CIMSA 2005 – IEEE International Conference on
Computational Intelligence for Measurement Systems and Applications
Giardini Naxos, Italy, 20-22 July 2005

A Soft Computing-Based Method for the Identification of Best Practices, with Application in the Petroleum Industry

Shahab D. Mohaghegh[1], Razi Gaskari[1]
[1]West Virginia University

***Abstract** – This paper introduces a new and novel methodology for fully data-driven best practices identification and analysis based on soft computing techniques. Using this new methodology "best practices" in any operation (industrial or otherwise) can be identified and recommendations can be made on how to conduct the operation in future in order to accomplish the best results. Since this is a fully data-driven process, no managerial or other biases will enter the process and it allows the data to speak for itself. In order for this method to be applicable it is required that sufficient amount of data for the process under investigation be available. The methodology uses neural network to build a representative predictive model of the process, fuzzy logic to provide analysis of the existing practices and genetic algorithms to identify the major trends. A recommendation matrix at the end of the process is developed that would serve as the foundation for making best practices recommendation. Application of this methodology to a problem in the petroleum industry is presented.*

***Keywords** – Soft Computing, Best Practices, Data Driven Solution, Petroleum Industry.*

I. INTRODUCTION

Identification of best practices in many industrial operations is gaining unprecedented momentum. Companies that have gathered large amounts of data now realize that they own a valuable commodity that can play an important role in increasing efficiency in their day to day operations. The question is how this vast amount of data can be used in order to help the company's bottom-line. This paper attempts to address this question by introducing a newly developed methodology that enables companies to deduce information and knowledge from the existing data. The deduced information and knowledge can then be used in developing business rules and decision making.

These data collected and preserved by companies cover all aspects of their business, from purely technical data that includes certain measurements to non-technical data such as those related to economics or human resources issues. Now that all this data is available, following questions may be asked:

- "What can we do with this data?"
- "How can the company get a return on its data collection and preservation investment?"
- "Are there stories hidden in the megabytes, or sometime gigabytes of data?"
- "The collected data is a reflection of the history of the operations that have taken place and sometime are still taking place. What can we learn from our past practices?"

As the volume of data increases, human cognition is no longer capable of deciphering important information from it by conventional techniques. Data mining and machine learning techniques based on soft computing must be used in order to deduce information and knowledge from the raw data that resides in the databases. The Intelligent Best Practices Analysis – IBPA (see Figure 1 for a flow chart) that is introduced here incorporates the state of the art in data mining and machine learning to assist professionals in making the most of their existing data.

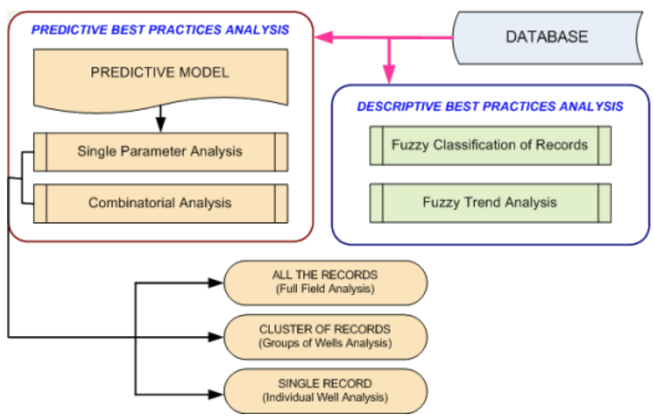

Figure 1. Intelligent Best Practices Flow Chart.

Inspired by state-of-the-art in data mining, Intelligent Best Practices Analysis is a combination of two sets of analysis, namely "Descriptive Analysis" and "Predictive Analysis". By relying on the resident data in the database, descriptive analysis can quickly identify and show any apparent patterns that exist. Predictive analysis, on the other hand, takes the problem to a new level by developing a comprehensive solution space and performing analysis based on the data residing on the database as well as the potential practices that could have been performed by interpolation between current and past practices. If the comprehensive solution space developed by the predictive analysis can be imagined as a terrain of hills and valleys in a hyper dimensional space, the existing practices (data in the database) would serve as discrete points on this terrain. The new solution space makes it possible to interpolate between existing practices and go into areas that have not yet actually been explored, but are logical extensions of the existing practices.

The Intelligent Best Practices Analysis starts by identifying a process outcome. The process outcome is a dependent variable in the database that is used in order to measure the degree of success of a particular practice (an objective function). For example "five-year cumulative production" would be a reasonable outcome for identifying the best practices in a petroleum production related process and "return on investment" would be an appropriate variable for identifying the best practices in a project evaluation process.

II. DESCRIPTIVE BEST PRACTICES ANALYSIS

Descriptive Best Practices Analysis tries to find and display patterns that exist in the database using a fuzzy averaging technique. It does not manipulate the data in any shape or form. It simply presents data in a new light that makes detection of existing trends and patterns possible. This is performed in two steps. First the process output (the dependent variable) is classified using fuzzy set theory [1]. For example when "five-year cumulative production" is used as the process outcome a well can be classified as poor, average or good (Figure 2). If "return on investment" is used as the process outcome the ROI of a project in the database can be classified as low, average or high. Upon completion of this step all the records in the database are classified using fuzzy membership functions.

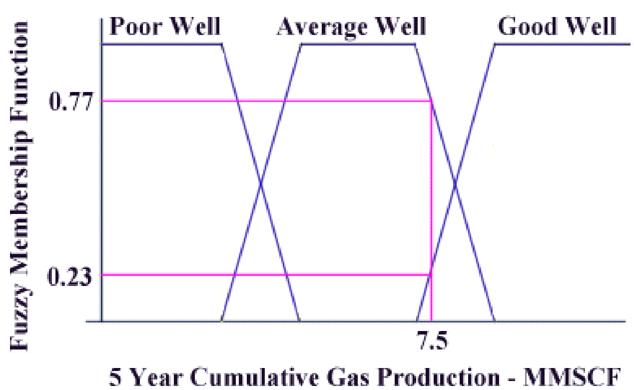

Figure 2. Fuzzy classification of a particular well (record) in the database.

Each of the parameters in the database (the independent variables) can be analyzed in order to see if a trend or pattern can be identified and then plotting them in the form of a bar chart. In Figure 3 while the average value of permeability for all of the wells in the database is about 0.09 md, poor wells have an average permeability of 0.04 md and good wells an average permeability of 0.33 md. This of course is an obvious trend and very well expected but for other parameters that are not necessarily so obvious, this kind of visualization of data can reveal important information.

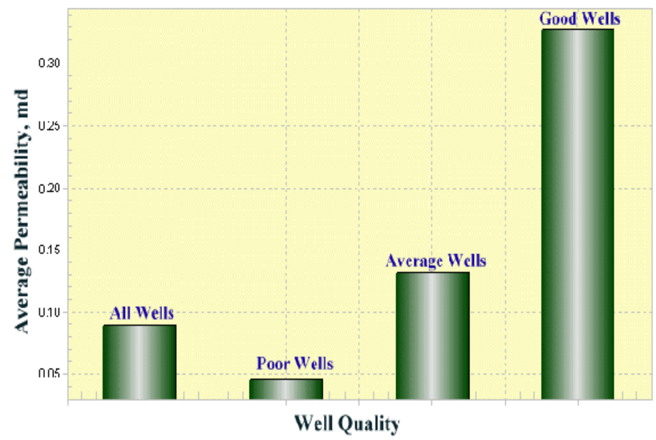

Figure 3. Fuzzy averaging of reservoir permeability for all the wells (records) in the database.

III. PREDICTIVE BEST PRACTICES ANALYSIS

The main idea behind the predictive analysis is to fill the gaps in the solution space in order to make a comprehensive analysis possible. As mentioned in the previous section the predictive model that is developed, calibrated and verified based on the existing data provides a continuous hyper - dimensional surface that is full of hills and valleys. This surface covers the entire solution space and the records in the database are discrete points on this surface. Goal of predictive analysis is two folds. First, to develop a predictive model that can accurately approximate this solution space. This goal is achieved using artificial neural networks [2]. The process outcome is used as the network's output and several independent variables are used as the network input. It is important to use the proper variables as the input to the neural network. The input variables have to sufficiently represent the process being modeled and include parameters that identify the uniqueness of each record in the database. Second goal of the multiple-parameter analysis is to exhaustively search through, and query the solution space in order to identify patterns that can be used as guides in the decision making process. This step of the analysis is performed using genetic algorithms [3]. The neural network model developed in the first step is used as the objective function of the genetic algorithms.

Once a predictive model is developed, single and multiple-parameter analysis is performed during which the effect of independent variables in the database on the process outcome is studied and recorded. From the recorded results of single and multiple-parameter analysis a recommendation matrix is generated. The recommendation matrix is then used in order to draw conclusions from the analysis and make operational recommendations pertaining to the best practices for the process being studied.

A. Single Parameter Analysis

Figure 4 demonstrates the sensitivity of process outcome as a function of one of one independent variable in a single-parameter analysis. The independent variable being studied in this figure is the relative reservoir quality index.

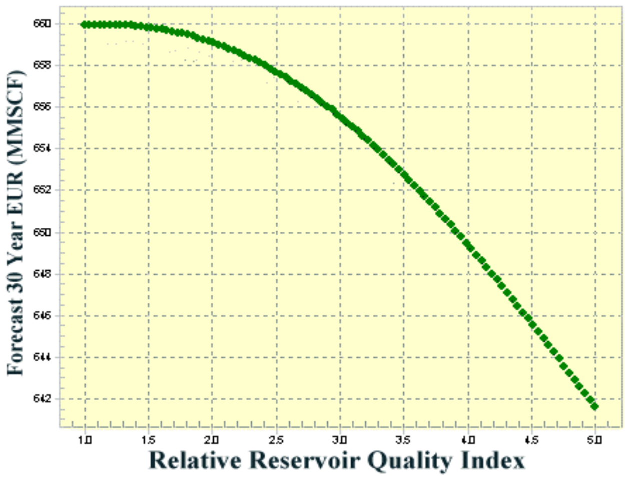

Figure 4. Sensitivity of outcome (dependent variable) to RRQI, an independent variable.

Performing single parameter analysis involves the following steps:

1. Select a parameter that is to be analyzed.
2. Identify its minimum and maximum value in the database.
3. Calculate the range of the parameter by subtracting the minimum from the maximum.
4. Calculate an increment by dividing the range calculated in step 3 by say, 100.
5. Starting from the minimum value for the parameter, run the predictive model 100 times, each time incrementing the value of the parameter by the increment calculated in step 4. Perform this step while all other parameters are held at their original value.

Upon completion of these steps we will end up with 100 runs of the model for each record in the database. If the values of the process outcome (model output) are plotted against the parameter being analyzed, the resulting plot will show the sensitivity of the process to a particular parameter as shown in Figure 4, above.

Similar analysis is performed on every record in the database. If the database has for example 1000 records, then there will be 1000 trends identified (one for each record) for a particular parameter. The shape of these trends then needs to be analyzed. For example if all or most of the record show the same trend (for example decreasing process outcome performance with increasing value of the parameter, as shown in Figure 4) then the conclusion for optimum value of this particular parameter can easily be made. The results of such trend analysis are recorded in the recommendation matrix that will be covered in the following sections.

B. Multiple-Parameter Analysis

Multiple-parameter analysis involves studying the sensitivity of process outcome to changes accruing in more than one parameter simultaneously. As the number of independent variables being studied in a multiple-parameter (combinatorial) analysis increase to more than two parameters, we no longer can picture the resulting solution surface. Therefore, genetic optimization is performed in order to capture the optimum values of multiple-parameters for this step. The question may arise that why an optimization routine is necessary at this point of the analysis. The reason is that we are looking for the "Best Practices" which in essence is the ultimate outcome of a process optimization procedure. Best Practices by definition are those that provide the best (optimum) results.

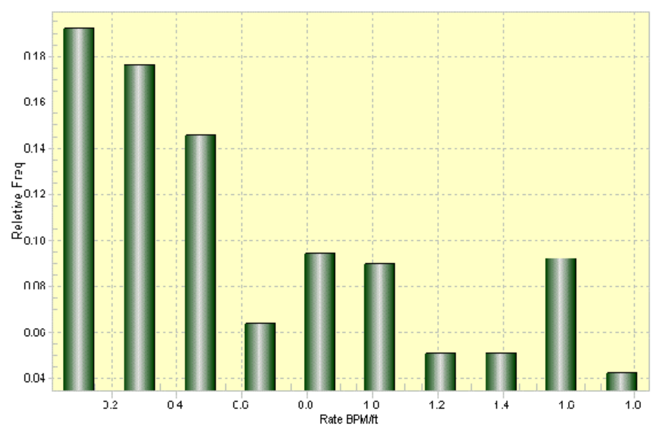

Figure 5. Frequency distribution of injection rate (an independent variable) after optimization is completed.

Figure 5 is an example of such an analysis. Multiple parameters such as amount of fluid being injected, type of injected fluid, amount of sand being injected, and injection rate were among the independent parameters that were involved in this analysis. In this figure it can be seen that from a combinatorial analysis point of view, the best practice for this particular field points toward lower injection rates of about 0.1 to 0.5 barrels of fluid per minute per foot of formation. Similar bar graphs can be generated for each of the parameters involved in the multiple-parameter analysis. Finally the result of the single parameter and combinatorial analysis is presented in form of a recommendation matrix as shown in Figure 6. The recommendation matrix summarizes the results and provides conclusions in the form of the “Best Practices”.

In order to perform multiple-parameter or combinatorial analysis these steps must be followed:

1. Identify the parameters that are going to be optimized. It is advised that the parameters have the following two characteristics:
    a. Be controllable. This simply means that the parameters should be those that can be controlled. For example when identifying best practices for hydraulic fracturing in oil or gas fields, porosity of the reservoir is not a controllable parameter while the type of fluid being used for the job is a controllable parameter.
    b. Be independent. If there are dependencies between parameters that are being optimized, then one must define those dependencies and code them into the optimization process, otherwise the result might not be meaningful.
2. Define any rules or constraints that should be imposed on the solutions while performing the optimization. As an example, a rule may state that only one of several available fluids can be used as the injection fluid and not a combination of several fluids.
3. Identify the number of population in each generation (number of solutions that will be used during the evolutionary process) and the number of generations that the solutions will be evolved in order to reach optimization.
4. Probability of genetic operations such as crossover, mutation, and inversion. These probability values will control the way each new generation is produced as a function of success and failure of the previous generation.
5. Identify convergence criterion. For example one convergence criterion can be that the best individual in the generation is not improving over several generations.
6. Perform genetic optimization.
7. Keep track of all the solutions in each generation. Select and save a certain number of top solutions from each generation.
8. Upon convergence, perform the following for each of the parameters that were involved in the optimization process:
    a. Use the top, say 10, solutions of each generation and record the values of the parameter being analyzed from each of these top solutions.
    b. Do the above step for all the generations.
    c. Do steps (a), and (b), for all the records (wells) in the database.
    d. Plot the frequency distribution for the values you have recorded for the parameter being analyzed.
    e. Analyze the trends that appears from the frequency distribution.

The process mentioned above uses the predictive neural model in order to identify the most appropriate value for each parameter that would result in optimum outcome and presents it in a graphical form in order to detect any apparent trends as the best practices.

C. Recommendation Matrix

Results of single parameter analysis and multiple-parameter (combinatorial) analysis are summarized in order to be used in the final stage of the analysis that is the Recommendation Matrix. The summarized form of the results of the analyses in the form of a recommendation matrix will provide information on the nature of the distribution of the parameters that would result in the best practices in a particular process. For example the result of the Figure 5 is summarized by tow simple identifiers. The distribution is "skewed" and it is skewed toward the lower values, which suggest that the lower injection rates are preferred.

Figure 6 shows an example of a typical recommendation matrix. In this particular recommendation matrix the type of fracturing fluid being used is being analyzed. The three possible fracturing fluids used in this operation (for an oil field in mid-continent United States) are water, oil and acid. The recommendation matrix includes three sections. In the first section the results of single parameter analysis is recorded. Three parameters in the first section are identified as percent of population, dominant trend and change in value. For example in the example shown in Figure 6 oil and acid are similar in that all the records show a moderate increase in process outcome (5 year cumulative oil production) when these two fluids are used. Water on the other hand shows different results. When water is used as the main fracturing fluid majority of the records (wells) show a small decrease in the production.

Parameter	Single Parameter Analysis			Combinatorial Analysis		Recommendations
	Percent Population	Dominant Trend	Change in Value	Distribution	Dominant Trend	
Water	Majority	Decreasing	Low	Skewed	Low Values	Use Not Recommended
Oil	All	Increasing	Moderate	Skewed	High Values	Use Recommended
Acid	All	Increasing	Moderate	Skewed	Low Values	INCONCLUSIVE

Figure 6. The Best Practices recommendation matrix.

As far as the combinatorial analysis are concerned, results from all the three fluids are skewed with only the oil being skewed to the higher values. Therefore, the recommendation will be to use oil as the fracturing fluid in this field, since oil is the only fluid that provides the same result for both single and combinatorial analysis. The recommendations can be either firm or inconclusive. For example if both single and combinatorial analysis point to the same direction then a conclusive recommendation can be made as the case is for water and oil in Figure 6. On the other hand if single and combinatorial analysis point to different directions, as is the case for acid in Figure 6, then the result would be inconclusive. Figure 7 shows the meaning of the terms that are used in the recommendation matrix.

SINGLE PARAMETER ANALYSIS	
PERCENT OF POPULATION	**MEANING**
ALL	More than 95% of records (wells) behave in a certain fasion
MAJORITY	More than 60% of records (wells) behave in a certain fasion
HALF & HALF	Between 45% to 55% of records (wells) behave in a certain fasion
DOMINANT TREND	
INCREASING	Use of this parameter causes an increase in process (model) outcome
DECREASING	Use of this parameter causes an decrease in process (model) outcome
MIX	Process (model) outcome is mixed (increase & decrease) for different records (wells)
CHANGE IN VALUE	
HIGH	The amount of increase in the process (model) outcome is high
LOW	The amount of increase in the process (model) outcome is low
MODERATE	The amount of increase in the process (model) outcome is moderate
COMBINATORIAL ANALYSIS	
DISTRIBUTION	
SKEWED	Probability Distribution Function for this parameter is skewed
NORMAL	Probability Distribution Function for this parameter is normal
UNIFORM	Probability Distribution Function for this parameter is uniform
DOMINANT TREND	
HIGH VALUES	Probability Distribution Function is skewed toward the high end of this parameter
AVERAGE VALUES	Probability Distribution Function has a normal behavior
LOW VALUES	Probability Distribution Function is skewed toward the low end of this parameter
NO TREND	The uniform Probability Distribution Function provides no trends for this parameter
RECOMMENDATIONS	
USE NOT RECOMMENDED	Try to avoid using this parameter
USE RECOMMENDED	Try using this parameter
INCONCLUSIVE	No recommendations can be made at this point for this parameter
USE LARGE VALUES	Higher values of this parameter is recommended
USE AVERAGE VALUES	Average values of this parameter is recommended
USE LOW VALUES	Lower values of this parameter is recommended

Figure 7**.** The key for the terms used in the recommendation matrix.

IV. A COMPLETE SET OF ANALYSES

In order to perform a complete set of Intelligent Best Practices Analysis, the process that was covered in the last three sections, namely Single Parameter Analysis, Combinatorial Analysis and finally the Recommendation Matrix is performed more than one time. First the process is performed for all of the records in the database as covered in the previous sections.

Upon completion of the analysis for the entire database the records in the database are classified and the process is repeated for each class (category or cluster) of the records. The classification can be based on several criteria. If there is a class indicator in the database it can be used as the basis for the classification. For example several companies may be operating in the same reservoir and the database may have a column that identifies the operator of the well that can be used as a natural (native) classifier. In the case where each record in the database represents a well in a field the classification of the records (wells) can be based on any predetermined classifier such as well quality, operating company, well locations, geology, reservoirs involved, or any other classification that makes sense. In cases that such native (natural) classifier is not present in the database, k-mean clustering or fuzzy c-mean clustering can be used in order to cluster the records in the database.

Once the classification of the records is completed, the best practices analysis is performed on each group of the records and the result are tabulated in recommendation matrices as shown in the previous sections. At the conclusion of this step there will be as many recommendation matrices as there are clusters (groups, classes or categories) in the database. Performance of this step is highly recommended since it has been observed that in cases where inconclusive results have been reached with all the records, more conclusive results can be identified for subsets of the records, and therefore more concrete recommendations can be made.

The last part of the analysis would be performed on individual records. Similar to the previous sections, single parameter analysis and combinatorial analysis can be performed for one record at a time. Based on the nature of the database and the type of process it describe this step can be important or redundant. In cases where an individual record describes an independent entity (such as a well in an oil or gas field) performing individual analysis can be very important and result in valuable information.

The relationship between these analyses is shown in Figure 8. This figure shows an inverted pyramid that increases in precision and decreases in averaging as one moves from the top of the pyramid (analyzing all the records in the database or full field analysis in the case oil industry) to cluster of records (groups of wells) and finally to the single record (individual wells).

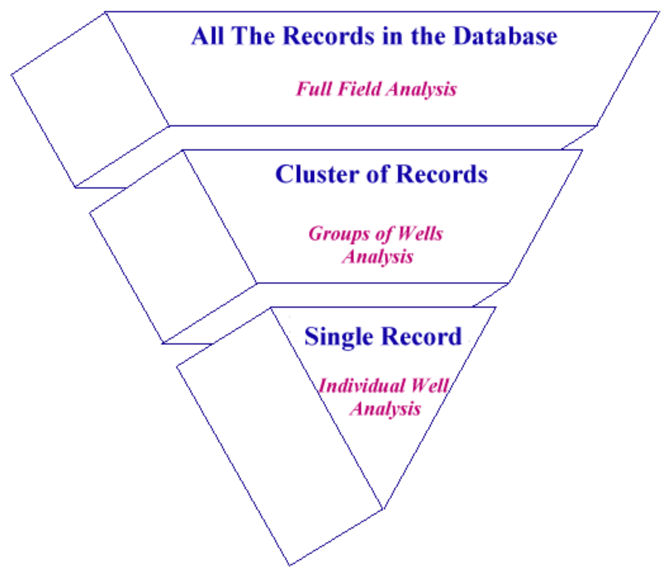

Figure 8. The relationship between different levels of the Intelligent Best Practices Analysis.

V. CONCLUSIONS

A new methodology has been developed and introduced for identification of best practices in any process that can be described by a data set. The new methodology is based on soft computing techniques such as artificial neural networks, genetic algorithms and fuzzy logic and has been named Intelligent Best Practices Analysis. Intelligent Best Practices Analysis (IBPA) is a two-step process that includes a descriptive and a predictive analysis. During the descriptive analysis fuzzy-averaging of parameters is performed in order to identify the trends that are present in the database. These trends and patterns usually provide a strong foundation for the best practices that are ultimately identified.

Predictive best practices analysis is a drill-down process that starts with all the records in the database and ends with individual records. During the complete database analysis the best practices are identified based on all the records in the database. The database is then divided into groups, based on different criteria, and each group is analyzed separately in order to either verify, refine or dispute the best practices that were identified during complete database analysis.

The last step of the predictive analysis is working with individual records and using the results of past two steps as a guide to enhance our understanding of each individual record. It has been observed that although the best practices that were identified during the complete database analysis and groups of records analysis would work for most of the records, there will be records in the database that would not necessarily follow the identified trends as expected. This makes analysis of individual records a highly recommended exercise.

This process was successfully applied to several oil filed related databases. For example when applied to hydraulic fracturing practices in an oil field in the mid-continent of the United States it identified following recommendations for the future hydraulic fracturing procedures in the aforementioned oil field:

- The recommended main fracturing fluid for the clastic formations in this field is “Diesel Oil”.
- The carbonate formations that produce mainly gas in this field respond positively to acid as the main fracturing fluids.
- In the carbonate formations “acid fracs” and not “acid jobs” are the stimulations that should be performed.
- A relatively low number of perforations (may be less than or equal to one shot per foot of pay thickness) would be the most appropriate practice of completion for wells in this field. This seems to be true for both clastic and carbonate formations.
- Higher proppant concentrations work better in this field. The recommendation is to use proppant concentrations of higher than 1.0 lbs/gal/ft.
- It is recommended to use average injection rates of less than or equal to 0.2 barrels of fluid per minute per foot of pay thickness while stimulating the formations in this filed. The combination of above three parameters, namely injecting higher proppant concentrations at lower injection rates into smaller numbers of perforations is targeted at avoiding increasing of the bottom-hole pressure during the treatment. While higher proppant concentrations provide a better conduit for the fluid flow and stronger support for keeping the fracture open for longer periods of time, its combination with lower numbers of perforations may contribute to higher bottom-hole pressures that might impede short and long term production. By injecting the treatment at lower injection rates we will try to keep the bottom-hole pressure low. It has been shown that there is a correlation between low bottom-hole treating pressures with higher production indicators [4].

REFERENCES

[1] Mohaghegh, S. D., ‘Virtual Intelligence Applications in Petroleum Engineering: Part 3 – Fuzzy Logic’. *Journal of Petroleum Technology*, Distinguished Author Series, November 2000, pp 82-87.

[2] Mohaghegh, S. D., ‘Virtual Intelligence Applications in Petroleum Engineering: Part 1 – Artificial Neural Networks. *Journal of Petroleum Technology*, Distinguished Author Series, September 2000, pp 64-73.

[3] Virtual Intelligence Applications in Petroleum Engineering: Part 2 – Evolutionary Computing. *Journal of Petroleum Technology*, Distinguished Author Series, October 2000, pp 40-46.

[4] Johnson, R. L., Dunn, K. P. Hopkins, C. W. and Conway, M. W. “The Pressure Dependence Ratio: A Bottomhole Treating Pressure Diagnostic Tool for Hydraulic Fracturing in Tight, Naturally Fractured Reservoirs,” SPE 77916. Presented at SPE Asia Pacific Oil & Gas Conference & Exhibition held in Melbourne, Australia, 8-10 October 2002.