

A least-square-driven functional networks type-2 fuzzy logic hybrid model for efficient petroleum reservoir properties prediction

Fatai Anifowose · Jane Labadin · Abdulazeez Abdulraheem

Received: 31 January 2012 / Accepted: 7 December 2012 / Published online: 18 January 2013
© Springer-Verlag London 2013

Abstract Various computational intelligence techniques have been used in the prediction of petroleum reservoir properties. However, each of them has its limitations depending on different conditions such as data size and dimensionality. Hybrid computational intelligence has been introduced as a new paradigm to complement the weaknesses of one technique with the strengths of another or others. This paper presents a computational intelligence hybrid model to overcome some of the limitations of the standalone type-2 fuzzy logic system (T2FLS) model by using a least-square-fitting-based model selection algorithm to reduce the dimensionality of the input data while selecting the best variables. This novel feature selection procedure resulted in the improvement of the performance of T2FLS whose complexity is usually increased and performance degraded with increased dimensionality of input data. The iterative least-square-fitting algorithm part of functional networks (FN) and T2FLS techniques were combined in a hybrid manner to predict the porosity and permeability of North American and Middle Eastern oil and gas reservoirs. Training and testing the T2FLS block of the hybrid model with the best and dimensionally reduced input variables caused the hybrid model to perform better with higher correlation coefficients, lower root mean square errors, and less execution times than the standalone T2FLS model. This work has demonstrated the promising

capability of hybrid modelling and has given more insight into the possibility of more robust hybrid models with better functionality and capability indices.

Keywords Hybrid artificial intelligence · Functional networks · Type-2 fuzzy logic · Petroleum reservoir · Least-square-fitting algorithm

1 Introduction

Porosity and permeability are two important oil and gas reservoir properties which have to do with the amount of fluid contained in a reservoir and its ability to flow. These properties have significant impact on petroleum field operations and reservoir management [1]. They are usually measured in the laboratory on plugs extracted from the core of wells drilled for oil and gas exploration. The data are valuable for relating the two properties whose values are not directly measurable from well logs since they are purely natural phenomena. Both porosity and permeability data serve as standard indicators of reservoir quality in the oil and gas industry. This process, called petroleum reservoir characterization, is used for quantitatively describing various reservoir properties in spatial variability by using available field and laboratory data. The process plays a crucial role in modern reservoir management: making sound reservoir decisions and improving the reliability of the reservoir predictions. These properties make significant impacts on petroleum field operations and reservoir management [2].

Since the laboratory measurements are usually costly and time-consuming, artificial intelligence (AI) techniques have been successfully used in the prediction of these properties to an acceptable degree of accuracy [1, 3–5]. However, each AI technique has certain limitations that would not make

F. Anifowose (✉) · J. Labadin
Faculty of Computer Science and Information Technology,
Universiti Malaysia Sarawak, 94300 Kota Samarahan,
Sarawak, Malaysia
e-mail: fanifowose@gmail.com

A. Abdulraheem
Department of Petroleum Engineering, King Fahd University
of Petroleum and Minerals, Dhahran 31261, Saudi Arabia

their use desirable in certain conditions such as in small dataset scenarios [6, 7] and high dimensionality of data conditions [8, 9]. Some AI techniques such as type-2 fuzzy logic are sensitive to the dimensionality of data [9, 10] such that even a little reduction in the number of features is expected to make a significant improvement in their performance indices. Hence, combining such techniques with others that have the capability of addressing some of the limitations in the former, such as reducing the data dimensionality, would be a welcomed development. Such hybridization tasks will improve the accuracy of predictions, which would in turn, increase the confidence in the results and hence improve the overall efficiency of oil and gas exploration and production activities.

The hybridization of two or more AI techniques, known as hybrid computational intelligence (HCI), to create a single integrated model, is becoming increasingly popular. The increased popularity lies in their extensive success in real-world complex problems such as in the characterization of oil and gas reservoirs [1, 6], network intrusion detection [11], biometric face recognition [12], bioinformatics [13, 14], financial credit risk assessment [15], multimedia processing [16], and control systems [17].

In this work, the desirable qualities of LS-fitting algorithm of functional networks (FN) and type-2 fuzzy logic (T2FLS) were combined to predict two properties of oil and gas reservoirs, namely porosity and permeability, with better performance indicators. Our motivations for this work include the quest for higher performance accuracy in the prediction of oil and gas reservoir properties; the recently increasing popularity of hybrid intelligent systems; the reported success of these hybrid systems in many real-world complex problems; the need to complement the weaknesses of one algorithm with the strengths of the other and hence to combine the cooperative and competitive characteristics of the individual techniques; and the existing theoretical and experimental justifications that hybrids produce more accurate results than the individual techniques used separately [2, 6, 18].

The considerable number of applications of hybrid techniques indicate, on one hand, the keen interest of researchers in this new concept and, on the other hand, the need for better components that will yield hybrid models that are simpler in architecture and more efficient in performance.

The rest of this paper is organized as follows: Sect. 2 presents some background knowledge on the petroleum reservoir characterization process, porosity, and permeability. Section 3 reviews some of the previous works related to this study. Section 4 describes the datasets, experimental methodology, and the criteria for the performance evaluation of this work. Section 5 presents the results with a detailed discussion. Conclusion is presented in Sect. 6 while our plan for further research is presented in Sect. 7.

2 Background knowledge

2.1 Petroleum reservoir characterization

Oil and gas reservoir characterization plays a crucial role in modern reservoir management. It helps to make sound reservoir decisions and improves the asset value of the oil and gas companies. It maximizes the integration of multi-disciplinary data and knowledge, and hence improves the reliability of reservoir predictions. The ultimate goal is “a reservoir model with realistic tolerance for imprecision and uncertainty” [2]. Furthermore, it is the process between the discovery phase of a reservoir and its management phase. The process integrates the technical disciplines of petroleum reservoir engineering, geology, geophysics, oil and gas production engineering, petrophysics, economics, and data management. The key objectives of reservoir characterization focus on modelling each reservoir unit, predicting well behavior, understanding past reservoir performance, and forecasting future reservoir conditions.

The petroleum reservoir characterization process usually involves a well logging technique that makes measurements in oil and gas wells that have been drilled. Probes are designed and lowered in these wells to measure the physical and chemical properties of rocks as well as the fluids contained in them. Much information can be obtained from samples of rock brought to the surface in cores or bit cuttings, or from other clues while drilling, such as the rate of penetration, reservoir pressure, and bottom-hole temperature. However, the greatest amount of information comes from well logs which results from the probes. The resulting measurements, known as geophysical well logs, are recorded graphically or digitally as a function of depth.

Although the most commonly used logs are for the correlation of geological strata and location of hydrocarbon zones, there are many other important subsurface parameters that need to be detected or measured. Also, different borehole and formation conditions can require different tools to measure the same basic property [19]. Some of the numerous uses of logs in petroleum engineering include the identification of potential reservoir rocks and bed thickness, determination of porosity, estimation of permeability, location of and quantification of the amount of hydrocarbons, and estimation of water salinity.

2.2 Porosity and permeability

Porosity is the percentage of voids and open spaces in a rock or sedimentary deposit. The greater the porosity of a rock, the greater its ability to hold water and other materials, such as oil, will be. It is an important consideration when attempting to evaluate the potential volume of hydrocarbons contained in a reservoir. Sedimentary

porosities are a complex function of many factors, including but not limited to, rate of burial, depth of burial, the nature of the carbonate fluids, and overlying sediments which may impede fluid expulsion. Different types of porosity are primary, secondary, fracture, and vuggy porosity. Porosity can be further categorized as effective and total porosity [20].

Although a rock may be very porous, it is not necessarily very permeable. Permeability is a measure of how interconnected the individual pore spaces are in a rock or sediment. It is a key parameter associated with the characterization of any hydrocarbon reservoir. In fact, many petroleum engineering problems cannot be solved accurately without having an accurate value of permeability. Permeability can also be categorized as absolute, effective, and relative permeability [21].

Together, these two reservoir properties are key indicators of reservoir viability and quality. They are also very essential attributes in the determination of other properties such as pressure–volume–temperature (PVT), hydraulic units, water saturation, oil/gas ratio, and wellbore stability.

3 Related works on hybrid computational intelligence in oil and gas

The application of HCI has been widely appreciated in petroleum engineering, as well as in other fields. Some of the areas of petroleum technology in which AI has been successfully used include seismic pattern recognition, porosity and permeability predictions, identification of sandstone lithofacies, drill bit diagnosis, and analysis and improvement of oil and gas well production [2, 6, 18, 22].

A number of hybrid models combined artificial neural networks (ANN) and fuzzy logic [1, 3, 4, 23]. In some of these models, the fuzzy logic component was used to select the best related well logs with core porosity and permeability data while the ANN component was used as a nonlinear regression method to develop transformation between the selected well logs and core measurements.

Both ANN and fuzzy logic, as implemented in those studies, have limitations that hamper their choice for such roles. ANN suffers from the following deficiencies [24]:

- There is no general framework to design the appropriate network for a specific task.
- The number of hidden layers and hidden neurons of the network architecture are determined by trial and error.
- A large number of parameters are frequently required to fit a good network structure.
- ANN uses pre-defined activation functions without considering the properties of the phenomena being modelled.

Fuzzy logic, on the other hand, especially T2FLS, was reported to be very complex and hence requires more time for execution when applied on high-dimensional data [10]. It also performs poorly when applied on datasets of small size [9, 25]. Hence, there is the need for another technique that does not have these limitations. It is necessary to state here that the work in [25] contained the initial results of the present work in its rather ad hoc implementation without much attention to the need to optimize the parameters. It was initial test run of the hybrid model. This work extends the previous by paying much more careful attention to the algorithms that were combined. The results presented here are much better.

Other hybrid techniques incorporated genetic algorithms (GA) in their components. These include [5] who developed a hybrid model of GA and fuzzy/neural inference system methodology that provides permeability estimates for all types of rocks in order to determine the volumetric estimate of permeability. Their proposed hybrid system consists of three modules: one that serves to classify the lithology and categorizes the reservoir interval into user-defined lithology types, a second module containing GA to predict permeability facies, and the third module that uses neuro-fuzzy inference systems to form a relationship for each permeability facies and lithology. Similar hybrid configurations incorporating GA were used by [26–31]. However, GA, with its exhaustive search algorithm and a very robust optimization algorithm, is known for its long execution time, its need for high processing power due to its complexity and sometimes inefficiency [32, 33].

Among the most popular hybrid techniques used in the modelling of oil and gas reservoir properties is the adaptive neuro-fuzzy inference system (ANFIS), which combines the functionalities of ANN and fuzzy logic techniques. This featured in the study of [34] where it was used to predict the permeability of tight gas sands using a combination of core and log data. Combining the limitations of ANN and fuzzy logic, as earlier discussed, in a single hybrid model could only result in a combined overall inefficiency despite their reported good performance.

Various studies to address these reported problems of ANN through the development of other algorithms such as cascade correlation and radial basis function did not improve its overall performance [35]. It has not been proved in the literature that the use of fuzzy logic and GA components in hybrid models was able to effectively neutralize the limitations of ANN.

Due to the limitations of ANN, fuzzy logic, and GA used in the reported hybrid techniques, the objective of this paper is to use a technique with simple architecture and more robust fitting algorithm to extract the most relevant and dimensionally reduced set of input variables for the use

of T2FLS model resulting in better performance with reduced time complexity and FN fits perfectly to this need.

An overview of FN, T2FL, and hybrid systems has been succinctly presented in [6]. More details about these techniques can further be found in [36–38] for FN, [1, 9, 10, 31] for T2FL and [3, 5, 22, 26] for hybrid systems.

4 Data description, experimental design, and model framework

4.1 Description of data

The same sets of porosity and permeability data from six wells (three for porosity and three for permeability) from previous studies [2, 6] were used in this work. Table 1 shows the six predictor variables for porosity from Site 1; Table 2 shows the eight predictor variables for permeability from Site 2, and the six datasets with their number of data points as well as their divisions into training and testing sets are shown in Table 3. Site 1 is a heterogeneous platform that is made up of carbonate and dolomite while Site 2 is majorly of carbonate and sandstone formations. Hence, the datasets are the representative of the major oil-bearing geological structures found in most parts of the oil-producing world.

4.2 Experimental design

The methodology employed in this study follows the standard computational intelligence approach to hybridization of AI techniques, and in the case of this study, a combination of the LS-fitting algorithm of FN and T2FLS techniques. In this work, the MATLAB codes for the iterative least-squares FN classifier obtained from the software repository of Enrique Castillo's AI Research Group [39] were partly used. For T2FLS, the Interval Type-2 version and the MATLAB codes provided at Mendel's software repository [40] were used. The use of Interval T2FLS was preferred over the general Type-2 version because, according to [9, 10], the former is too complicated as the process of calculating the meet

Table 1 Predictor variables for Site 1 well log for porosity

| Predictors for porosity | | |
|-------------------------|----|---------------|
| 1 | CR | Core |
| 2 | TP | Top interval |
| 3 | GD | Grain density |
| 4 | GV | Grain volume |
| 5 | LT | Length |
| 6 | DM | Diameter |

Table 2 Predictor variables for Site 2 well log for permeability

| Predictors for permeability | | Full meaning |
|-----------------------------|------|------------------------------|
| 1 | GR | Gamma ray log |
| 2 | PHIE | Porosity log |
| 3 | RHOB | Density log |
| 4 | SWT | Water saturation |
| 5 | RT | Deep resistivity |
| 6 | MSFL | Microspherically focused log |
| 7 | NPHI | Neutron porosity log |
| 8 | CALI | Caliper log |

Table 3 Division of datasets into training and testing

| | Site 1 (porosity) | | | Site 2 (permeability) | | |
|-----------------|-------------------|-----|----|-----------------------|-----|-----|
| | 1 | 2 | 3 | 1 | 2 | 3 |
| Wells | 1 | 2 | 3 | 1 | 2 | 3 |
| Data size | 415 | 285 | 23 | 355 | 477 | 387 |
| Training (70 %) | 291 | 200 | 16 | 249 | 334 | 271 |
| Testing (30 %) | 124 | 85 | 7 | 106 | 143 | 116 |

operations for each fired rule and the procedures for type reduction are prohibitive, time-consuming, and memory-intensive. Hence, Interval T2FLS is more practical for implementation than the general Type-2.

The hybrid model was designed to benefit immensely from the strength of the LS-based FN by complementing some of the weaknesses of T2FL such as increased complexity of implementation with increased input data dimensionality and extracting the best variables to improve its performance, and hence to combine the cooperative and competitive characteristics of the individual techniques.

4.2.1 Framework of the hybrid model

The proposed hybrid model is composed of two blocks containing, respectively, LS-FN and T2FLS. The LS-FN block, with its least-squares-fitting algorithm, was used to select the best variables from the input data and then the input data were divided into training and testing subsets using the stratified sampling approach previously used in [2, 6, 18, 25, 36]. The dimensionally reduced variables of both the training and testing subsets are applied on the T2FLS block for training and prediction. A schematic diagram of this model is shown in Fig. 1. Only the part of FN that performs the least-square fitting as part of its training procedure was used in this work. No testing occurred in the FN block. Each dataset was passed through the FN block and the output contains only the variables that were found to be most relevant to the target (porosity and permeability). These are shown in Tables 4 and 5.

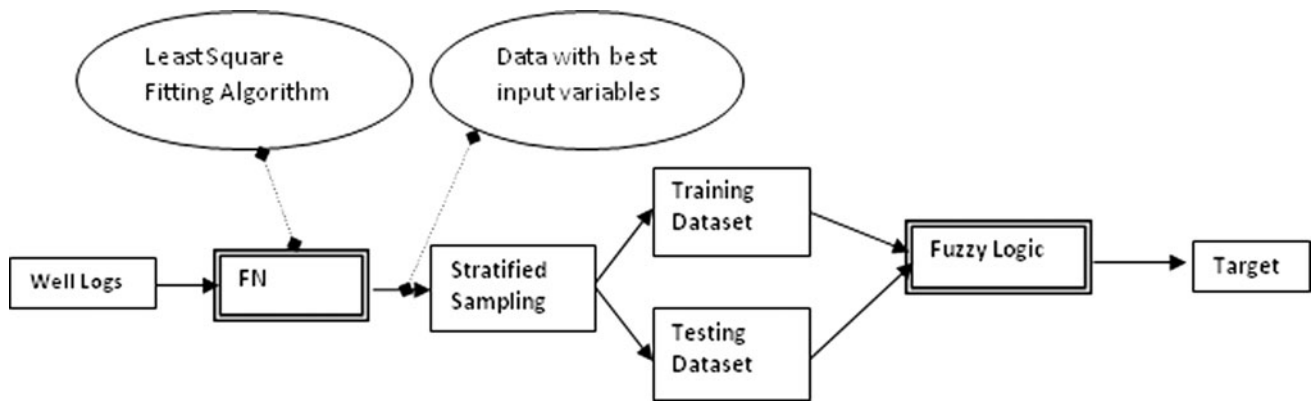


Fig. 1 Design framework of the hybrid model

Table 4 Initial and selected parameters for porosity

| | |
|----------------------------|--|
| Initial variables | 6: core, top interval, grain density, grain volume, length, and diameter |
| Best subset selected by FN | 4: top interval, grain density, grain volume, and length. |

Table 5 Initial and selected parameters for permeability

| | |
|----------------------------|--|
| Initial variables | 8: GR, PHIE, RHOB, SWT, RT, MSFL, NPHI, and CALI |
| Best subset selected by FN | 4: PHIE, RHOB, SWT, and MSFL |

For the LS–FN component, the iterative least-squares-fitting algorithm segment of the FN technique was used for the implementation of the hybrid model. This algorithm has the ability to learn itself and to use the input data directly, by minimizing the sum of squared errors, in order to obtain the parameters, namely the number of neurons and the type of kernel functions needed for the fitting process. It works by building an initial model, simplifying the model, and selecting the best parameters for the simplified model. The backward–forward function of the algorithm takes different combinations of the input variables and determines their correlation with the target variable. Using the criteria of the mean square errors and minimum description length (MDL), the functions iteratively add and remove one variable at a time to determine the most correlated of the set of variables with the target variable. The model that evolves corresponds to those variables that gave the least values mean square errors and the MDL.

With model initialization, the value of porosity and permeability of a well is determined by the parameters shown in Tables 1 and 2. The parameters (represented here by x, y, z for simplicity) were modelled in an initial network like the one shown in Fig. 2 and were then reduced to the simplified version equivalent of the initial network as shown in Fig. 3.

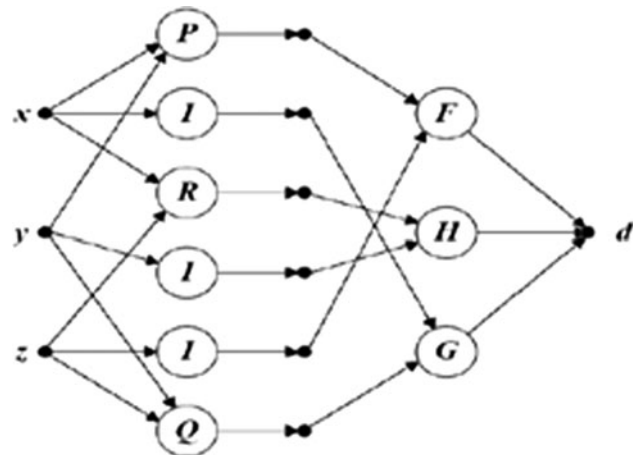


Fig. 2 Initial topology of FN corresponding to the combined functional equations [36]

With model selection, part of the steps in FN training is the model selection procedure using the MDL principle. This measure allows comparisons not only of the quality of different approximations, but also of different FN models. It is also used to compare models with different parameters due to the use of a penalty term for overfitting. Moreover, it is distribution-independent making it a convenient method for solving the model selection problem. Accordingly, the best FN model for a given problem corresponds to the one with the smallest description length value. This was calculated using the backward-elimination and forward-selection methods.

The backward-elimination process starts with the complete model with all parameters, and it sequentially removes the one that will lead the model to the smallest value of the MDL measure, repeating the process until there is no further improvement in the measure. Next, the forward-selection process is applied, but starting from the final model of the backward process, and it sequentially adds the variable that leads to the smallest value of MDL

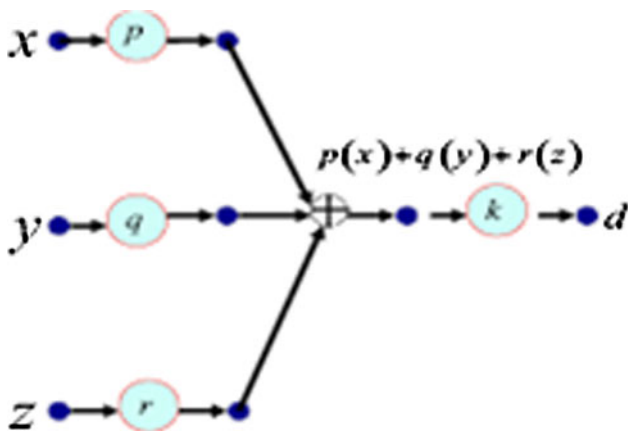


Fig. 3 Simplified network [36]

measure. This process is repeated until there is no further improvement in MDL measure obtained either by removing or adding a single variable. The learning process of the LSFN includes obtaining the neural functions from a set of training data based on minimizing the sum of squared errors between the input and the target output by suggesting an approximation to each of the functions and selecting the best among them [36–38]. For simplicity, we used polynomial functions with degree = 1.

Generally, the least-square algorithm attempts to find the least of the mean differences between the actual observations, y and the outputs of the evolved models, X , called the residuals and denoted as:

$$r = y - X(\alpha) \tag{1}$$

The aim is to find the α 's that make the residuals as small as possible. Hence, minimizing the sum of the squares of the residuals:

$$\|r\|^2 = \sum_1^m r_i^2 \tag{2}$$

MDL is generalized as follows: For a given set of hypothesis H and dataset D , the aim is to find the subset of hypothesis in H that learns D most.

Let $H^{(1)}, H^{(2)}, \dots, H^{(k)}$ be a list of candidate models each containing a set of point hypotheses. The best point hypothesis $H \in H^{(1)} \cup H^{(2)} \cup \dots \cup H^{(k)}$ to explain the data D is the one which minimizes the sum $L(H)$, the length of the description of the hypothesis. The model to explain D adequately is the smallest model containing the selected H . This helps to avoid overfitting while choosing a trade-off between goodness of fit and complexity of the models involved. Such trade-offs lead to much better predictions of test data than one would get by adopting the “simplest” (one degree) or most complex ($n - 1$ -degree) polynomial. MDL provides one particular means of achieving such a trade-off [41].

Interval T2FLS with Gaussian membership functions was used because they are simple to implement and at present, it is very difficult to justify the use of any other kind. When the T2FLS is Interval Type-2, all secondary grades (flags) are equal to 1 and the rules were extracted directly from the input data using the back-propagation (steepest descent) method. This option was preferred over the extraction of rules from a subset of the data, because this ensures all data points are completely represented in the rule base. This leads to more time spent in implementation, but better accuracy and efficiency are guaranteed. To be able to use T2FLS effectively, the standard steps as suggested by [9, 10] were followed.

For inferencing, the rules are of the general form:

$$R1: \text{IF } x_1 \text{ is } F_1^l \text{ and } x_2 \text{ is } F_2^l \text{ and } \dots \text{ and } x_p \text{ is } F_p^l \text{ THEN } y \text{ is } G^l \tag{3}$$

representing a Type-2 fuzzy relationship between the input space $x_1 \times x_2 \times \dots \times x_p$ and the output space Y of the system. The membership function of this Type-2 relation is denoted by:

$$\mu_{F_1^l x_1 \dots x_p F_p^l \rightarrow G^l}(x, y) \tag{4}$$

where $F_1^l x_1 \dots x_p F_p^l$ denotes the Cartesian product of $F_1^l, F_2^l, \dots, F_p^l$ and $x = \{x_1, x_2, \dots, x_p\}$. This was then converted to a Type-2 fuzzy set by applying the extension principle, details of which can be found in [9, 10].

The output set corresponding to each rule of the T2FLS is a Type-2 set. There is then the need to reduce it to 1, in order to give room for defuzzification. The type-reducer combines all these output sets to a sum of 1 using this formula:

$$C_A = \frac{\sum_{i=1}^N X_i \mu_A(x_i)}{\sum_{i=1}^N \mu_A(x_i)} \tag{5}$$

The type-reduced set of a T2FLS is the centroid of a Type-2 output set for the FLS. The type-reduced set was defuzzified to get a crisp output from the T2FLS. The most natural way of doing this is by finding the centroid of the type-reduced set. The learning parameter, alpha, was set to 0.1 as recommended in [9, 10].

The T2FLS was run independently and simultaneously with the hybrid model under the same input data conditions. In order to ensure the most fairness in the results, several iterations were made and the average values of the results were taken. This is necessary due to the behavior of the stratified sampling approach used to divide the input data into training and testing sets. Since the input data were randomized during the stratification processes, hence, slightly different results were obtained for each run of the experiments.

4.2.2 Justification for the proposed framework

With the excellent LS-fitting capabilities of FN and the ability of T2FLS to handle uncertainties that might be present in data, combining LS-FN and T2FLS is expected to overcome the limitations of the previously reported studies of hybrid techniques [9, 10, 24, 32–34]. Basically, the LS–FN block of the proposed hybrid will select the best subsets of the input data for the training and testing of the T2FLS block. This will ensure that the T2FLS block is supplied with both the best of the variables and a version of the same data with reduced dimensionality. The latter would reduce the complexity of the T2FLS processing and the former will ensure that the variables that are irrelevant to the prediction process are filtered out and hence would not pollute the overall performance of the proposed hybrid model.

To our knowledge, LS–FN has not been used as a best-model selector in any of the reported hybrid techniques. With the recent advances in data acquisition tools in the oil and gas industry such as logging while drilling (LWD) and measurement while drilling (MWD), many logs are produced and there is the need to extract from these logs only those variables that are most relevant to our target reservoir properties.

4.3 Computing environment and model evaluation

The computing environment used for the simulation in this study consists of a 32-bit MATLAB 2010a software that runs on a 64-bit Personal Computer with Windows 7 Professional Edition version 2009. The processor is based on Intel Core 2 Quad CPU with a speed of 3 GHz and a RAM size of 6 GB.

The performance of the models was evaluated using the correlation coefficient (CC), root mean-squared error (RMSE), and execution time (ET). CC measures the statistical correlation between the predicted and actual values. RMSE is one of the most commonly used error measures of success for numeric prediction as it computes the average of the squared differences between each predicted value and its corresponding actual value. ET is simply the total time taken for a technique to run from the beginning to its end using the CPU time.

5 Experimental results and discussion

All the results obtained from the simulation are shown in Tables 6, 7, 8, 9, 10, and 11. Tables 6 and 7 presented the comparative correlation coefficients for the porosity and permeability wells, respectively. Tables 8 and 9 showed the comparative root mean square errors while

Table 6 Correlation coefficient results for porosity wells

| Wells | Training | | Testing | |
|-------|----------|----------|----------|----------|
| | T2FL | FN-T2FL | T2FL | FN-T2FL |
| 1 | 0.83569 | 0.868561 | 0.756021 | 0.844328 |
| 2 | 0.742781 | 0.836016 | 0.686023 | 0.814673 |
| 3 | 0.804283 | 0.817494 | 0.788061 | 0.921438 |

Table 7 Correlation coefficient results for permeability wells

| Wells | Training | | Testing | |
|-------|----------|----------|----------|----------|
| | T2FL | FN-T2FL | T2FL | FN-T2FL |
| 1 | 0.791102 | 0.812469 | 0.758849 | 0.794979 |
| 2 | 0.859088 | 0.891122 | 0.846123 | 0.86618 |
| 3 | 0.702918 | 0.812158 | 0.731261 | 0.79865 |

Table 8 Root mean square error results for porosity wells

| Wells | Training | | Testing | |
|-------|----------|---------|---------|---------|
| | T2FL | FN-T2FL | T2FL | FN-T2FL |
| 1 | 7.25699 | 6.90762 | 7.26232 | 7.10594 |
| 2 | 7.968 | 6.78391 | 8.29779 | 7.35157 |
| 3 | 5.58098 | 5.49339 | 7.42699 | 6.69004 |

Table 9 Root mean square error results for permeability wells

| Wells | Training | | Testing | |
|-------|----------|---------|---------|---------|
| | T2FL | FN-T2FL | T2FL | FN-T2FL |
| 1 | 0.78905 | 0.75328 | 0.82061 | 0.79909 |
| 2 | 0.66425 | 0.63462 | 0.71684 | 0.66506 |
| 3 | 0.95728 | 0.92315 | 0.96747 | 0.93699 |

Table 10 Execution time results for porosity wells

| Wells | Training (s) | | Testing (s) | |
|-------|--------------|---------|-------------|---------|
| | T2FL | FN-T2FL | T2FL | FN-T2FL |
| 1 | 185.79 | 178.3 | 37.02 | 34.06 |
| 2 | 87.14 | 85.75 | 17.08 | 14.5 |
| 3 | 0.33 | 0.27 | 0.05 | 0.04 |

Table 11 Execution time results for permeability wells

| Wells | Training (s) | | Testing (s) | |
|-------|--------------|---------|-------------|---------|
| | T2FL | FN-T2FL | T2FL | FN-T2FL |
| 1 | 164.78 | 153.59 | 32.41 | 29.76 |
| 2 | 407.36 | 333.01 | 84.99 | 63.31 |
| 3 | 187.61 | 179.74 | 36.88 | 34.98 |

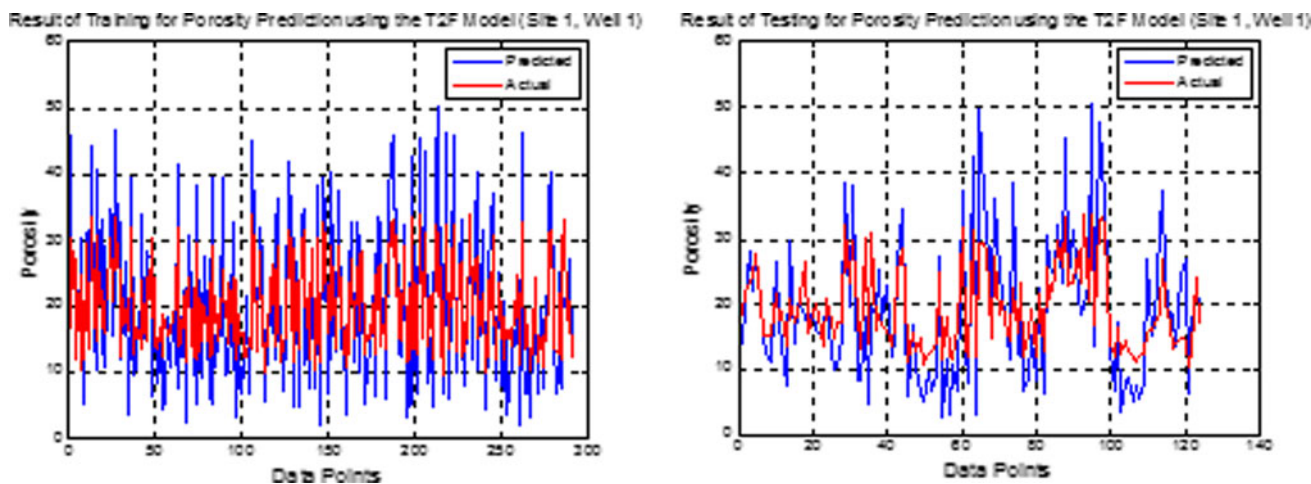


Fig. 4 Sample of actual-predicted porosity for training and testing with T2FL

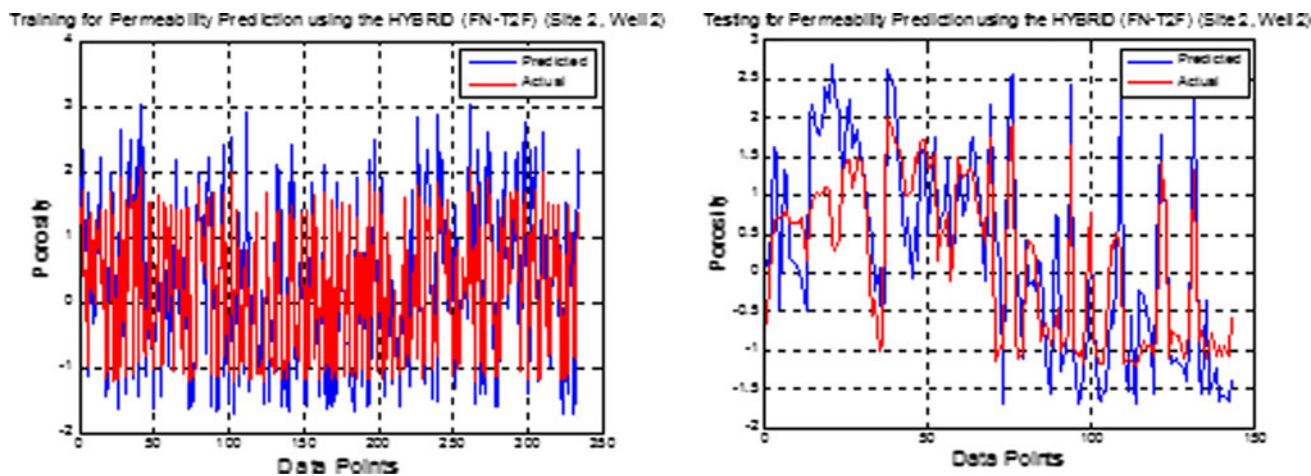
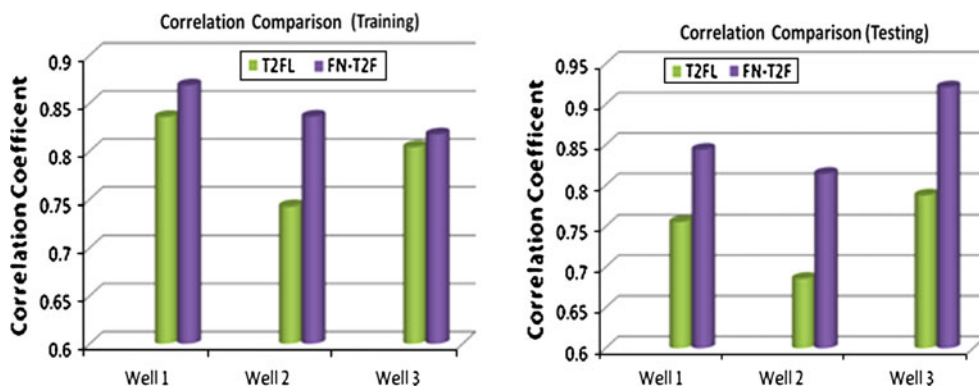


Fig. 5 Actual and predicted porosity of training and testing for Site 2 well 2 dataset using FN-T2FL hybrid model

Fig. 6 Porosity training and testing correlation coefficients for all wells



Tables 10 and 11 displayed the comparative execution time for porosity and permeability wells, respectively. Figures 4 and 5 showed the plots of the actual and predicted values of porosity and permeability with T2FLS and the LS-FN-T2FLS hybrid models from Site 1

well 1 and Site 2 well 2 datasets. Figures 6, 7, 8, 9, 10, 11, and 12 showed the comparisons of the porosity and permeability predictions indicating the comparative performance of the models with respect to CC, RMSE, and ET.

Fig. 7 Permeability training and testing correlation coefficients for all wells

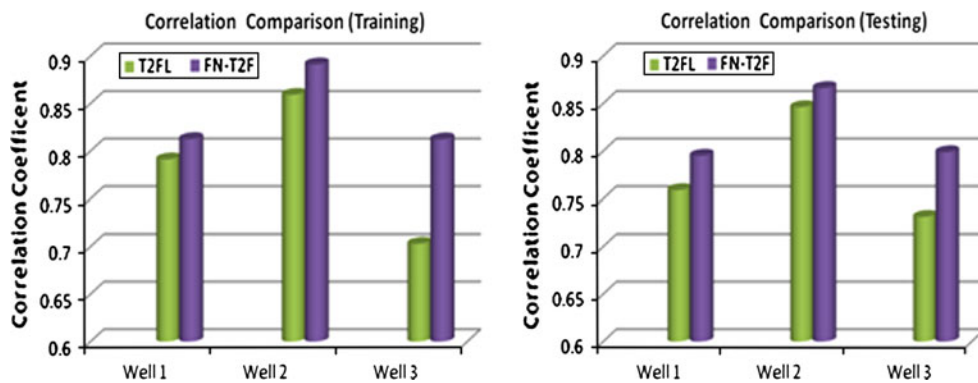


Fig. 8 Root mean square errors for porosity training and testing for all wells

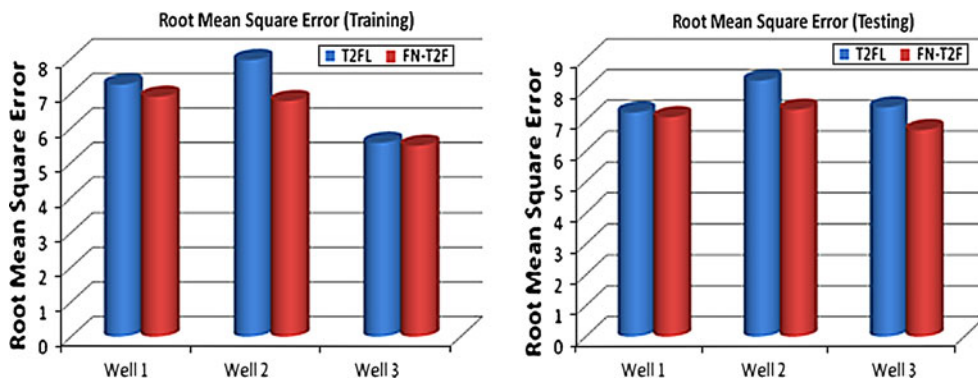


Fig. 9 Root mean square errors for permeability training and testing for all wells

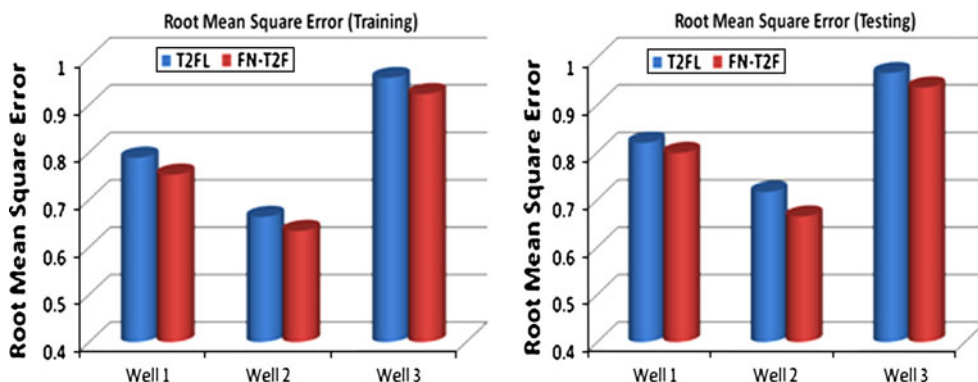
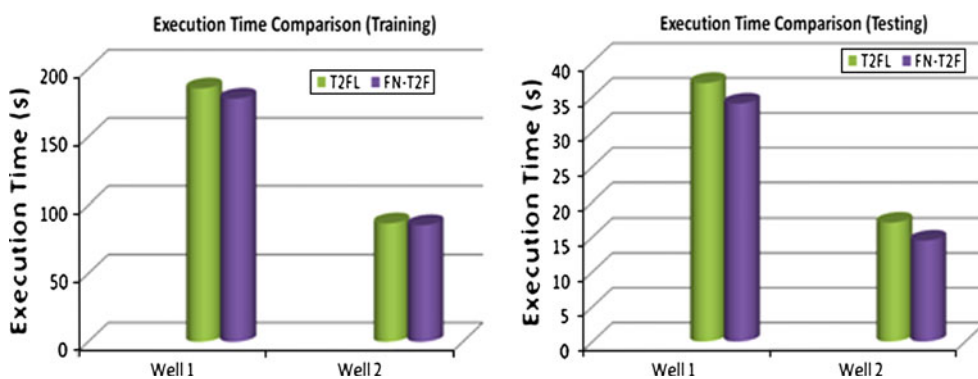


Fig. 10 Execution time for porosity training and testing for wells 1 and 2



The results showed the better performance of the LS-FN-T2FLS hybrid over the original T2FLS technique. From the visual analysis of the plots, the plot lines of the actual and

the predicted values are closer for LS-FN-T2FLS (Fig. 5) than in the original T2FLS (Fig. 4). This indicates the superior performance of the FN-T2FLS hybrid model.

Fig. 11 Execution time for porosity training and testing for well 3

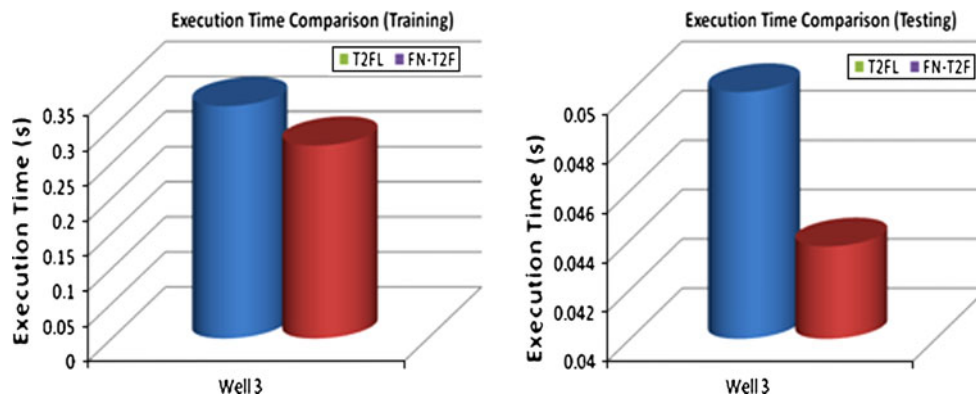
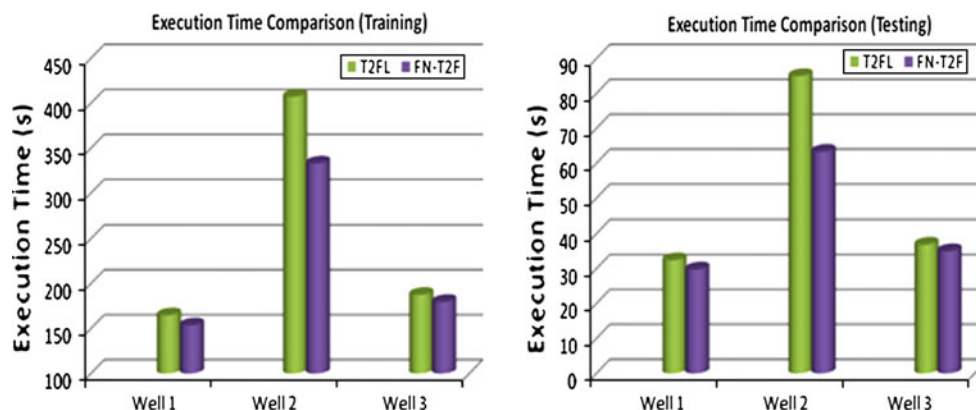


Fig. 12 Execution time for permeability training and testing for all wells



In terms of CC, Figs. 6 and 7 show that the predictions of the hybrid model are more correlated with the actual values of the target variable. Though the predictions are not perfect since real-life field datasets were used, the performance of the hybrid model is more significantly noticed over T2FLS. Since it is expected that a higher correlation should have the less error, Figs. 8 and 9 show the hybrid model having less RMSE than the T2FLS.

Despite that there are more internal processes involved in the hybrid model than in T2FLS, the former runs faster in terms of CPU time than the latter. It is shown in Figs. 10, 11, and 12 that it took the hybrid model less time for training and testing than the T2FLS.

The better performance exhibited by the FN-T2FL hybrid in Figs. 6, 7, 8, 9, 10, 11, and 12 can be attributed to the role of the LS-FN block in the extraction of the most relevant input variables for the training and testing of the T2FLS block. This ensures that the T2FLS block handled only the best of the datasets and hence is not corrupted by the redundant and irrelevant variables from the original datasets.

It would have been expected that the hybrid model would need more time for execution than the T2FLS since there are almost two techniques (LS-FN and T2FLS) and two processes (best subset selection and data stratification) that were executed before the T2FLS block, only the segment that performs the model selection procedure using the

LS-fitting algorithm was used in this work. The dimensionally reduced dataset that was used by the T2FLS block from the output of the LS-FN block also ensured that the execution of the T2FLS block is less complex than the original T2FLS model executed in parallel with the hybrid; hence, less work was done in handling the input–output data matrix.

6 Conclusions

A design framework for the hybridization of LS-FN and T2FLS was implemented and presented. The hybrid model was tested with six datasets from different geological formations to predict the porosity and permeability of oil and gas reservoirs. From the results presented, the following conclusions could be reached:

- The hybrid model consisting of the combination of LS-FN and T2FLS performed better than the original T2FLS in terms of higher correlation coefficients, lower root mean square errors, and reduced execution time.
- The better performance of the LS-FN-T2FLS hybrid model was due to the dual role of LS-FN to select the best and most relevant input variables and the consequent reduction in the dimensionality of the data that were used by the T2FLS block.

- Hence, the subset selection improved the correlation of the hybrid model while the reduced dimensionality of the input data reduced the time and space complexity of the T2FLS block, thereby reducing its processing time.
- Both the higher correlation and the reduction in processing time would be great assets toward the actualization of hybrid models in real-time reservoir modelling.

The results of this study have shown that FN, with its iterative least-square-fitting algorithm, is a good candidate for building hybrid computational intelligence tools that will offer improved performance and increased robustness. This study has further confirmed that the application of hybrid techniques in petroleum reservoir characterization will increase the accuracy of petroleum exploration activities and hence increase the production of more oil and gas since a fractional increase in the correlation of this hybrid model would result in the greater precision of the prediction of reservoir properties in considerably less time, which would, in turn, increase the production capacity of oil and gas reservoirs. This would have an overall positive effect on the production of the energy the world needs.

7 Further research

Another similar hybrid technique, but with support vector machines (SVM) replacing the Type-2 fuzzy systems block, will be implemented. The performance of this will then be compared with the standalone SVM and the LS-FN-T2FLS model implemented in this paper. The two hybrids will, in turn, be compared with the existing adaptive neuro-fuzzy inference system for possible proof of performance improvement.

Acknowledgments The authors would like to thank the Universiti Malaysia Sarawak and King Fahd University of Petroleum and Minerals for providing the resources used in the conduct of this study. The authors would also like to acknowledge the support provided by King Abdulaziz City for Science and Technology (KACST) through the Science & Technology Unit at King Fahd University of Petroleum & Minerals (KFUPM) for funding this work under Project No. 08-OIL82-4 as part of the National Science, Technology and Innovation Plan.

References

- Jong-Se L (2005) Reservoir properties determination using fuzzy logic and neural networks from well data in offshore Korea. *J Pet Sci Eng* 49:182–192
- Helmy T, Anifowose F, Faisal K (2010) Hybrid computational models for the characterization of oil and gas reservoirs. *Int J Exp Syst Appl* 37:5353–5363
- Ferraz IN and Garcia CB (2005) Lithofacies recognition hybrid bench. In: Proceedings of the 5th international conference on hybrid intelligent systems, IEEEXplore/ACM digital library
- Salim C (2006) A fuzzy art versus hybrid NN-HMM methods for lithology identification in the Triassic province. In: Proceedings of the 2nd international conference on information and communication technologies. IEEEXplore 1:1884–1887
- Xie D, Wilkinson D, Yu T (2005) Permeability estimation using a hybrid genetic programming and fuzzy/neural inference approach. In: Proceedings of the SPE annual technical conference and exhibition, Dallas, Texas, USA
- Anifowose F and Abdullaheem A (2010) Prediction of porosity and permeability of oil and gas reservoirs using hybrid computational intelligence models. Paper SPE-126649. In: Proceedings of the society of petroleum engineers North Africa technical conference and exhibition, Cairo, Egypt, February
- Chang FM (2008) Characteristics analysis for small data set learning and the comparison of classification methods. In: Proceedings of the 7th WSEAS international conference on artificial intelligence, knowledge engineering and data bases, Cambridge. Kazovsky L, Borne P, Mastorakis N, Kuri-Morales A and Sakellaris I (eds) Artificial Intelligence Series. World Scientific and Engineering Academy and Society (WSEAS), Stevens Point, Wisconsin, pp 122–127
- Van BV, Pelckmans K, Van HS, Suykens JA (2011) Improved performance on high-dimensional survival data by application of Survival-SVM. *Bioinformatics* 27(1):87–94
- Mendel JM (2003) Type-2 fuzzy sets: some questions and answers. *IEEE Connect Newsl IEEE Neural Netw Soc* 1:10–13
- Mendel JM, Robert II, Liu F (2006) Interval type-2 fuzzy logic systems made simple. *IEEE Trans Fuzzy Syst* 14(6):808–821
- Peddabachigari S, Abraham A, Grosan C, Thomas J (2011) Modelling intrusion detection system using hybrid intelligent systems. *J Netw Comput Appl* 30:114–132
- Mendoza O, Licea G, Melin P (2007) Modular neural networks and type-2 fuzzy logic for face recognition. In: Proceedings of the annual meeting of the north american fuzzy information processing society, pp 622–627
- Jin B, Tang YC, Yan-Qing Z (2007) Support vector machines with genetic fuzzy feature transformation for biomedical data classification. *Info Sci* 177:476–489
- Helmy T, Al-Harhi MM, Faheem MT (2012) Adaptive ensemble and hybrid models for classification of bioinformatics datasets. *Trans Fuzzy Neural Netw Bioinform Glob J Technol Optim* 3:20–29
- Lean Y, Kin KL, Shouyang W (2006) Credit risk assessment with least squares fuzzy support vector machines. In: Proceedings of the 6th IEEE international conference on data mining workshops, IEEEXplore, pp 823–827
- Evaggelos S, Giorgos S, Yannis A, Stefanos K (2006) Fuzzy support vector machines for image classification fusing Mpeg-7 visual descriptors. Technical Report. Image, Video and Multimedia Systems Laboratory, School of Electrical and Computer Engineering, National Technical University of Athens, Greece
- Bullinaria JA, Li X (2007) An introduction to computational intelligence techniques for robot control. *Ind Robot* 34(4): 295–302
- Helmy T, Anifowose F (2010) Hybrid computational intelligence models for porosity and permeability prediction of petroleum reservoirs. *Int J Comput Intell Appl* 9(4):313–337
- Accessscience Encyclopaedia Article (2011) Well Logging. Available online: <http://www.accessscience.com/abstract.aspx?fid=744100&referURL=http%3a%2f%2fwww.Accessscience.com%2fcontent.aspx%3fid%3d744100>. Accessed on 28th March
- Schlumberger Oil Field Glossary (2011) Porosity, Available online: <http://www.glossary.oilfield.slb.com/Display.cfm?Term=porosity>. Accessed on April 20
- Schlumberger Oil Field Glossary (2011) Permeability, Available online:

- <http://www.glossary.oilfield.slb.com/Display.cfm?Term=permeability>. Accessed on April 20
22. Anifowose F (2009) Hybrid artificial intelligence models for the characterization of oil and gas reservoirs: concept, design and implementation. VDM Verlag, Germany
 23. Singh V, Singh TN (2006) A neuro-fuzzy approach for prediction of Poisson's ratio and young's modulus of shale and sandstone. In: Proceedings of the 41st United States symposium on rock mechanics (USRMS), Golden Rocks
 24. Petrus JB, Thuijsman F, Weijters AJ (1995) Artificial neural networks: an introduction to ann theory and practice. Springer, Netherlands
 25. Anifowose F and Abdulaheem A (2010) A functional networks-type-2 fuzzy logic hybrid model for the prediction of porosity and permeability of oil and gas reservoirs. Paper #1569334237. In: Proceedings of the 2nd international conference on computational intelligence, modelling and simulation (CIMSIm 2010), IEEEEXPlore, pp 193–198
 26. Park HJ, Lim JS, Roh U, Kang JM, Min BH (2006) Production-system optimization of gas fields using hybrid fuzzy-genetic approach. In: Proceedings of the SPE Europec/EAGE annual conference and exhibition, Vienna, Austria
 27. Lim J, Park H, Kim J (2006) A new neural network approach to reservoir permeability estimation from well logs. In: Proceedings of the SPE Asia pacific oil and gas conference and exhibition, Adelaide, Australia
 28. Mohsen S, Morteza A, Ali YV (2007) Design of neural networks using genetic algorithm for the permeability estimation of the reservoir. *J Pet Sci Eng* 59:97–105
 29. Zarei F, Daliri A, Alizadeh N (2008) The use of neuro-fuzzy proxy in well placement optimization. In: Proceedings of the SPE intelligent energy conference and exhibition, Amsterdam, The Netherlands
 30. Al-Anazi A, Gates I, Azaiez J (2009) Innovative data-driven permeability prediction in a heterogeneous reservoir. In: Proceedings of the SPE EUROPEC/EAGE annual conference and exhibition, Amsterdam, The Netherlands
 31. Shahvar MB, Kharrat R, Mahdavi R (2009) Incorporating fuzzy logic and artificial neural networks for building hydraulic unit-based model for permeability prediction of a heterogenous carbonate reservoir. In: Proceedings of the international petroleum technology conference, Doha, Qatar
 32. Gao W (2012) Study on new improved hybrid genetic algorithm. In: Zeng D (ed) *Advances in information technology and industry applications*. Lecture Notes in Electrical Engineering 136, pp 505–512
 33. Bies RR, Muldoon MF, Pollock BG, Manuck S, Smith G, Sale ME (2006) A genetic algorithm-based, hybrid machine learning approach to model selection. *J Pharmacokinet Pharmacodyn* 33(2):195–221
 34. Ali L, Bordoloi S, Wardinsky SH (2008) Modelling permeability in tight gas sands using intelligence and innovative data mining techniques. In: Proceedings of the SPE annual technical conference and exhibition, Denver, Colorado
 35. Bruen M, Yang J (2005) Functional networks in real-time flood forecasting: a novel application. *Adv Water Res* 28:899–909
 36. El-Sebakhy E (2009) Software reliability identification using functional networks: a comparative study. *Exp Syst Appl* 36(2): 4013–4020
 37. El-Sebakhy E, Hadi AS, Kanaan FA (2007) Iterative least squares functional networks classifier. *IEEE Trans Neural Netw* 18(2): 1–7
 38. Castillo E, Gutiérrez JM, Hadi AS, Lacruz B (2001) Some applications of functional networks in statistics and engineering. *Technometrics* 43:10–24
 39. Castillo E (2011) Software repository. Artificial Intelligence Research Group. Available online: <http://ccaix3.unican.es/~AIGroup>, Accessed March 25
 40. Mendel J (2011) Matlab codes for type-2 fuzzy logic system. Accessed March 25. sipi.usc.edu/~mendel/software/
 41. Grünwald P (2005) A tutorial introduction to the minimum description length principle. In: Grünwald P, Myung IJ, Pitt M (eds) *Advances in minimum description length: theory and applications*. MIT Press, Cambridge