Wong, K.W., Tikk, D., Biro, G. and Gedeon, T. (2002) Intelligent well log data analysis: a comparison study.
In: 1st International Conference on Fuzzy Systems and Knowledge Discovery, 18-22 November 2002, Singapore

INTELLIGENT WELL LOG DATA ANALYSIS: A COMPARISON STUDY

K.W. Wong, D. Tikk**, G. Biró*** and T.D. Gedeon**

* School of Information Technology
Murdoch University
South St, Murdoch
Western Australia 6155
Email: {kwong | tgedeon} @ murdoch.edu.au

** Department of Telecommunication & Telematics
Budapest University of Technology and Economics
H-1111 Budepest, Sztoczek u. 2, Hungary
Email: tikk@ttt.bme.hu

*** Department of Informatics
Eötvös Loránd Science University
H-1117 Budapest, Magyar Tudósok körútja 2., Hungary
E-mail: georgejr@alvilag.hu

ABSTRACT

In this paper we compare three different soft computing methods used as the well log data analysis model in petroleum engineering. Due to the diversely behaving nature, namely, that each region has a unique geophysical characteristic it is difficult to build a universal model relating the mathematical behaviour of the measured variables. This is the reason why soft computing techniques may be favourable to be applied in such case. We describe, investigate and compare a neural network, a fuzzy, and a neuro-fuzzy based method on a particular real data set.

1. INTRODUCTION

In petroleum reservoir modelling, boreholes are drilled at different locations around the region. Then well logging instruments are lowered into each borehole to collect data typically every 150mm or so of depth. These data are known in the industry as well log data. A very intense processing of this data is carried out in order to commence an evaluation of the reservoir's potential.

Well logging instruments used in the measurement of well log data fall broadly into three main categories: electrical, nuclear and acoustic [1]. Examples are Gamma Ray (GR), Resistivity (RT), Spontaneous Potential (SP), Neutron Density (NPHI) and Sonic interval transit time (DT). There are over fifty different types of logging tools available for different requirements. Physical rock samples from various depths are obtained by using a coring barrel to recover intact cylindrical samples of reservoir rock. These samples are then sent to a laboratory and examined using various physical and chemical processes. Data obtained from this phase are known as core data in the log analysis process. Although core data is the most accurate way of assessing the hydrocarbon of a well, they are very difficult and expensive to obtain. Means of providing good prediction of the petrophysical properties is necessary to avoid spending excessive amounts of money on coring. Therefore it is important to establish an accurate well log data analysis procedure to provide reliable information for the log analyst.

In well log analysis, the objective is to establish an accurate interpretation model for the prediction of petrophysical characteristics such as porosity, permeability and volume of clay for uncored depths and boreholes around the region [2,3]. Such information is essential to the determination of the economic viability of a particular well or reservoir to be explored.

A large number of techniques have been introduced in order to establish an adequate interpretation model over the past fifty years. Nevertheless, conventional derivation of a well log data analysis model normally falls into one of the two main approaches: empirical and statistical.

In the empirical approach, mathematical functions relating the desired permeability based on several well log data inspired by theoretical concepts are used [4,5]. This approach has long been favoured in the field and much effort has been made to understand the underlying petroleum engineering principles. However, the unique geophysical characteristic of each region prevents a single formula from being universally applicable.

Statistical techniques are viewed as more practical approaches [6,7]. The common statistical technique used is multiple regression analysis. The simplest form of

regression analysis is to find a relationship between the input logs and the petrophysical properties. The derived regression equations are then used for well log analysis. However, a number of initial assumptions of the model need to be made. Assumptions must also be made as to the statistical characteristics of the log data.

Over the past decade, another technique that has emerged as an option for well log analysis is the Artificial Neural Network (ANN). Research has shown that an ANN can provide an alternative approach to well log analysis with improvement over the traditional methods [8,9,10]. Most of the ANN based well log analysis models have used the Multi-layer Neural Network (MLNN) utilising the backpropagation learning algorithm. Such networks are commonly known as Backpropagation Neural Networks (BPNNs). A BPNN is suited to this application, as it resembles the characteristics of regression analysis in statistical approaches.

Fuzzy Logic (FL) that is capable to express the underlying characteristics of a system in human understandable rules is also used. A fuzzy set allows for the degree of membership of an item in a set to be any real number between 0 and 1. This allows human observations, expressions and expertise to be modelled more closely. Once the fuzzy sets have been defined, it is possible to use them in constructing rules for fuzzy expert systems and in performing fuzzy inference.

This approach seems to be suitable to well log analysis as it allows the incorporation of intelligent and human knowledge to deal with each individual case. However, the extraction of fuzzy rules from the data can be difficult for analysts with little experience. This could be a major drawback for use in well log analysis. If a fuzzy rule extraction technique is made available, then fuzzy systems can still be used for well log analysis [11,12].

With the emergence of intelligent techniques that combine ANN and fuzzy together have been applied successfully in well log analysis [13]. These techniques used in building the well log analysis model normally address the disadvantages encountered in ANN and fuzzy system.

The purpose of this paper is to conduct a comparison study of the results generated from an ANN, fuzzy and neuro-fuzzy technique. This paper gives a fair summary of the advantages and disadvantages of each type of intelligent well log data analysis model.

2. NEURAL NETWORK DATA ANALYSIS MODEL

Backpropagation Neural Network (BPNN) is the most widely used neural network system and the most well known supervised learning technique [14]. Back propagation is a systematic method for training multilayer ANN. It has been implemented and applied successfully to various problems. A basic BPNN consists of an input, an output and one or more hidden layers. Each layer is made up of a number of neurons that are connected to all the neurons in the next layers. However, the output layer will only generate the results of the network.

The objective of training BPNN is to adjust the weights so that application of a set of inputs will produce the desired set of outputs. A training set containing a number of desired input and output pairs is used. The input set is presented to the input layer of BPNN. A calculation is carried out to obtain the output set by proceeding from the input layer to the output layer. After this stage, feed forward propagation is done. At the output, the total error (the sum of the squares of the errors on each output cell) is calculated and then back propagated through the network. The total error, E, can be calculated using:

$$E = \sum_{k=1}^{K} \left(\frac{1}{2} \sum_{i=1}^{N_L} [T_i(k) - O_i^L(k)]^2\right) \qquad (1)$$

where K is the number of patterns, L is the layer number, T is the expect target, and O is the actual output

A modification of each connection weight is done and new total error is calculated. This back-propagated process is repeated until the total error value is below some particular threshold. At this stage, the network is considered trained. After the BPNN has been trained, it can then be applied to predict other cases.

As the most important factor of using BPNN is the ability to generalise, validation techniques used in [10] and [15] are used to ensure the generalisation capability of the well log analysis model.

3. SUGENO AND YASUKAWA'S MODEL

The goal of the Sugeno–Yasukawa (SY) fuzzy modelling method [16] is to create a transparent, viz. linguistic interpretable fuzzy rule based model from input–output sample data. Here we describe the original method with some modification proposed in [17]. These modifications concern approximation of membership functions, rule creation from sample data points, and selection of important variables. The construction of the rule base is performed in two main steps: the *identification* and the build-up of the *qualitative model*. The former can be further divided into two tasks: the structure identification and parameter identification. Having an identified model at hand, linguistic labels can be assigned to the finalized fuzzy sets in the rules in the qualitative modelling phase. In this paper we focus solely on the identification step.

Table 1: Classification of identification

Structure identification I	a: input candidates b: input variables
Structure identification II	a: number of rules b: partition of the input space
Parameter identification	

In [16] the authors classified the structure identification task into two types. The type I structure identification consists of finding the input candidates of

the system and finding its actual variables that affect the output. The type II structure identification covers the determination of the number of rules and the partition of the (usually) multidimensional input space. The identification task is summarized in Table 1.

3.1 Identification of Input Variables

The structure identification of type Ib concerns the selection of input variables that influence truly the output. This means that one has to choose a set of effective variables among a finite set of original variables. In [16], they used the regularity criterion (RC) method [18], but in [17] the authors proposed another, more reliable method for variable selection [19]. RC is a heuristic method that selects a set of inputs among the possible candidates (see more details in Section 4). It is performed between identification of type II and parameter identification steps. The outcome of RC depends on the identification of type II (see also Figure 1). In [17] it is shown that RC heavily depends on the implementation details of the method such as, e.g., the approximation of membership function and its parameters of RC itself, therefore they suggest another method that is based on the interclass separability criterion, and is performed once, before identification of type II.

3.2 Determination of the Number of Rules and the Input Partition

Usually in the design of a fuzzy system the rule antecedents and the partition of the input domain are determined. This (dense) rule base design methodology results in exponentiality in terms of the number of rules. To avoid this significant drawback, the SY method proceeds oppositely: first the partition of the output space is determined, which is done by clustering the whole output data set by the fuzzy c-means clustering (FCMC) [20]. The optimal number of clusters is determined by means of the following criterion [21]:

$$S(C) = \sum_{k=1}^{N} \sum_{i=1}^{C} (m_{ik})^m \left(\left\| y^k - v_i \right\|^2 - \left\| v_i - \bar{y} \right\|^2 \right) \qquad (2)$$

where N is the number of data to be clustered; C is the number of clusters, $C \geq 2$; y_k is the kth $\bar{y}$ is the average of data y_k; v_i is the centre of the ith cluster (vector); m_{ik} is the membership degree of the kth datum with respect to the ith cluster; m is the fuzzy exponent, $m > 1$.

As a result of the clustering, every output datum is associated with a membership degree in all the clusters B_i, i=1,…,C. From an output fuzzy clusters B_i we can induce a fuzzy cluster A_i in the multi-dimensional input space. This cluster can be projected onto the axis of variables; hence defining the antecedent fuzzy sets in each input dimension. Starting from a cluster B_i, and assuming that we have two input variables x_1 and x_2, we usually obtain a rule like

If x_1 is A_{i1} and x_2 is A_{i2} then y is B_i

We remark that although this notation implies that the number of rules is identical with the number of output clusters, it can happen that this is not the case.

In the original paper [16] it is proposed, in general, to approximate the (non-convex) input clusters with trapezoidal membership functions. On the other hand they remarked that more than one input cluster could belong to an output cluster. As a solution they suggest to ``form carefully two convex fuzzy clusters" in the input space. Although, these details seem to be not very important from the methodology aspect of the SY fuzzy modelling, they affect significantly the performance of the model being constructed; therefore in our comparison study we implemented the detailed algorithms presented in [17].

3.3 Parameter Identification

Parameter identification step can be accomplished in two stages in the fuzzy model design. In [16], it proposed to repeat it in every input candidate evaluation step, but this is mostly superfluous and time consuming. Performing it may be enough after the important input variables have been identified. At this stage we have to measure the performance of the rough fuzzy model. For this purpose the following performance index (PI) is used:

$$PI = \sum_{k=1}^{N} \left(y_k - \hat{y}_k \right)^2 / N \,, \qquad (3)$$

where $\hat{y}_k$ is the model output for the kth sample datum. In the case of fuzzy model, the parameters are those of the membership functions. Having trapezoidal membership functions it means 4 parameters for each antecedent $p_1 \leq p_2 \leq p_3 \leq p_4$. In the parameter identification step they adjusted these four values in an iterative algorithm. 5\% of the width of the actual input space was applied as adjusting value. In [17], the authors proposed an adaptive adjusting value, which solution improves somewhat the performance of the method. The flowchart on Figure 1 shows the overall design of the identification steps of the SY modelling.

4. SY WITH NEURAL NETWORK

This method similar to Sugeno and Yasukawa's model except that it uses backpropagation neural networks for inference engine not fuzzy systems. The set of inputs is selected the same way as the SY method with RC method. The input–output observations are divided into two sets. Two BPNN learns these two sets of input–output observations with respect to the set of selected inputs. The method measures the quality of the inference engines with a cross probe as

$$RC = \left[\sum_{k=1}^{|A|} \left(y_k^A - y_k^{AB} \right)^2 / |A| + \sum_{k=1}^{|B|} \left(y_k^B - y_k^{BA} \right)^2 / |B| \right] / 2 \,, \qquad (4)$$

where $|\cdot|$ denotes the size of a data group, superscript A and B mean that the corresponding value concerns data

group *A*, resp. *B*; and finally y_k^{AB} (y_k^{BA}) is the model output for the group *A* (*B*) input estimated by the model identified using group *B* (*A*) data. It decides that different set of inputs is needed or not. At the end, BPNN learns the full input–output observation set w.r.t. the set of selected inputs.

5.CASE STUDY AND DISCUSSIONS

A typical well consisting of a set of 127 data has been used for the case study. In this study, only three output rock matrices are used to in the comparison. The rock matrices are sandstone (MAT-1), limestone (MAT-2) and dolomite (MAT-3). The input logs used in this work are bulk density (RHOB), neutron (NPHI), uninvaded zone resistivity (RT), gamma ray (GR), caliper (CALI), photoelectric (PEF), invaded zone resistivity (RXO), sonic travel time (DT) and spontaneous potential (SP).

The best BPNN configuration chosen for this case consists of 9 input neurons, 18 hidden neurons and 3 output neurons. 72 of the available data sets are used for training and 54 are used for validation.

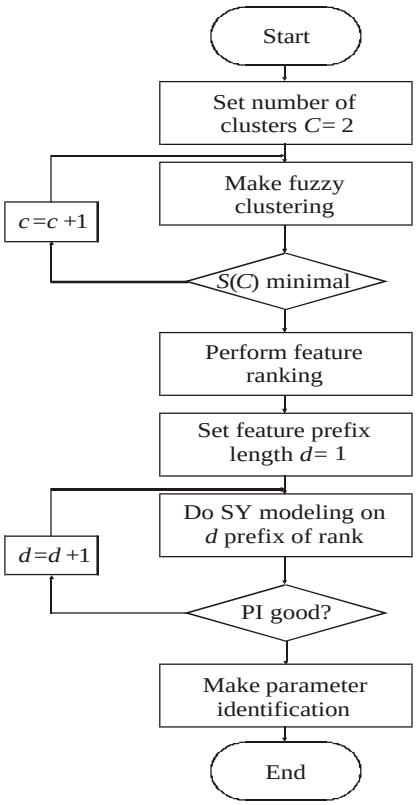

Figure 1. Flowchart of SY method [17]

We performed experiments with different settings of the SY model. We applied various methods to approximate trapeze shaped membership functions, and different feature selection algorithms. Because this technique is able to build MISO systems, we have to construct three different models for the 3 estimated output variables. As a consequence, it happens that the selected true inputs differ in the 3 models. To alleviate this drawback we can fix a certain set of variables as true inputs for all the three output variables. For MAT-1 we obtained the best result with moderate slope trapeze approximation (see details in [17]) and with setting all inputs as true ones. For MAT-2 it was with steep slop trapeze approximation and true inputs SP and PEF. For MAT-3 it was with moderate slope trapeze approximation and with true inputs PEF, RXO and RT.

SY with neural network model and regularity criterion input selection algorithm found RXO, PET, NPHI; RXO, SP, GR; and GR, respectively, the true inputs for each output variable (MAT-1, MAT2, and MAT-3). The network configuration was 10-10-3. The weight parameters of the final network were obtained after 25 learning iterations.

The prediction results as shown in Table 2 could show that BPNN can give the best prediction results among the three. The main disadvantages of using BPNN found in this comparison study is that the parameters and the configuration of the BPNN is very difficult to realise and is time consuming. Although literature has given some formulas in determining the parameters and configuration, it is observe that it is not true for this case study. Thus trial-an-error is used. In order to obtain an accurate prediction model, a number of configurations and parameters have been trial to obtain the model presented.

After a BPNN is trained, it acts like a “black box” with only weight connections between the nodes. Unlike an empirical expression with limited terms and coefficients, the analyst would have difficulty in understanding the vast number of weights involved and how the network performs a task. In addition, if some weights of the BPNN are modified, the effects on the output are unpredictable.

The SY model shows quite good performance but works considerable slower than SY with NN. The required time for computation is typically between 2 and 10 minutes depending on the trapeze approximation method and number of selected true inputs. The steeper a trapeze approximation version is the less space is left for membership function modification and improvements. Obviously, increasing number of input variables requires longer computational time when parameter identification is performed. The number of rules is 4, 7, and 3, respectively for each output variable. An obvious advantage of this technique is that it offers an insight into the model. Membership functions are easy to interpret and understand for users.

We remark here that classical BPNN network is unable to select the true inputs. When dummy variables are present this may influence considerably the performance of the method. When BPNN with fixed network configuration is combined with the SY input selection methodology, this drawback is alleviated. We note that this technique is the fastest among the three. The calculation time is about 30 seconds.

Table 2: Mean Square Errors (MSE)

Intelligent Techniques	MSE (MAT-1)	MSE (MAT-2)	MSE (MAT-3)
BPNN	0.223	0.235	0.172
SY model	0.425	2.01	1.39
SYNN	0.86	3.61	1.94

6. CONCLUSION

In this paper, we have compared the three intelligent techniques used in well log data analysis. From the results, BPNN produce the most accurate prediction. However, it does suffer some disadvantages as outlined in the previous section. The accuracy of the fuzzy rule extraction technique, SY Fuzzy modelling, and the neuro-fuzzy technique, SYNN could be improved by some parameters adjustment. This will be explored in the next paper. The future work on this is to integrate the results generated from these three intelligent techniques to produce a more accurate and sensible prediction that can be used confidently in the well log data analysis for petroleum engineering.

7. ACKNOWLEDGEMENT

This work was supported by the Australian Research Council, by the Hungarian Scientific Research Foundation (OTKA) Grants No. D034614, T34212, and T 34233.

8. REFERENCES

[1] M. Rider, *The Geological Interpretation of Well Logs*, Second Edition, Whittles Publishing, 1996.

[2] E.R. Crain, *The Log Analysis Handbook Volume 1: Quantitative Log Analysis Methods*, Penn Well Publishing Company, 1986.

[3] G.B. Asquith and C.R. Gibson, *Basic Well Log Analysis for Geologists*, The American Association of Petroleum Geologists, 1982.

[4] M.R.J. Wyllie, and W.D. Rose, "Some Theoretical Considerations Related to the Quantitative Evaluation of the Physical Characteristics of Reservoir Rock from Electric Log Data," *Journal of Petroleum Technology*, vol. 189, pp. 105-110, 1950.

[5] S.P. Kapadia, and U. Menzie, "Determination of Permeability Variation Factor V from Log Analysis," *SPE Technical Report 14402*, 1985.

[6] W.A. Wendt, S. Sakurai, and P.H. Nelson, "Permeability Prediction from Well Log using Multiple Regression", in L.W. Lake, and H.B. Caroll, *Reservoir Characterization*, Academic Press, pp. 181-221, 1986.

[7] J.M. Hawkins, "Integrated Formation Evaluation with Regression Analysis," *Proceedings of Petroleum Computer Conference*, pp. 213-223, 1994.

[8] D.A. Osborne, "Neural Networks Provide More Accurate Reservoir Permeability", *Oil and Gas Journal*, 28, pp. 80-83, 1992.

[9] P.M. Wong, I.J. Taggart, and T.D. Gedeon, "Use of Neural Network Methods to Predict Porosity and Permeability of a Petroleum Reservoir," *AI Applications*, 9(2), pp. 27-37, 1995.

[10] C.C. Fung, and K.W. Wong, "Petrophysical Properties Interpretation Modelling: An Integrated Artificial Neural Network Approach," *International Journal of Systems Research and Information Science*, vol. 8, pp. 203 - 220, 1999.

[11] K.W. Wong, D. Myers, and C.C. Fung, "A Generalised Neural-Fuzzy Well Log Interpretation Model With A Reduced Rule Base", *The Sixth International Conference of Neural Information Processing ICONIP*, Perth, vol. 1, pp. 188 - 191, 1999.

[12] H. Kuo, T.D. Gedeon, and P.M. Wong, "A Clustering Assisted Method for Fuzzy Rule Extraction and Pattern Classification," *Proceedings of 6th International Conference on Neural Information Processing*, vol. 2, pp. 679-684, 1999.

[13] Y. Huang, T.D. Gedeon, and P.M. Wong, "An Integrated Neural-Fuzzy-Genetic-Algorithm Using Hyper-surface Membership Functions to Predict Permeability in Petroleum Reservoirs," *Engineering Applications of Artificial Intelligence*, 14(1), pp. 15-21, 2001.

[14] D.E. Rumelhart, G.E. Hinton, and R.J. Williams "Learning Internal Representation by Error Propagation" in *Parallel Distributed Processing*, vol. 1, Cambridge MA: MIT Press, pp. 318-362, 1986.

[15] K.W. Wong, C.C. Fung, and H. Eren, "A Study of the Use of Self Organising Map for Splitting Training and Validation Sets for Backpropagation Neural Network," *Proceedings of IEEE Region Ten Conference (TENCON) on Digital Signal Processing Applications*, 1996, pp. 157-162.

[16] M. Sugeno and T. Yasukawa, "A fuzzy logic based approach to qualitative modeling" *IEEE Trans. on Fuzzy System*s, vol. 1, no. 1, pp. 7–31, 1993.

[17] D. Tikk, Gy. Biró, T.D. Gedeon, L.T. Kóczy, and J.D. Yang, "Improvements and critique on Sugeno and Yasukawa's qualitative modeling" *IEEE Trans. on Fuzzy System*s, to appear in 2002.

[18] J. Ihara, "Group method of data handling towards a modeling of complex system IV" *Systems and Contro*l, vol. 24, pp. 158–168, 1980, (in Japanese)

[19] D. Tikk and T. D. Gedeon, "Feature ranking based on interclass separability for fuzzy control application" *Proc. of the Int. Conf. on Artificial Intelligence in Science and Technology (AISAT'2000*), V. Karri and M. Negnevitsky, Eds., Hobart, Tasmania, Australia, 2000, pp. 29–32.

[20] J.C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithm*s, Plenum Press, New York, 1981.

[21] Y. Fukuyama and M. Sugeno, "A new method of choosing the number of clusters for fuzzy c-means method" in *Proc. of the 5th Fuzzy System Symposiu*m, 1989, pp. 247–250, (in Japanese)