between 0 (very confident) and $E_{\max} = \log n$ (not confident). For a four-class problem, the maximum entropy is 0.602. Note that $\hat{\mu}_i$ has to be normalized such that $\sum_i^n \hat{\mu}_i = 1$.

Entropy measures only the confidence of the classification task. However, it is important to obtain the prediction of the original output variable y and its confidence bounds. The major contribution of this work is to introduce a back-transformation of the fuzzy membership values into minimum and maximum $(\min\text{–}\max)$ predictions of variable y for each input vector.

Fig. 2 shows an example of the back-transformation process. For each fuzzy set, the $\min\text{–}\max$ predictions, $(\min_i, \max_i)$, $i = 1, \ldots, n$, can be obtained from the corresponding triangular

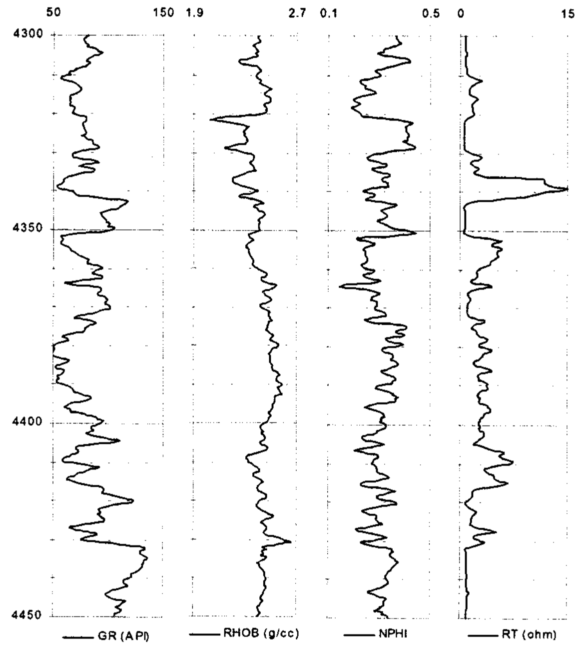

Fig. 3. Well logs.

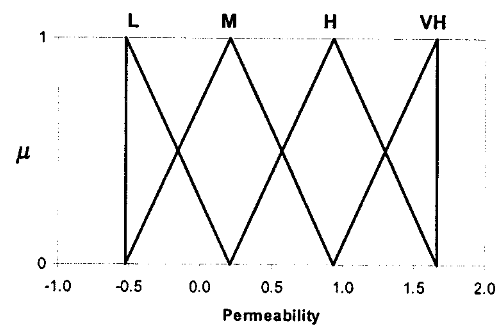

Fig. 4. Fuzzy transformation of the logarithm of permeability.

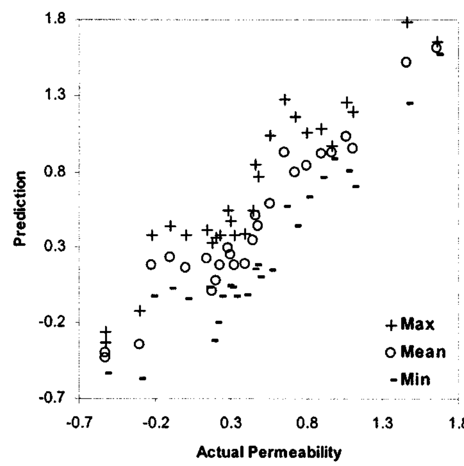

Fig. 5. Permeability prediction and confidence bounds for the 27 training samples.

function. The final $\min\text{–}\max$ predictions, $(\overline{\min}, \overline{\max})$, can be weighted by the membership values

$$\overline{\min} = \frac{\sum_i^n \hat{\mu}_i \min_i}{\sum_i^n \hat{\mu}_i} \tag{4}$$

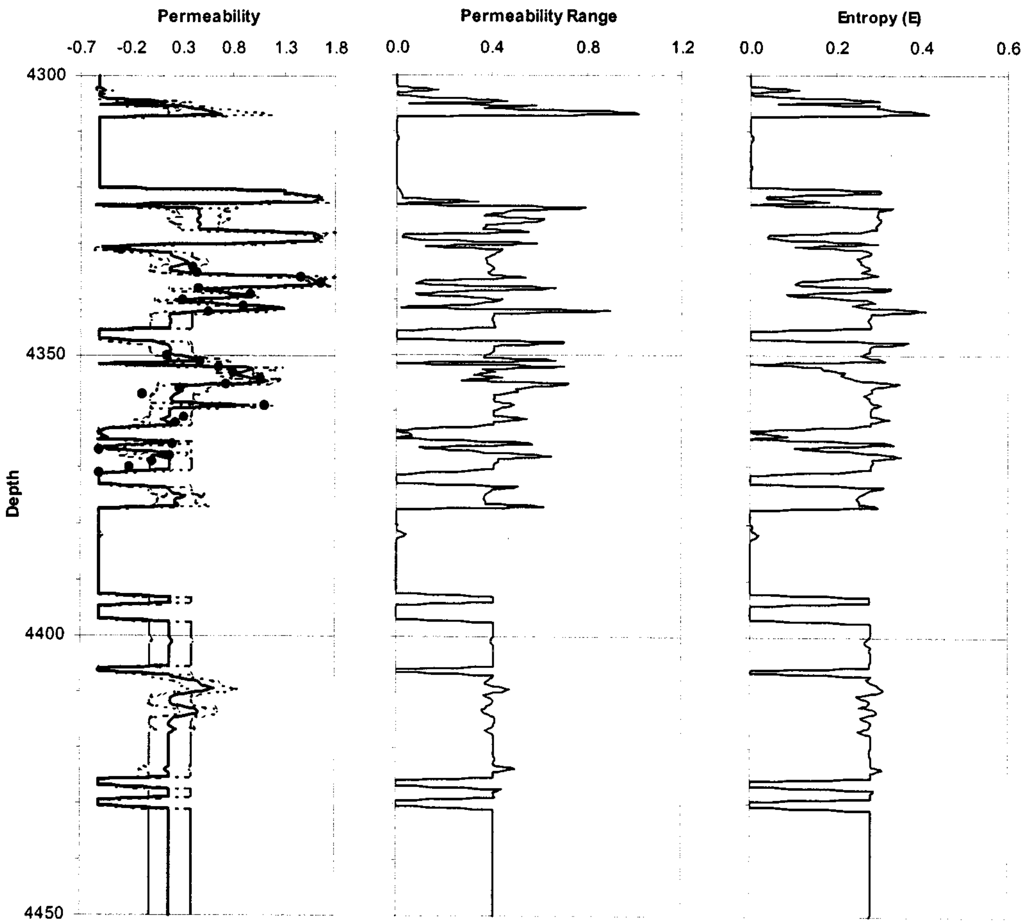

Fig. 6. Permeability and confidence logs.

and

$$\overline{\max} = \frac{\sum_{i}^{n} \hat{\mu}_i \max_i}{\sum_{i}^{n} \hat{\mu}_i}. \tag{5}$$

Note that, for symmetrical purposes, the "dummy" $\min_l$ and $\max_h$ values are required at both edges. These dummy points have no significant impact on the final $\min$–$\max$ values because the corresponding membership values would be small. The mid-point between the $\min$–$\max$ values can be taken as the final prediction, $\hat{y} = (\overline{\min} + \overline{\max})/2$. More vigorous defuzzification techniques can also be used to obtain the final prediction.

From the diagram, it is clear that the confidence bounds would be narrow if the E value were small. For example, if $\hat{\mu}_l \approx 0$, $\hat{\mu}_m \approx 1$, $\hat{\mu}_h \approx 0$, $E \approx 0$ and $\overline{\min} \approx \overline{\max} \approx \hat{y}$. This becomes a very confident prediction. On the other hand, if $\hat{\mu}_l = \hat{\mu}_m = \hat{\mu}_h = 1/3$, $E = E_{\max} = \log 3$ and the final $\min$–$\max$ values become the simple averages of all the $\min$–$\max$ predictions.

The basic uncertainty model presented in this paper was originated from the fuzzy membership functions derived by the geoscientist. Although they may be too subjective, the uncertainty model is more explicit and hence easier to refute than the conventional way of using implicit Gaussian-type model in the system.

It is important to note that this paper used both entropy and range of predictions to quantify the confidence of the predictions using the neural network output activations. They measure if the trained neural network is confident in the predictions. Thus they in fact measure only the relative (not absolute) confidence with respect to the quality and quantity of the training patterns. We therefore should be aware of the limitations and check if the results conform to the expert knowledge. If the training set is representative, the confidence bounds become useful for making practical decisions.

III. Field Examples

A. Permeability From Well Logs

The first case study used a data set from a well in an Indonesian reservoir [14]. The well contains 27 core permeability values from 4330 to 4370 feet. Four well logs are available: gamma ray (GR), density (RHOB), neutron (NPHI), and deep resistivity (RT). Fig. 3 displays the well logs. The objective of this case study is to develop a permeability transform from the four well logs and to generate a continuous permeability log for the well.

The target permeability (in logarithmic scale) was first transformed into four linguistic variables: "L" (low), "M" (medium), "H" (high) and "VH" (very high). Three equal intervals were used to design the cut-offs as shown in Fig. 4. The problem became a four-inputs–four-outputs problem. We used a standard backpropagation neural network with adaptive learning rate and momentum.

Since the data set was small, no blind testing was performed. We used all the 27 data input vectors to train the network. Six hidden neurons gave the smallest error on the training set. The results on the training set are shown in Fig. 5. A simple analysis shows that 93% of the (weighted) minimum predictions are lower than the actual values,

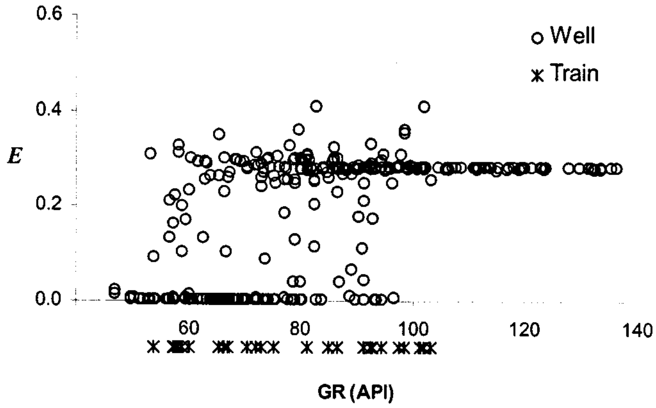

Fig. 7. Scatter-plot of $\boldsymbol{E}$ and $\boldsymbol{GR}$ values.

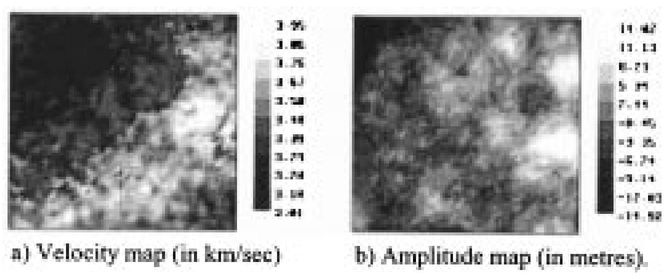

Fig. 8. Seismic attribute maps.

and 93% of the (weighted) maximum predictions are higher than the actual values. Also, 85% of the actual values fall within the confidence bounds. The mid-point of the $\min$–$\max$ predictions were taken as the final permeability (mean) predictions and the R-square was 0.93.

Fig. 6 shows the permeability log and the confidence bounds for the whole well. The permeability log matched well with the extreme values. Note that the permeability range (maximum minus minimum predictions) log shows a direct relationship with the entropy (E) log as expected. In fact, E is a useful extrapolation indicator. Fig. 7 shows a scatter-plot of E versus gamma ray (GR) for both the training set and the whole well data set. The points below the x-axis show the GR values of the training set. The plot shows that the E value becomes large when GR values of the well are beyond the range of the training set.

B. Porosity Mapping From Seismic Attributes

The data used came from a real reservoir with 294 wells [15]. A large-scale 2-D seismic survey was carried out and seismic velocity and amplitude data were obtained on a 70 × 70 grid system (Fig. 8). Average porosity was available at each well and the corresponding seismic attributes were obtained. The objective of this case study is to develop a porosity transform from the seismic attributes so that a 2-D porosity map can be generated for the whole area. We use the spatial coordinates (easting and northing) and the two seismic attributes as inputs and the target output is porosity. As in the first study, we transformed the porosity into the same four linguistic variables (Fig. 9).

Since the dataset is relatively large, we can improve the generalization performance of the neural network using early-stopping [13]. The 294 training data vectors were randomly divided into three data sets: 200 for training, 50 for optimizing the network parameters (e.g., no. of hidden neurons, stopping criteria) and 44 for blind testing. The optimum hidden size was 8. The results for the blind test samples are shown in Fig. 10. A simple analysis shows that 95% of the (weighted) minimum predictions are lower than the actual values, and 98% of the (weighted) maximum predictions are higher than the actual values. Also, 93% of the actual values fall within the confidence bounds. The mid-point of the $\min$–$\max$ predictions were taken as the final porosity (mean) predictions and the R-square was 0.91. The results were extremely encouraging. Fig. 11 shows the predicted porosity and confidence maps.

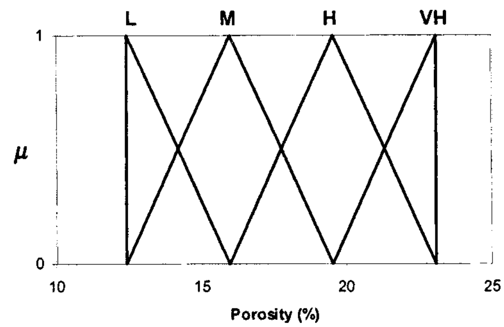

Fig. 9. Fuzzy transformation of porosity.

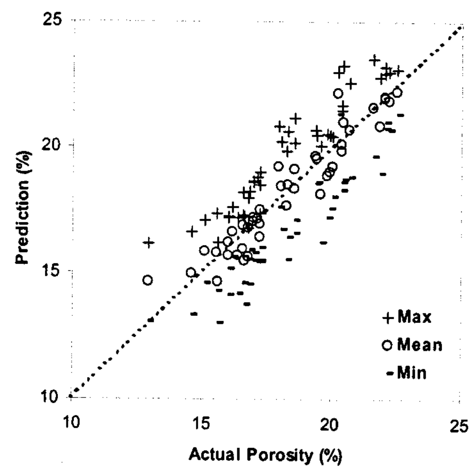

Fig. 10. Porosity prediction and confidence bounds for the 44 blind test samples.

IV. Conclusions

This paper presents a simple technique to estimate confidence bounds of predictions from conventional neural network regression models. The proposed methodology involves pre-processing of the target outputs by transforming the values into linguistic classes. A conventional learning algorithm is used and the predicted outputs are the membership function values. A simple post-processing of the outputs back-transforms the values into the original variable and the confidence bounds. From the two petroleum data sets (permeability from well logs and porosity mapping from seismic attributes) presented, the results show that the petrophysical predictions are realistic. The proposed methodology is general and useful, and can be implemented in simple spreadsheets without altering any existing neural network code.

As the design of the fuzzy membership functions is subjective, different experts may prefer different membership functions. The modeling of such vague decisions may provide further insights into the un-

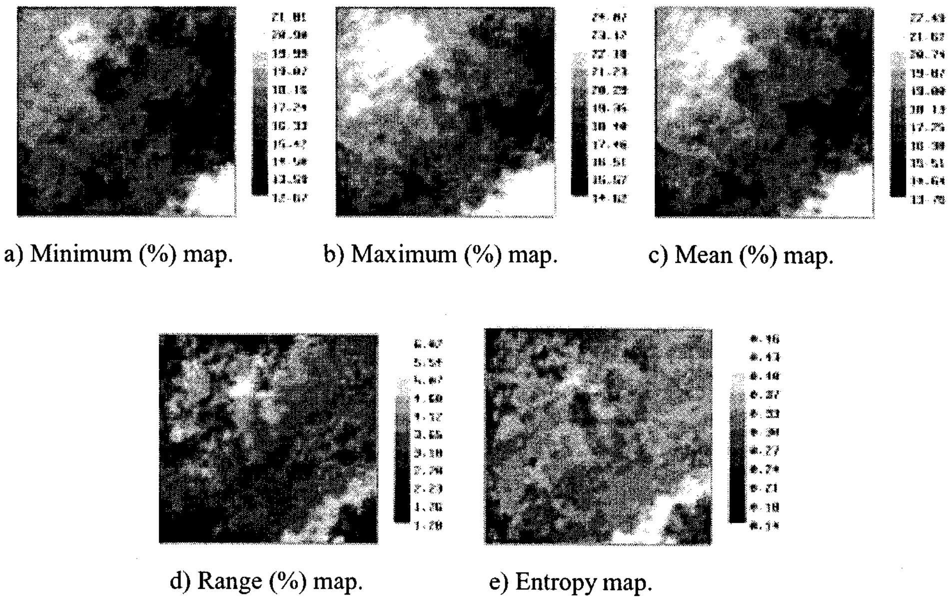

Fig. 11. Predicted porosity maps and confidence maps.

certainty of the petrophysical predictions. In the near future, we will examine the potential use of "rough" neurons [16] for implementing a range of membership values associated with a fuzzy segment such as multiple uncertainty models can be implemented.

References

[1] P. M. Wong, D. Tamhane, F. Aminzadeh, and M. Nikravesh, "Soft computing for intelligent reservoir characterization," in *Proc. SPE Asia Pacific Conf. Integr. Modeling Asset Manag.*, 2000.
[2] L. Brieman, "Bagging predictors," *Mach. Learn.*, vol. 24, pp. 123–140, 1994.
[3] J. G. Carney and P. Cunningham, "Confidence and prediction intervals for neural network ensembles," in *Proc. IJCNN'99 Conf.*, vol. 2, 1999, pp. 1215–1218.
[4] A. J. C. Sharkey, "On combining artificial neural networks," *Connect. Sci.*, vol. 8, pp. 299–314, 1996.
[5] P. M. Wong, M. Jang, S. Cho, and T. D. Gedeon, "Multiple permeability predictions using an observational learning algorithm," *Comput. Geosci.*, vol. 26, pp. 907–913, 2000.
[6] D. Mackay, "A practical Bayesian framework for backpropagation networks," *Neural Computat.*, vol. 4, pp. 448–472, 1992.
[7] G. Chryssolouris, M. Lee, and A. Ramsey, "Confidence interval prediction for neural network models," *IEEE Trans. Neural Networks*, vol. 7, pp. 229–232, Jan. 1996.
[8] D. Qu, A. G. Bruce, and P. M. Wong, "Bayesian neural networks: A new tool for permeability prediction," in *Advances in Intelligent Systems: Theory and Applications*, M. Mohammadian, Ed. Amsterdam, The Netherlands: IOS, 2000, pp. 1–6.
[9] L. Yang, T. Kalvi, M. Carlin, S. Clausen, and P. F. M. de Groot, "An evaluation of confidence bound estimation methods for neural networks," in *Proc. ESIT'00 Conf.*, 2000, pp. 322–329.
[10] L. Zadeh, "Fuzzy sets," *Inform. Contr.*, vol. 8, pp. 338–353, 1965.
[11] J. C. Bezdek, "Fuzzy models—What are they, and why?," *IEEE Trans. Fuzzy Syst.*, vol. 1, pp. 1–6, Feb. 1993.
[12] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagation errors," *Nature*, vol. 323, pp. 533–536, 1986.
[13] C. M. Bishop, *Neural Networks for Pattern Recognition*. New York: Oxford Univ. Press, 1995.
[14] D. Tamhane, P. M. Wong, and F. Aminzadeh, "Integrating linguistic descriptions and digital signals in petroleum reservoirs," *Int. J. Fuzzy Syst.*, vol. 4, pp. 586–591, 2002.
[15] P. M. Wong and S. A. R. Shibli, "Combining multiple seismic attributes with linguistic reservoir qualities for scenario-based reservoir modeling," in *Proc. SPE Asia Pacific Oil Gas Conf. Exh.*, 2000.
[16] P. Lingras, "Fuzzy-rough and rough-fuzzy serial combinations in neurocomputing," *Neurocomputation*, vol. 36, pp. 29–44, 2001.