Applied Soft Computing 13 (2013) 1085–1098

Contents lists available at SciVerse ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

# Evolving artificial neural network and imperialist competitive algorithm for prediction oil flow rate of the reservoir

Mohammad Ali Ahmadi [a,*], Mohammad Ebadi [b], Amin Shokrollahi [c], Seyed Mohammad Javad Majidi [c]

[a] Department of Petroleum Engineering, Ahwaz Faculty of Petroleum Engineering, Petroleum University of Technology, Kut Abdollah, Ahwaz, Iran
[b] Department of Petroleum Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran
[c] Department of Chemical and Petroleum Engineering, Sharif University of Technology, Tehran, Iran

ARTICLE INFO

*Article history:*
Received 27 September 2011
Received in revised form 28 August 2012
Accepted 11 October 2012
Available online 22 October 2012

*Keywords:*
Artificial neural network
Imperialist competitive optimization
Oil flow rate
Fuzzy logic
Hybrid
Evolutionary algorithms

ABSTRACT

Multiphase flow meters (MPFMs) are utilized to provide quick and accurate well test data in numerous numbers of oil production applications like those in remote or unmanned locations topside exploitations that minimize platform space and subsea applications. Flow rates of phases (oil, gas and water) are most important parameter which is detected by MPFMs. Conventional MPFM data collecting is done in long periods; because of radioactive sources usage as detector and unmanned location due to wells far distance. In this paper, based on a real case of MPFM, a new method for oil rate prediction of wells base on Fuzzy logic, Artificial Neural Networks (ANN) and Imperialist Competitive Algorithm is presented. Temperatures and pressures of lines have been set as input variable of network and oil flow rate as output. In this case a 1600 data set of 50 wells in one of the northern Persian Gulf oil fields of Iran were used to build a database. ICA-ANN can be used as a reliable alternative way without personal and environmental problems. The performance of the ICA-ANN model has also been compared with ANN model and Fuzzy model. The results prove the effectiveness, robustness and compatibility of the ICA-ANN model.


## 1. Introduction

Undoubtedly, every technical aspect of production engineering in petroleum industry is a strong function of measuring accurately and precisely the multi-phase flow rate in the upstream oil and gas operations. The importance of the referred topic is highlighted during designing and running the well completion procedures. In more details, economical optimization of the wellhead equipment to maximize the hydrocarbon production and minimize the other unfavorable fluids production as well as lessening the total cost of production processes is generally accessible when the top ulterior rate of production have already been determined based on the capability of the supposed reservoir and technology related constraints [1–5]. For instance, in a field including numbers of already drilled wells, the relevant production data, existing in an extended sheet, such as: oil flow rate, temperature and pressure and the subsequent conclusions can channel the production experts and engineers towards making a more reliable decision about the completion plan of the ongoing drilling well. Besides, a more significant aspect of measuring the amount of flow rate is emerged in well testing, a mathematical approach based on the recordation of pressure versus time [6,7]. The well testing is an extremely useful tool to diagnosis the current conditions of the considering reservoir, and then scheming for the most brilliant and favorable target [8,9]. According to the definition, in the best scenario, the consequent detections evolving from different well testing methods can be applied at that present time and there is no chance to program thoughtfully for the future thanks to the nature of the well testing techniques. The remarked blind spot can be compensated by implementing different mathematical solutions, and specially the statistics, which provide this opportunity for well testing users to apply well test conventional solutions to prepare the most appropriate preventions. Artificial intelligence based models can be taken as one of the most favorable numerical and inverse tactics which can be tuned in well testing to make it capable of predicting the future flow rate by using the other easy to access parameters [10–12]. To sum up, it can be concluded that monitoring the flow rate is a crucial, vital and critical task which must correctly, strictly and validly be done to observe its key effect on other production related obstacles.

Temperature, pressure and flow rate are the most prominent variables acting the leading roles in production process from reservoir to production and separation unit [13]. Reservoir behavior simulation, production forecasting and separation process controls are deeply, highly and strongly affected by well production parameters (temperature, pressure and flow rates).

A soft sensor is a conceptual device whose output or inferred variable can be modeled in terms of other parameters that are

* Corresponding author. Tel.: +98 9126364936.
*E-mail address:* ahmadi6776@yahoo.com (M.A. Ahmadi).

1568-4946/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.asoc.2012.10.009

**Nomenclature**

*Acronyms*

| | |
|---|---|
| AI | artificial Intelligence |
| ANN | artificial Neural Network |
| BP | back propagation |
| C-o-A | center of area |
| FL | fuzzy logic |
| FDT | fuzzy decision tree |
| GA | genetic algorithm |
| ICA | imperialist competitive algorithm |
| L | low |
| H | high |
| M | medium |
| MSE | mean square error |
| MW | molecular weight |
| MFs | membership functions |
| NN | neural network |
| pdfs | probability distribution functions |

*Variables*

| | |
|---|---|
| $c_n$ | cost of $n$-th imperialist |
| $C_{\mathrm{n}}$ | normalized cost in Eq. (4) |
| $D_i$ | colony '$i$' |
| $d$ | distance between a colony and an imperialist |
| $f$ | cost function |
| $G$ | number of training samples |
| $m$ | number of output nodes |
| $N_{\mathrm{country}}$ | number of initial countries |
| $N_{\mathrm{imp}}$ | number of imperialists |
| $N_{\mathrm{col}}$ | number of colonies |
| $P$ | vector of possession probability for all empires |
| $P$ | pressure (Psi) |
| $P_i$ | country '$i$' |
| $q$ | flow rate |
| $R$ | uniform distribution of possession probability |
| SG | specific Gravity |
| $T_j(k)$ | actual output |
| $T$ | temperature (F) |
| To.$C_n$ | total cost of $n$-th empire |
| $Y_j(k)$ | expected output |
| $\mu(x)$ | membership function |
| $\alpha_j$ | designing values for the MF associated with the $j$-th linguistic term of an attribute |
| $I_j$ | $j$-th interval |
| $m_j$ | number of values in $I_j$ |
| $x_i$ | values of each data in $j$-th interval |
| $\pi(x)$ | possibility distribution |
| $E_{\mathrm{a}}$ | ambiguity |
| $S$ | subset hood |
| $C$ | class |
| $E$ | fuzzy evidence |
| $G$ | class ambiguity |
| $w$ | weight |

*Greek letter*

| | |
|---|---|
| $\beta$ | an arbitrary coefficient in Eq. (7) |
| $\zeta$ | a coefficient for total cost of an empire |

*Subscripts*

| | |
|---|---|
| Min | minimum |
| Max | maximum |

relevant to the same process. According to Rallo et al. [14], artificial neural networks are capable of being used as soft sensor building approach. The ANN is a popular, nonlinear, nonparametric tool in well log analysis. This technique has been increasingly applied to predict reservoir properties through using well log data [15,16].

To determine accurately and precisely the network structure and its relevant parameters such as connecting weights, some evolutionary algorithms such as genetic algorithm (GA), particle swarm optimization, hybrid genetic algorithm and particle swarm optimization, stochastic particle swarm optimization, unified particle swarm optimization, back propagation (BP), pruning algorithm, and simulated annealing can be used for this determination [17–34]. There are some unavoidable disadvantages of the above mentioned approaches channeling researchers to put forth great attempts to propose more potential algorithms. For instance, BP (as a gradient descend method) is a very popular optimization method, but it is plagued by slow convergence and susceptibility to local minima, while GA has its inherent disadvantages of the pre-maturity and the unpredictability of the result. Hence, other approaches to improve NN training introduced. Recently, a new algorithm has been proposed by Atashpaz-Gargari and Lucas [35] that has inspired from a socio-political phenomenon, called Imperialist Competitive Algorithm (ICA). This method has two great aspects; (a) high ability of this algorithm to search the global optimization even when facing with nonlinear optimization problems and (b) fast convergence speed.

In this paper, potential application of feed-forward Artificial Neural Network (ANN) gaining from ICA as a great optimizing algorithm is suggested to forecast the oil flow rate. The ICA is implemented in this current study to decide on initial weights of the factors employed in neural network. The developed ICA-ANN model is examined with using experimental results of an oilfield located in the Persian Gulf. Furthermore, results acquired from the developed ICA-ANN methodology were fully compared with the experimental oil flow rate data while the back propagation neural network, genetic algorithm-neural network, particle swarm optimization-neural network and Fuzzy model, using the same observed data, have been discussed in further details too.

## 2. Artificial neural networks

Neural networks, or more precisely artificial neural networks (ANNs), are a branch of artificial intelligence methods. Artificial neural networks were inspired by biological neural networks [36,37]. They were made of a large number of simple computing components, called nodes or neurons that arranged to form an input layer, one or more hidden layers and an output layer [38]. They further include interconnections between the nodes of successive layers through the so-called weights [36,39–42]. The role of these weights is to modify the signal carried from one node to the other and enhance or diminish the influence of the specific connection. Each neuron in the hidden layer receives weighted inputs from each neuron in the previous layer plus one bias [36,39–41,43]. The internal weights of the network are adjusted in the course of an iterative process termed training and the algorithm used for this purpose so-called training algorithm. The error back-propagation (BP) algorithm is the most common form of learning, utilized today in artificial neural networks. There exist many network architectures but Multilayer Perception is the most popular among them. The number of nodes in the feed forward neural network input layer is equal to the number of inputs in the process, whereas the number of output nodes is equal to the number process output [19,39,41,44].

Basically, the back-propagation training procedure is intended to obtain an optimal set of the network weight, which minimizes an error function. The commonly employed error function is the mean squared error (MSE) as defined by:

$$\mathrm{MSE} = \frac{1}{2}\sum_{k=1}^{G}\sum_{j=1}^{m}[Y_j(k) - T_j(k)]^2 \quad (1)$$

where $m$ is the number of output nodes, $G$ is the number of training samples. $Y_j(k)$ and $T_j(k)$ are the expected output, and the actual output, respectively. Generally, minimizing the MSE is the priority of training an ANN.

In this article, the available data are divided into two assortments including the training data and the testing data. The internal weights and biases are calculated from training set. Training has been continued whenever error on the validation set starts to increase. Test set is used to evaluate neural network performance. ANNs have been used for many chemical and petroleum engineering applications such as steady state and dynamic process modeling, process identification, yield maximization, nonlinear control and fault detection and diagnosis [19,39,41,44].

In this study, an artificial neural network was implemented to construct a predictive model to forecast the oil flow rate of the reservoir. Based on expert knowledge of the first author and statistical performance criteria, the best ANN architecture was: 2–7–1 (2 input units, 7 hidden neurons, 1 output neuron) [24–34]. Artificial neural network model was trained by implementing back propagation procedure according to Levenberg–Marquardt algorithm to estimate oil flow rate by presenting two parameters (pressure ($P$), temperature ($T$)) as inputs of the model. Sigmoid and linear functions are assigned the transfer functions in hidden and output layers, respectively.

## 3. Fuzzy logic

To discuss about real and current issues of the world we do live in, it is vitally needed to consider precisely and accurately the related ambiguity, uncertainty and vagueness [45–49]. This matter can be done with a high level of quality by fuzzy logic which was initially introduced by Zadeh [50]. This theory is an extension of binary logic [51–54]. In fuzzy set theory every object is a matter of degree which is indicated by membership functions (MFs) [56,58–62]. The amount of degree is always between 0 and 1 [63,64]. In this research there are 3 involved parameters which are pressure ($P$) and temperature ($T$) as the inputs and flow rate ($q$) as the output. In order to apply fuzzy logic in this case, primarily, these factors must be characterized flexibly by suitable membership functions. To do this step, 1100 of 1600 series of data were selected randomly as the training part and the rest of them (500 series) to test the model. During doing researches relevant to fuzzy logic, it is extremely crucial to consider he facilitation of a smooth transition between the real world and the fuzzy model, so to fulfil this principle it was decided to use piecewise linear MFs which can make an appropriate connection between the nature of the real problem and the fuzzy model [65–68]. The general functional form of a piecewise linear MF is given by (Fig. 1):

$$\mu_j(x)\begin{cases} 0, & x \le \alpha j, 1, \\ \dfrac{x-\alpha_{j,1}}{\alpha_{j,2}-\alpha_{j,1}} & \alpha j, 1 < x < \alpha j, 2 \\ \dfrac{x-\alpha_{j,2}}{\alpha_{j,3}-\alpha_{j,2}} & \alpha j, 2 < x < \alpha j, 3 \\ 1, & x = \alpha_{j,3} \\ \dfrac{x-\alpha_{j,3}}{\alpha_{j,4}-\alpha_{j,3}}, & \alpha j, 3 < x < \alpha j, 4 \\ \dfrac{x-\alpha_{j,4}}{\alpha_{j,5}-\alpha_{j,4}}, & \alpha j, 4 < x < \alpha j, 5 \end{cases} \quad (2)$$

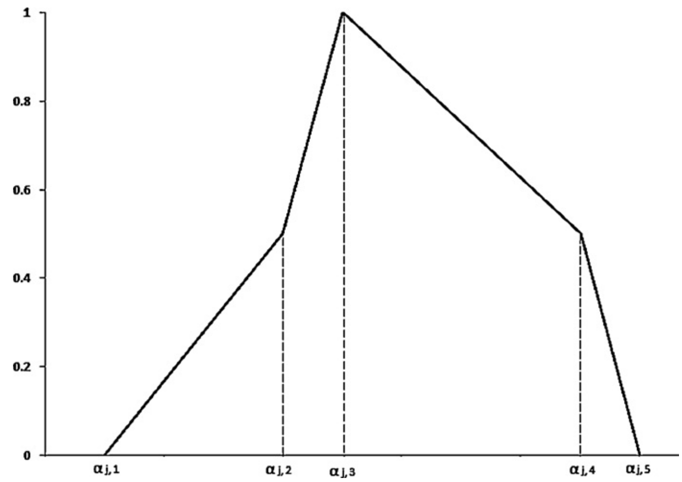


**Fig. 1.** General form of linear piecewise membership function.

**Table 1**
The relevant information on the three clusters for flow rate.

| Description | Low ($L$) | Medium ($M$) | High ($H$) |
|---|---|---|---|
| Interval | $L \le 3190$ (367) | $3190 \le M \le 10109$ (367) | $10109 \le$ H (366) |

where $[\alpha_{j,1}, \alpha_{j,2}, \alpha_{j,3}, \alpha_{j,4}, \alpha_{j,5}]$ are the designing values for the MF associated with the $j$-th linguistic term of an attribute. Thus, the associated support for the $j$-th linguistic term is indicated by the interval $[\alpha_{j,1}, \alpha_{j,5}]$.

Normally, the constructions of MFs are proposed by experts' knowledge, which can be different in production engineering problems. To defeat this obstacle, it was decided to construct the MFs automatically [69–73]. Hence, in the current paper, there are 1100 series of data to define the MFs and consequently, model fuzzy decision tree (FDT). Each attribute is divided to 3 subsets. It follows that three intervals are identified for an initial partitioning of each attribute, defined as Low ($L$), Medium ($M$) and High ($H$) levels. Using a simple equal-frequency partitioning (clustering) method was carried out in which the middle point between intervals in different clusters is defined as the initial interval boundary point. In other words, there are 1100 series of data for the training part which have been divided to main 3 parts based on this fact that in each set there must virtually be the equal number of data [74] (Table 1).

After defining the intervals for each of attributes, it is necessary to construct MFs. In order to do this step automatically, the suggested approach is firstly to make estimated distributions in the form of probability density functions (pdfs) for each of the intervals of the decision attribute; with the link between probability distributions and MFs (through possibility distributions) which has been an active subject matter since the research of Zadeh [65,75]. Next, instead of holding full details connected to pdf expressions, in this paper only the centre of area (C-o-A) values for each pdf and the interval boundary values, which were found previously, are used as the key parameters to set MFs [76].

To build the estimated distributions for each of the decision classes, the procedures of [77,78] was exactly followed as bellow and the ensuing results were gained.

$$\mathrm{pdf}_j(x) = \frac{1}{\sqrt{2m_j\pi(\max(I_j)-\min(I_j))}}\sum_{i=1}^{m_j}\exp\left[-\frac{m_j}{2}\left(\frac{x-x_i}{\max(I_j)-\min(I_j)}\right)^2\right] \quad (3)$$

where $m_j$ is the number of values in $I_j$ which is the interval related to the linguistic term and $x_i$ represents the value of each data in

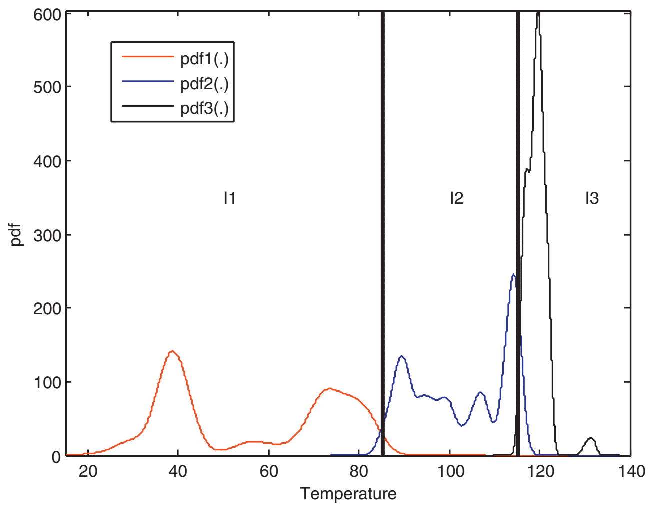


**Fig. 2.** Estimated distribution related to intervals I1, I2, I3 for temperature.

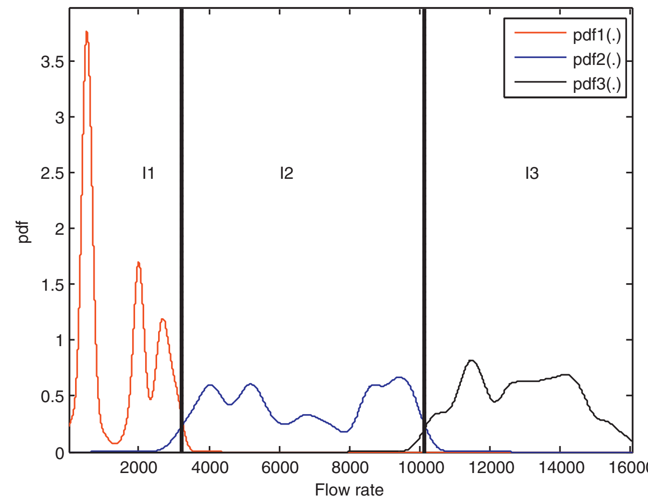


**Fig. 4.** Estimated distribution related to intervals I1, I2, I3 for flow rate.

the interval and $x$ is the variable. In the demystified mode, the pdf for each interval which is directly connected to one linguist term is determined through firing the above function for the total domain of the assumed parameter while the data that had already dedicated to the supposed interval through dividing all data to main 3 parts must one by one be included to the formula. In other word, during the programming of this function it is necessary to use two "for" loops which the firs one is for number of data which are in the considered interval and is currently shown by $m_j$ and the second one is for total domain of the parameter (Figs. 2–4).

Consequently, according to the boundaries and C-o-As, the MFs were calculated (Figs. 5–7).

### 3.1. Description of the inductive fuzzy decision tree method

The applied methods results in producing series of fuzzy rules which present the aforementioned parameters in linguistic terms [79]. Its advantages can be summarized in measuring cognitive uncertainties and minimizing classification ambiguity when fuzzy evidences are present (Fig. 8) [77]. A membership function $\mu(x)$ of a fuzzy variable $Y$ defined on $X$, can be assumed as a possibility

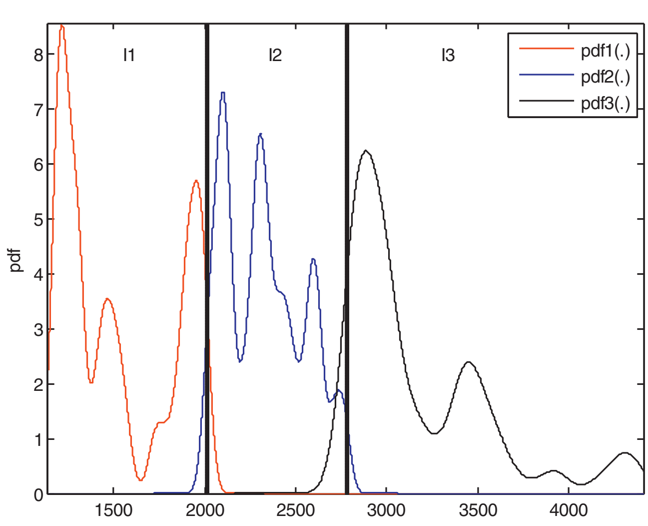


**Fig. 3.** Estimated distribution related to intervals I1, I2, I3 for pressure.

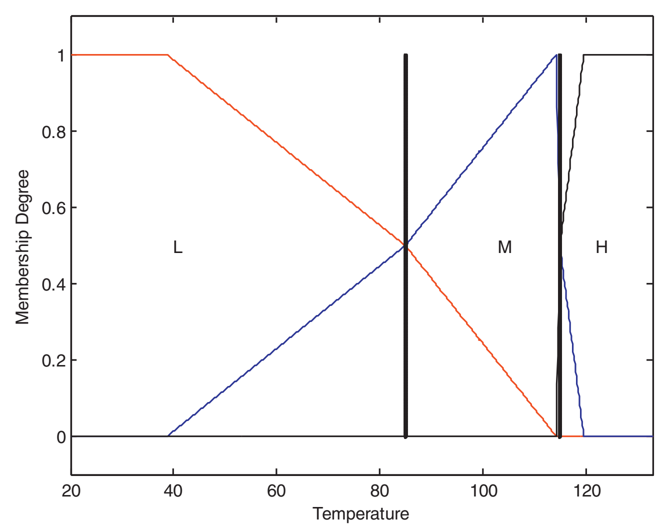


**Fig. 5.** Membership functions of the decision classes *L*, *M*, and *H* connected with temperature.

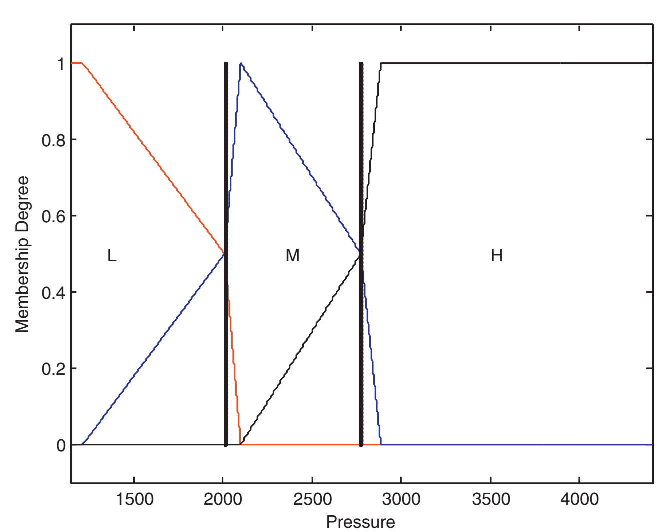


**Fig. 6.** Membership functions of the decision classes *L*, *M*, and *H* connected with pressure.

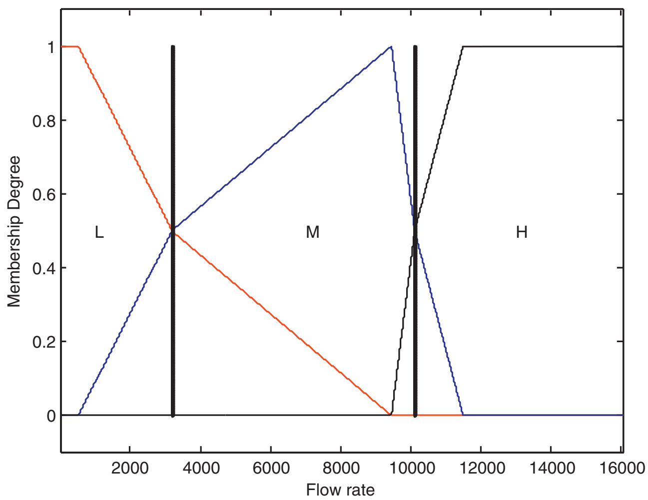


**Fig. 7.** Membership functions of the decision classes $L$, $M$, and $H$ connected with flow rate.

distribution of $Y$ on $X$, that is $\pi(x) = \mu(x)$, for all $x \in X$. The possibility measure $E_\alpha(Y)$ of ambiguity is defined as:

$$E_\alpha(Y) = g(\pi) = \sum_{i=1}^{n} (\pi_i^* - \pi_{i+1}^*)\ln i, \tag{4}$$

where $\pi^* = \{\pi_1^*, \pi_2^*, \ldots, \pi_n^*\}$ is the permutation of the possibility distribution $\pi = \{\pi(x_1), \pi(x_2), \ldots, \pi(x_n)\}$ sorted so that $\pi_j^* \geq \pi_{i+1}^*$ for $i = 1,..., n$ and $\pi_{n+1}^* = 0$ [80,81]. The ambiguity of attribute $A$ is then calculated as:

$$E_\alpha(A) = \frac{1}{m}\sum_{i=1}^{m} E_\alpha(A(u_i)) \tag{5}$$

where $m$ is the number of cases and term of $E_a$ $(A(u_i))$ has been discussed in the same paragraph. The formula has also roots in the formula No. (4) in which the dedicated degrees by MFs for each set must firstly be normalized and then $E_a(Y)$ is generated. After that, through the formula No. (5) an averaging operation is done to produce $E_a(A)$. Where $E_\alpha(A(u_i)) = g(\mu T_s(u_i)/\max_{1\leq j\leq s}(\mu T_j(\mu_i)))$, with $T_j$ the linguistic scale used within an attribute for $m$ cases. Overlapping exist when there is overlapping between linguistic terms. The degree which $A$ is a subset of $B$ is measured through fuzzy subset hood $S(A,B)$ and is given by [82]:

$$S(A, B) = \frac{\sum_{u\in U} \min(\mu_A(u), \mu_B(u))}{\sum_{u\in U}(\mu_A(u)} \tag{6}$$

which all attributes are over the same set of objects.

The possibility of classifying an object to class $C_i$, according to the given fuzzy evidence $E$ can be defined as:

$$\pi = (C_i|E) = \frac{S(E, C_i)}{\max_j S(E, C_j)} \tag{7}$$

where $S(E,C_i)$ performs the degree of truth for the classification rule (that is $E \Rightarrow C_i$). The classification ambiguity based on a single piece of evidence (a fuzzy value for an attribute) is defined as:

$$G(E) = g(\pi(C|E)) \tag{8}$$

$G(P|F)$ Is the classification ambiguity with fuzzy partitioning $P = \{E_1, \ldots, E_k\}$ on the fuzzy evidence $F$ which is the weighted average of classification ambiguity with each subset of partition:

$$G(P|F) = \sum_{i=1}^{k} w(E_i|F)G(E_i \cap F) \tag{9}$$

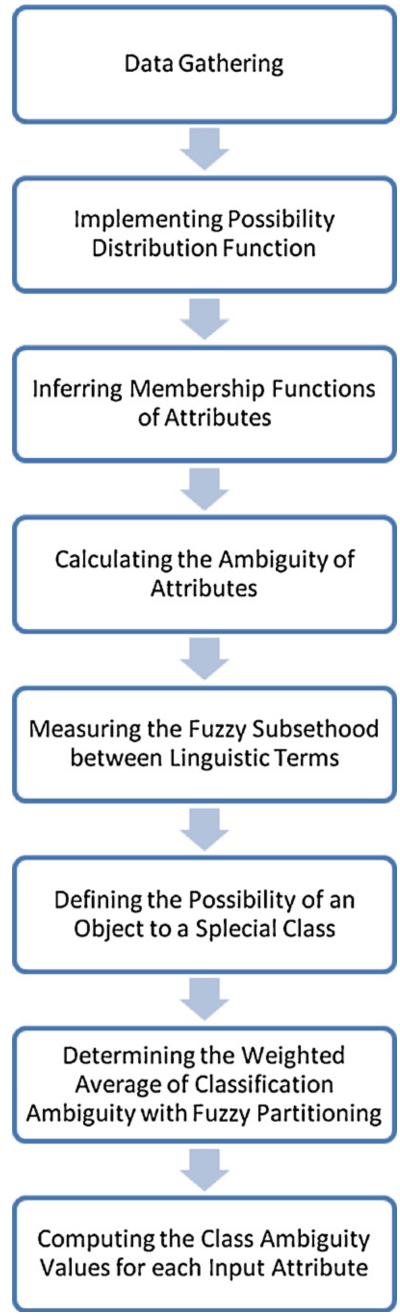


**Fig. 8.** Flowchart of Fuzzy Decision Tree.

where $G(E_i \cap F)$ is the classification ambiguity with fuzzy evidence $E_i \cap F$, and where $w(E_i|F)$ is the weight $(w(\cdot))$which indicates the relative size of subset $E_i \cap F$ in $F$:

$$w(E_i|F) = \frac{\sum_{u\in U} \min(\mu_{E_i}(u), \mu_F(u))}{\sum_{j=1}^{k}\left(\sum_{u\in U} \min(\mu_{E_i}(u), \mu_F(u))\right)} \tag{10}$$

In short, depend on the lowest level of ambiguity attributes are assigned to nodes. A node is terminated (becomes a leaf node), if degree of subset hood (based on the intersection of the nodes from the root) is higher than a truth value β (in this current paper it has been equated to 0.7) [69,77,78].

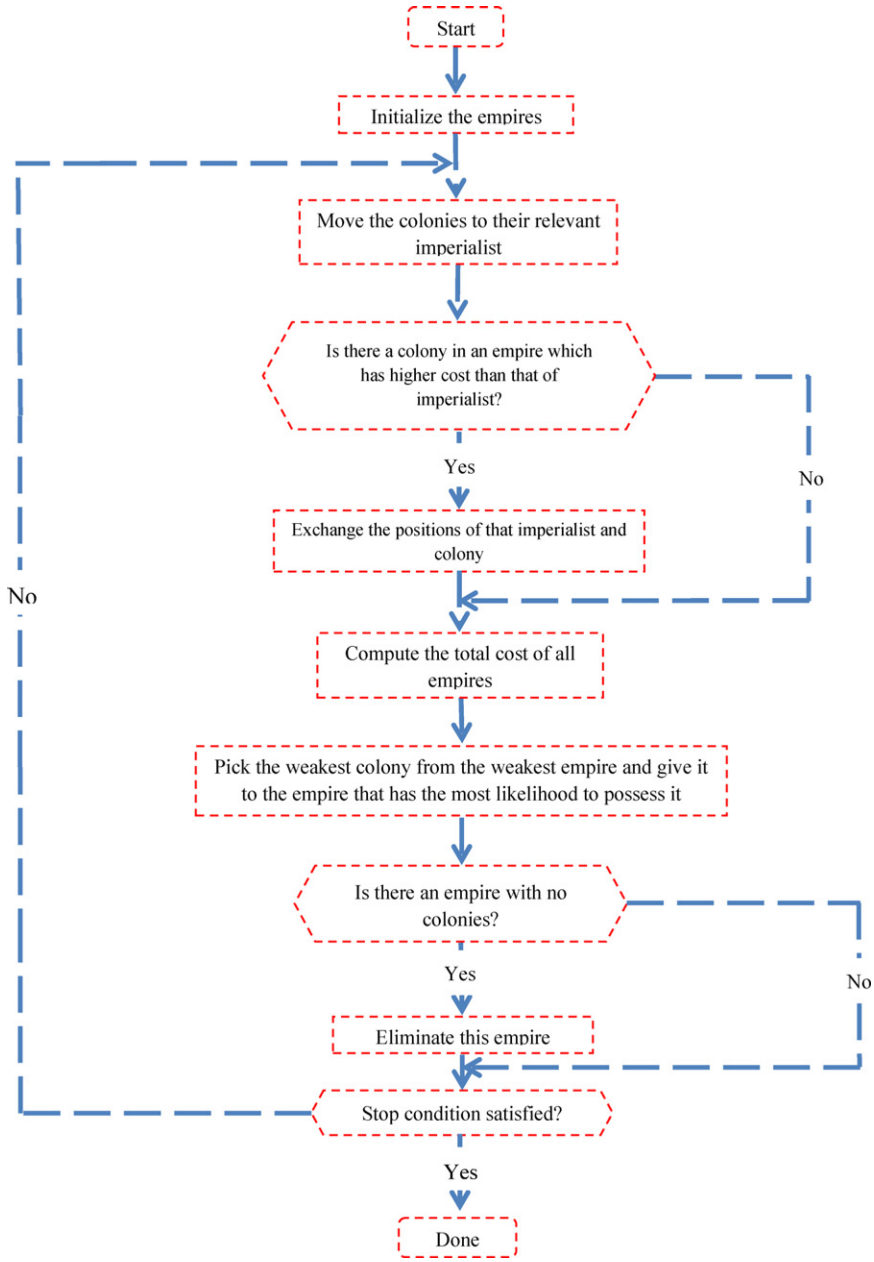


**Fig. 9.** Flowchart of the ICA algorithm [35].

## 4. Imperialist competitive algorithm

New evolutionary algorithm based on human being's socio-political evolution was expressed by Atashpaz-Gargari and Lucas [35] and named Imperialist Competitive Algorithm (ICA). Imperialist competitive algorithm (ICA) is population based algorithm and like other evolutionary algorithms the ICA starts with initial populations called countries. Countries split in two types; colony and imperialist (in optimisation terminology, countries with the least cost) which together form empires. Attempting of the imperialists to gain more colonies named imperialist competitive process. Due to competition process the powerful imperialists will be aggravated in the power and the vice versa about the weak ones. If an empire missed all of its colonies then an empire assumed to be collapsed. The most powerful imperialist remains in the world and all the countries are colonies of this unique empire at the end of algorithm. Power and position of the imperialist and colonies at this level of computation are the same (Fig. 9) [35]. The implementation procedures of our proposed matching strategy based on ICA are described as follows.

### 4.1. Generating initial empire

The end of optimization procedure is to forecast the best solution of the discussed issue in terms of its connected variables. These parameters create an array of variable values which is supposed to be optimized. Country in imperialist competitive algorithm (ICA) is corresponding to "chromosome "In genetic algorithm (GA) terminology. In an $N_{var}$ – dimensional optimization issue, a country is a $1 \times N_{var}$ array. This array is defined by:

$$\text{Country} = [P_1, P_2, P_3, \ldots, P_{N_{var}}] \tag{11}$$

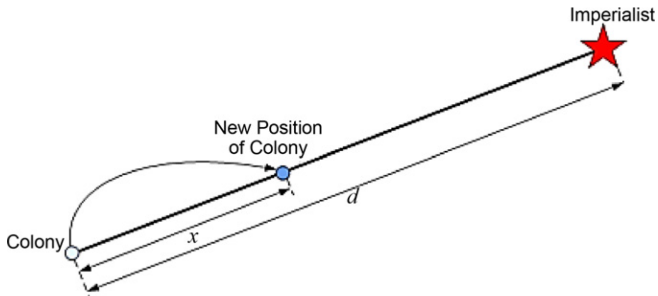


**Fig. 10.** Movement of colonies toward their relevant imperialist [35].

Cost function of any country was evaluated by following expression:

$$\text{Cost} = f(\text{country}) = f([P_1, P_2, P_3, \ldots, P_{N_{\text{var}}}]) \tag{12}$$

Different parameters such as number of initial countries ($N_{\text{country}}$), number of imperialist ($N_{\text{imp}}$) and a number of the remaining country are colonies that each belongs to an empire ($N_{\text{col}}$) were assigned to begin imperialist competitive algorithm. The initial colonies of an empire are in convenience with their empire [35]. To split proportionally the colonies among imperialists, the normalized cost of an imperialist is expressed as follow:

$$C_{\text{n}} = c_n - \max_i\{C_i\} \tag{13}$$

where $c_n$ is the cost of $n$-th imperialist and $C_{\text{n}}$ is its normalized cost. Through having the normalized cost of all imperialist, the power of each imperialist is expressed as follow [35]:

$$P_n = \left| \frac{C_{\text{n}}}{\sum_{i=1}^{N_{\text{imp}}} C_i} \right| \tag{14}$$

In the other side, the normalized power of an imperialist is determined based on its colonies. Then, the initial number of an imperialist will be [35]:(15)$N \cdot C_{\text{n}} = \text{round}\,\{P_n \cdot N_{\text{col}}\}$

Where $N{\cdot}C_{\text{n}}$ represents the initial number of colonies of $n$-th empire and $N_{\text{col}}$ is the number of all colonies. In order to split the colonies among imperialists, $N{\cdot}C_{\text{n}}$ of the colonies is randomly selected and assigned to each imperialist. The colonies together with the imperialist form the $n$-th empire [35].

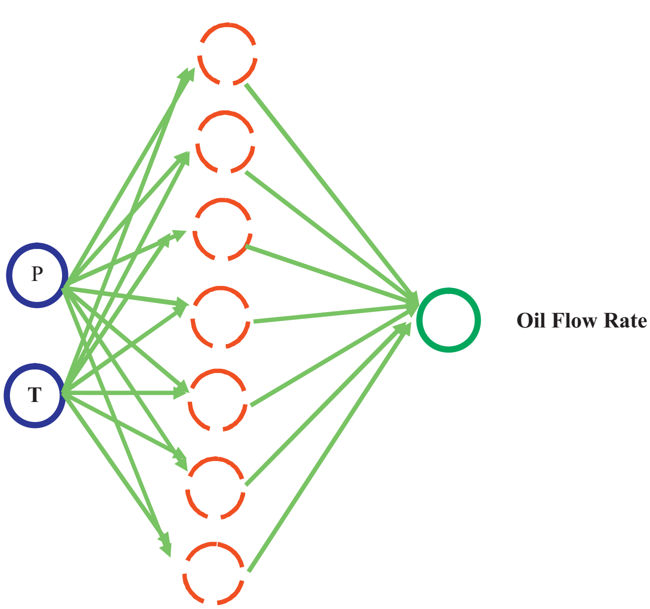


**Fig. 11.** Architecture of three layers ANN.

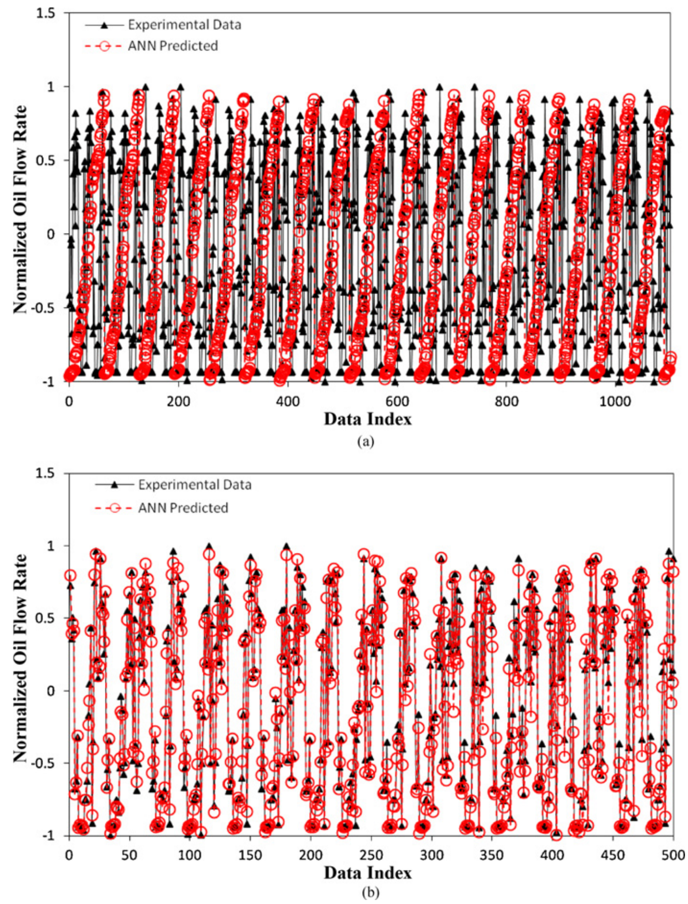


**Fig. 12.** Comparison between measured and predicted oil flow rate (ANN): (a) Train, (b) Test.

### 4.2. Moving colonies of an empire toward the imperialist

The imperialist countries try to enhanced their colonies and make them a part of their imperialists. Movement of all colonies towards their proper imperialist implemented to represent this fact. This movement behaviour is shown in Fig. 10. As can be seen from Fig. 10, the colony moves toward the imperialist by $x$ (is a random variable with uniform distribution) units [35].

$$x \sim U(0, \beta \times d) \tag{16}$$

where $d$ is distance between an imperialist and colony and $\beta$ is a number greater than 1. When a colony achieve a position with lower cost that of its imperialist, the imperialist and colony change their position. Henceforth, new position of imperialist which implemented through the algorithm and colonies start moving towards this position [35].

### 4.3. The total power of an empire

Power of the imperialist and its colonies have most impact on the total power of an empire. Due to this fact, total cost of an empire is obtained by following expression:

$$T \cdot C_{\text{n}} = \text{Cost}(\text{imperialist}_n) + \xi \text{mean}\{\text{cost}(\text{colonies of impire}_n)\} \tag{17}$$

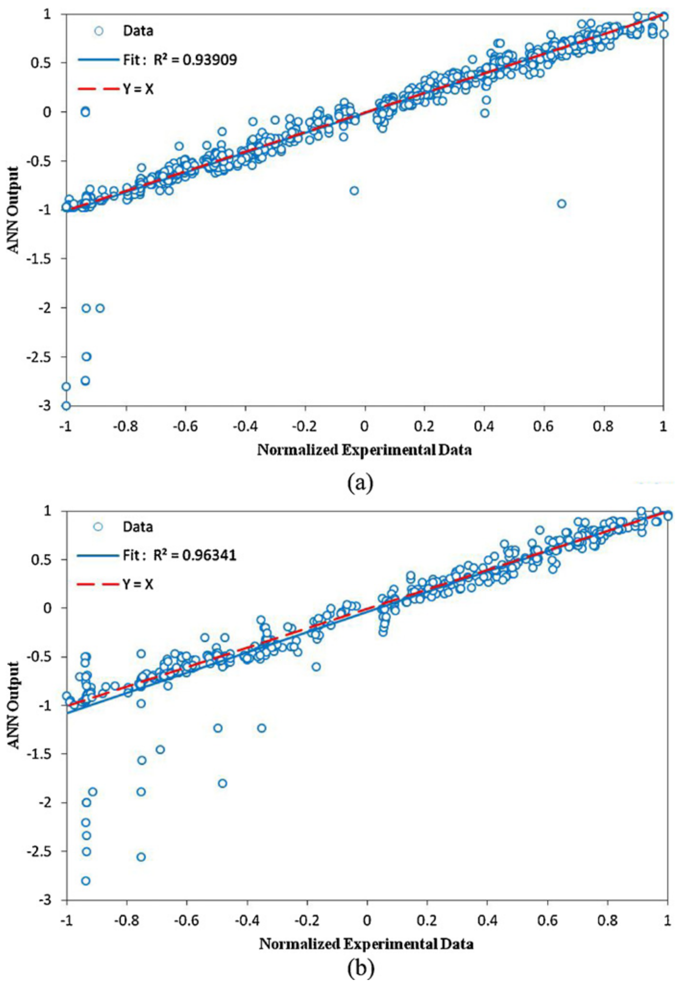


**Fig. 13.** The performance of predicted values of flow rate vs. the measured, (a) Test, (b) Train.

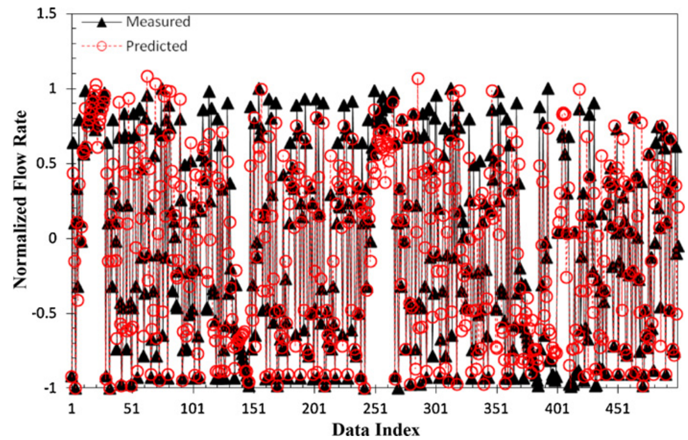


**Fig. 15.** The comparison between predicted and measured values of flow rate for each sample of testing set.

where the total cost of the $n$-th empire is assigned to $T{\cdot}C_n$, and $\xi$ is a positive number which is considered to be less than 1. A minuscule value for $\xi$ implies that the total power of an empire to be obtained by just the imperialist and increasing it will accentuate the importance of the colonies in determining the total power of an empire. The value of 0.1 for $\xi$ is an appropriate value in most of the implementations [35].

### 4.4. Imperialistic competition

All empires make efforts to gain colonies of the other empires and handle them. The imperialistic challenge gradually brings about a decrease in the power of weaker empires and an increase in the power of more vigorous ones. This challenge is approached by just selecting some (usually one) of the weakest colonies of the weakest empires and making a competition among all empires to possess this colonies [35].

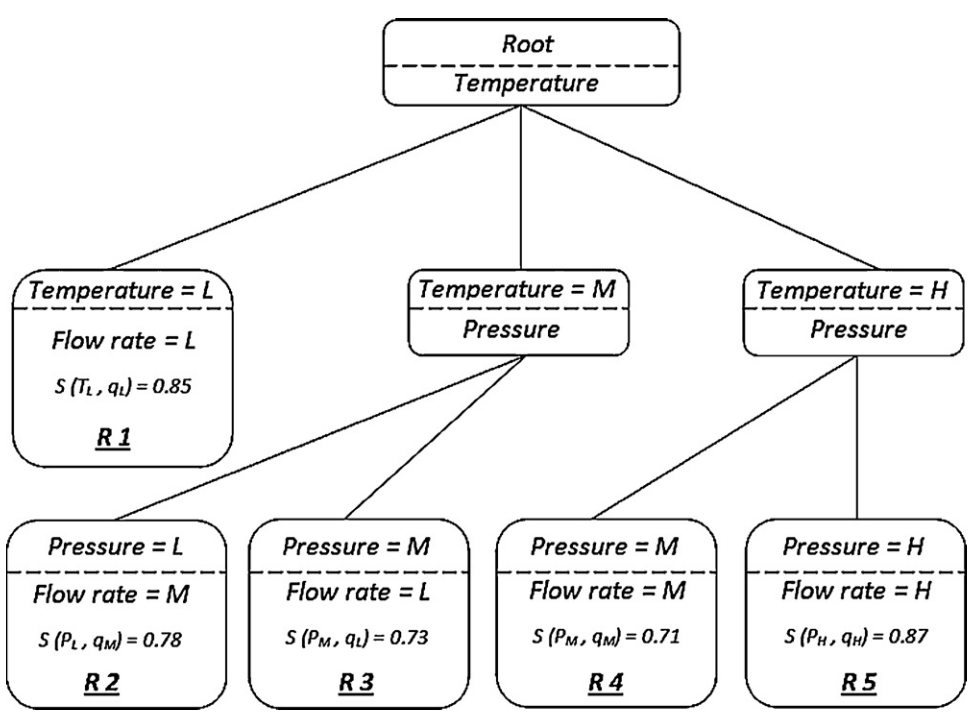


**Fig. 14.** Schematic structure of fuzzy decision for flow rate.

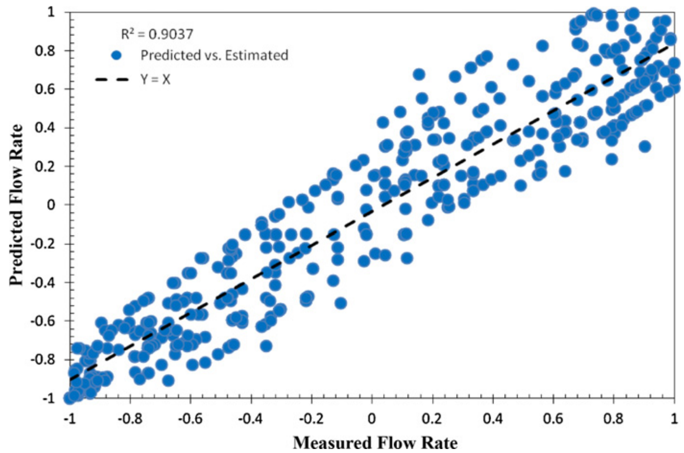


**Fig. 16.** The performance of predicted values of flow rate vs. the measured.

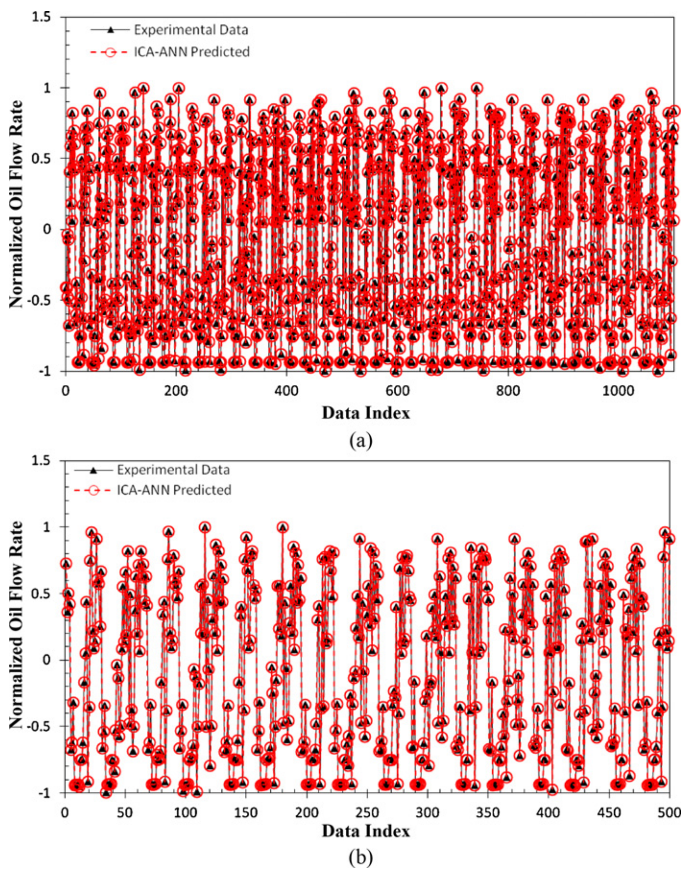


**Fig. 17.** Comparison between measured and predicted oil flow rate (ICA-ANN): (a) Train, (b) Test.

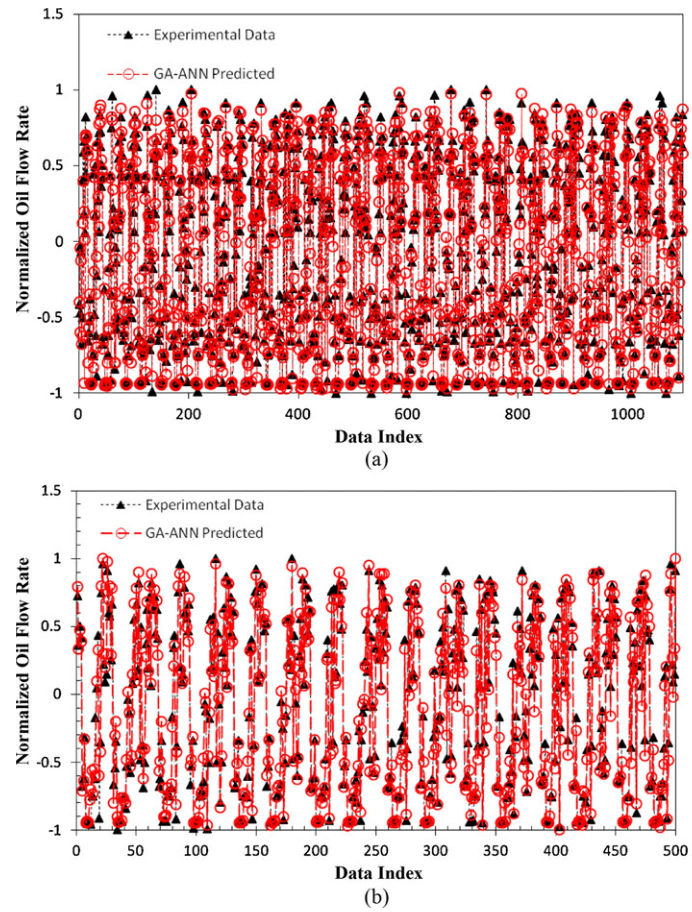


**Fig. 18.** Comparison between measured and predicted oil flow rate (GA-ANN): (a) Train, (b) Test.

The possession probability of each empires figure out based on its total power while start the competition process. The normalized total cost is determined by following expression [35]:

$$N \cdot T \cdot C_n = T \cdot C_n - \max_i\{T \cdot C_i\} \tag{18}$$

where the total cost and normalized total cost of $n$-th empire are assigned to $T{\cdot}C_n$ and $T{\cdot}C_n$, respectively. While having the normalized total cost, the possession probability of each empire is obtained by following equation [35]:

$$P_{p_n} = \left| \frac{N \cdot T \cdot C_n}{\sum_{i=1}^{N_{imp}} N \cdot T \cdot C_i} \right| \tag{19}$$

To distribute introduced colonies among empires, vector **P** is formed as:

$$\mathbf{P} = [P_{p_1}, P_{p_2}, P_{p_3}, \ldots, P_{p_{N_{imp}}}] \tag{20}$$

Then, vector which numbers randomly created and uniformly distributed with same size vector as **P** named vector **R** as follow [35],

$$\mathbf{R} = [r_1, r_2, r_3, \ldots, r_{N_{imp}}] \tag{21}$$

After that, by subtracting **R** from **P** vector **D** is created:

$$\mathbf{D} = \mathbf{P} - \mathbf{R}[D_1, D_2, D_3, \ldots, D_{N_{imp}}] \tag{22}$$

Referring to vector **D**, the mentioned colony (colonies) is controlled to an empire whose correspond index in **D** is maximized [35].

Powerless empire will encounter in the imperialistic competition process and their colonies will be split between other empires. At the end of the imperialist competitive algorithm, all the empires except the most powerful one will encompass and all the colonies will be handling by this unique empire. It should be noted that here, imperialist and colonies have the same position and power in this stage [35].

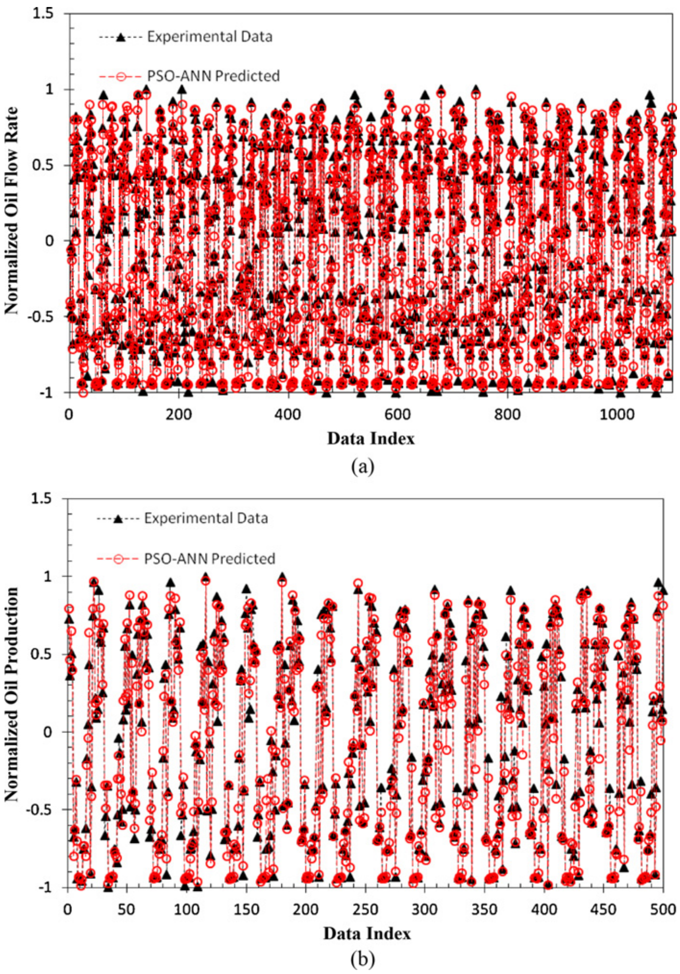


**Fig. 19.** Comparison between measured and predicted oil flow rate (PSO-ANN): (a) Train, (b) Test.

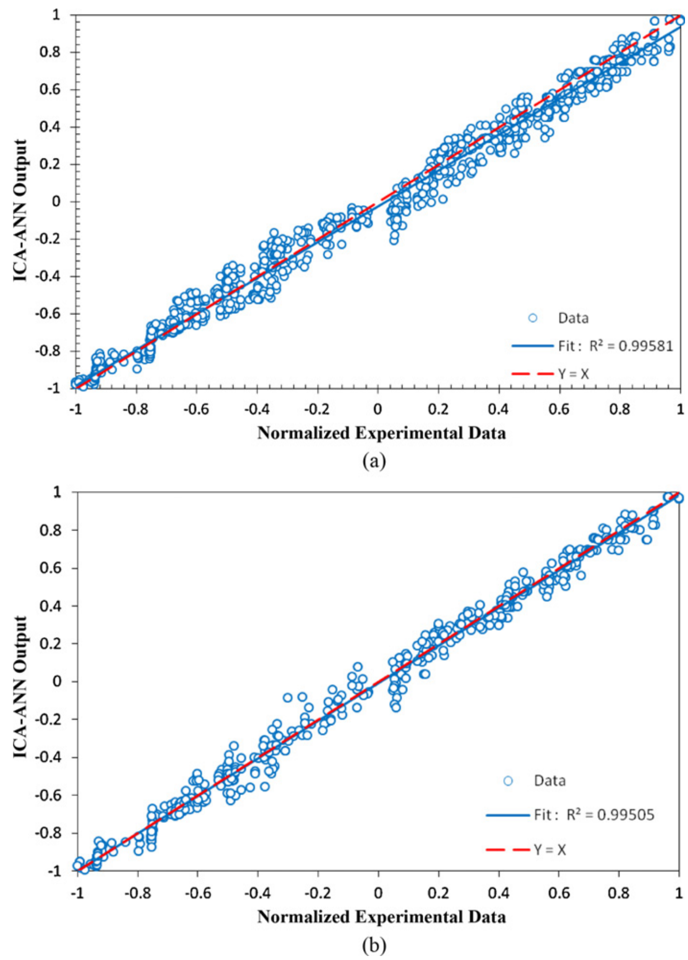


**Fig. 20.** The performance of predicted values of flow rate vs. the measured, a) Train b) Test.

## 5. Case study

One of northern Persian Gulf oil reservoir is a strongly folded anticlinal structure about 32 km long and varies in width from 2.5 km to 5 km at the original water oil contact (WOC). The folding of the Southwest flank is somewhat steeper than that of the Northeast flank due to the north-eastern direction of the thrust which caused the folding, The original GOC in the main central dome was at 1015 feet subsea, and the average original WOC of the two flanks was at about 3087 feet subsea. More than 50 wells have been drilled in this structure. In this research we implemented 1600s production data set of these wells to develop an intelligent approach to estimate oil production flow rate of the petroleum reservoir. To defeat this obstacle, we implemented two production variables such pressure ($P$) and temperature ($T$) as an inputs of introduced predictive model.

## 6. Results and discussions

### 6.1. ANN model

In this research, an artificial neural network (ANN) was carried out to construct a predictive model to estimate the amount of oil flow rate of reservoir. One of the most important parameter in obtaining high performance predictive model based on artificial neural network (ANN) is Data selection. Therefore, to achieve high performance neural network model a good training data set which can represent all of the possible conditions is required (Fig. 11).

Quality and accuracy of the implemented data have a crucial role on neural network performance. Hence, implemented data in artificial neural network training should have the maximum precision and the minimum uncertainty. To give high precision and accurate predictive model, an implemented dataset should be represent all properties of the modeled system behavior such as system complexity, restrictions and etc. [83]. In this text, it is assumed that the implemented data which carried out to build predictive model is statistically representative of the system behavior.

Different parameters such as problem type, system complexity, probable relations between the parameters, user experience, precision, neural network type, and the time consumption of solve have affect on data numbers in artificial neural network. Randomly selection of implemented data is prevalent in ANN implications. In such cases, neural network cannot be trained sufficiently due to very low proportion of the training data. Therefore, obtained outputs of neural network model will not be acceptable. Hence, presenting a sufficient portion of data to model can warranted the accuracy and precision of the training and testing phases in neural network model. Also it should be noted that here, If a small proportion of data for testing phase is selected, the results testing phase cannot be valid. This is due to the fact that sufficient data with different ranges should be picked up to confirm that the network is working well. In testing phase new data set was implemented to introduce into network model to figure out accuracy and robustness of model. It should be mentioned here; new data set in testing phase was not

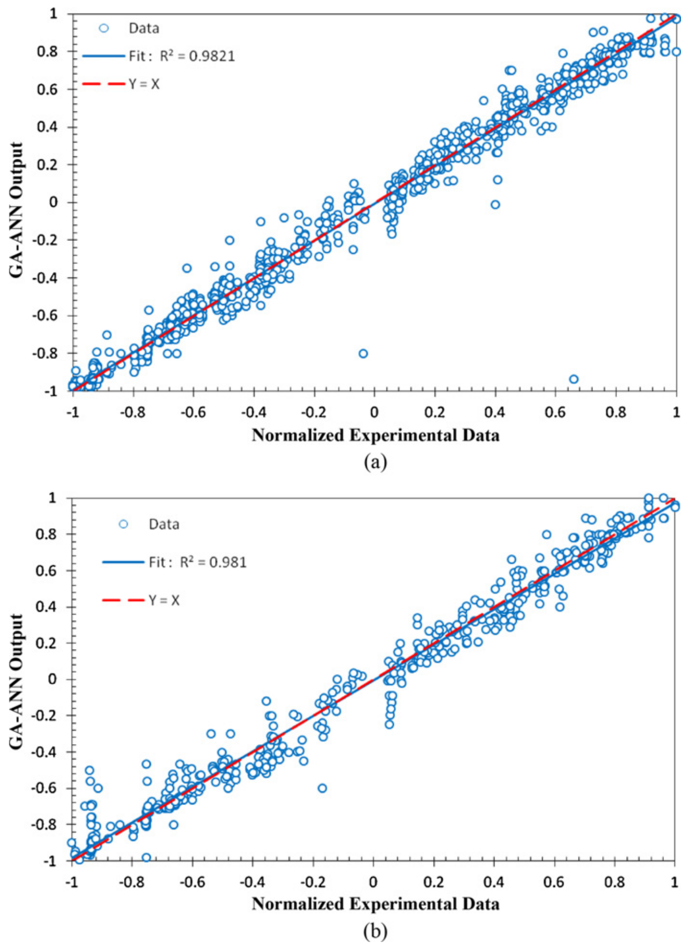


**Fig. 21.** The performance of predicted values of flow rate vs. the measured, (a) Train, (b) Test.

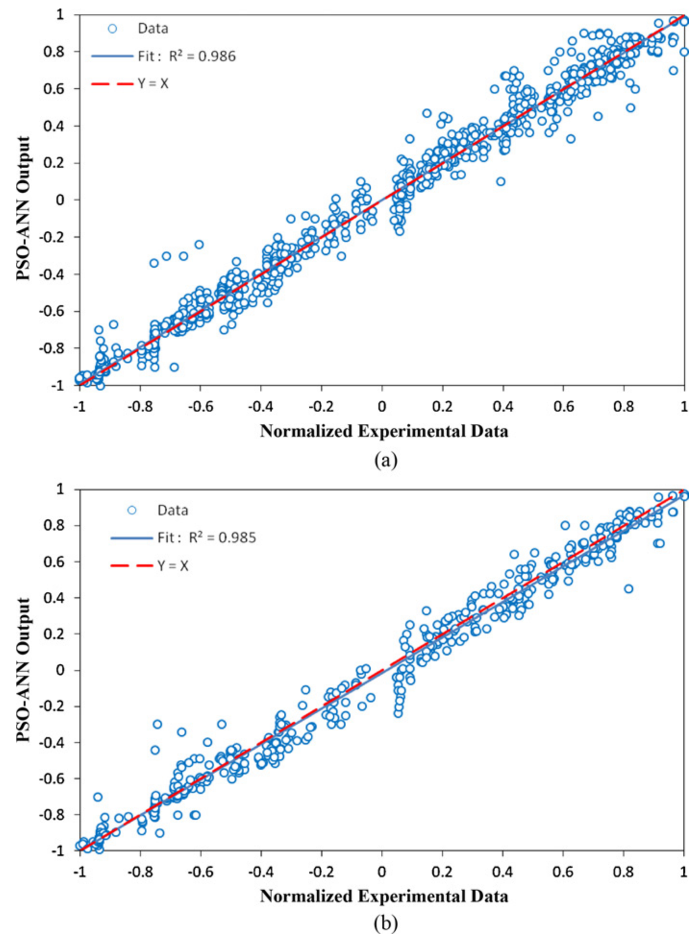


**Fig. 22.** The performance of predicted values of flow rate vs. the measured, (a) Train, (b) Test.

**Table 2**
Class ambiguity values for each input attribute.

| Attribute | Temperarure | Pressure |
|---|---|---|
| $G(E)$ | 0.1932 | 0.5264 |

implemented in training phase. Therefore, the data implemented in this study were split into two distinguished parts: training (1100 data set of all data) and testing (500 data set of all data). The oil flow rate prediction of the reservoir in the training and test phase are shown in Fig. 12. Fig. 13 show the extent of the match between the measured and predicted oil flow rate values by artificial neural network in terms of a scatter diagram.

### 6.2. Fuzzy model

After calculating the mentioned method, the succeeding results obtained as shown below (Table 2) (Fig. 14):

After completing the construction of FDT and extracting rules, 500 series of data in the set of testing were feed to the model and $R$-square of 0.898 was calculaterd for responses (Figs. 15 and 16).

### 6.3. ICA-ANN model

Imperialist competitive algorithm (ICA) was implemented in this research to optimize the connection weights of the introduced neural network system. To assess this end, the connection weights of the neural network were chosen as variables of an optimization issue. It should be noted that here, every weight in the network was initially set in the range of $[-1, 1]$. Also, The Mean Square Error (MSE) was carried out as a cost function in the imperialist competitive algorithm. Back propagation is a gradient descent algorithm on the error space which most willingly gets deceive into a local optimum making it entirely dependent on initial (weight) settings. To overcome this obstacle different optimization algorithm such as genetic algorithm (GA), particle swarm optimization (PSO) and imperialist competitive algorithm (ICA) can be implemented due to global searching ability of them. Minimizing of the cost function is the main end of the imperialist competitive algorithm which implemented in this study. In the simulations conducted, the number of imperialists and the colonies considered are 4 and 40, respectively; and the parameter $\beta$ was set to 2.

Two different statistical indexes were implemented to figure out performance and accuracy of each model. The most popular statistical indexes which implemented to this research are mean square error (MSE) and correlation coefficient ($R^2$) (see Appendix A). Generally, lowest value of mean square error and highest value of correlation coefficient is desirable for selecting best model. Reportedly, for $R^2$ value greater than 0.9 illustrates best efficiency of model, while a $R^2$ value between 0.8 and 0.9 indicates a good performance and for value less than 0.8 demonstrate an unsatisfactory efficiency. To show strength of introduced model (ICA-ANN), different popular optimization algorithms such as particle swarm optimization (PSO) and genetic algorithm (GA) was implemented to optimize connection weights of neural network and results of these implementations are shown in Figs. 17–22. Forecasted oil flow rate of the reservoir in training and testing stages of ICA-ANN, GA-ANN

**Table 3**
Comparison between the performances of ICA-ANN and ANN model.

| | ICA-ANN | PSO-ANN | GA-ANN | ANN | Fuzzy |
|---|---|---|---|---|---|
| MSE | 0.0030392 | 0.012613 | 0.04516 | 0.091343 | 0.0073664 |
| $R^2$ | 0.99505 | 0.985 | 0.981 | 0.93909 | 0.9037 |

and PSO-ANN models are shown in Figs. 17–19, respectively. Also extent of the match between the experimental and forecasted oil flow rate of ICA-ANN, GA-ANN and PSO-ANN models are shown in Figs. 20–22, respectively. As can be seen in Figs. 20–22, ICA-ANN model has best accuracy and precision in forecasting oil flow rate of the reservoir in comparison with GA-ANN, PSO-ANN and fuzzy logic approach. Mean square error (MSE) and correlation coefficient ($R^2$) of five different models in testing phase are reported in Table 3.

## 7. Conclusions

A hybrid imperialist competitive algorithm and artificial neural network algorithm which effectively combines the local searching ability of the back propagation (BP) method with the global searching ability of ICA was implemented for oil flow rate prediction. Experimental data from a oil reservoir were used to examine the proposed ICA-ANN model. Based up on the results obtained from this study the following main conclusions can be drawn:

In this work, we have expressed an Imperialist Competitive algorithm evolved artificial neural network. In Our methodology, Imperialist Competitive algorithm was implemented to optimized connection weights of proposed neural network, while combines the local searching ability of the imperialist competitive algorithm and global searching ability of neural network. To figure out robustness of the imperialist competitive algorithm in optimizing connection weights of neural network must popular optimization algorithms such as particle swarm optimization and genetic algorithm were implemented. To prevent premature convergence, the imperialist competitive parameters are carefully assigned to optimize the artificial neural network. In comparison between introduced ICA-ANN model and other intelligent models, the same experimental data has performed in each predictive model and based on correlation coefficient and mean square error, performance of the proposed ICA-ANN model is better than that of the BP-ANN, GA-ANN, PSO-ANN and fuzzy logic approaches. The determination of the optimal neural network topology is an obstacle when considering the combination of neural networks and imperialist competitive algorithm for oil flow rate prediction. Based on statistical indexes were introduced in this study and try and error method our neural network topology is determined manually. Optimization of artificial neural network topology by implementing different optimization algorithms such as imperialist competitive algorithm will be a part of our future work.

## Acknowledgment

Special thanks go to Dr. Nazanin Ahmadi at Sharif university of Technology for her continuous support and fruitful discussions about fuzzy approach.

## Appendix A.

To evaluate performance of presented model to forecast oil flow rate of the reservoir two must popular statistical parameters such $R^2$ and MSE were implemented through this research. The expression to evaluate the above parameters and also the corresponding description are as follows.

Statistical index which gives information on goodness of fit in a model is represent by $R^2$ value. According to regression analysis knowledge or matching of a model, the correlation coefficient ($R^2$) is a statistical criterion to display how well the model line approaches the actual data points. It is clear that for $R^2$ of 1.0 illustrates the model line perfectly match the actual data. Mathematical representation of the $R^2$ is described as follow [87,88]:

$$R^2 = \frac{\sum_{i=}^{n}(\text{Flow}_i^{\text{P}} - \overline{\text{Flow}}^{\text{M}})^2}{\sum_{i=}^{n}(\text{Flow}_i^{\text{M}} - \overline{\text{Flow}}^{\text{M}})^2} \tag{A1}$$

where $\text{Flow}^{\text{M}}$ and $\text{Flow}^{\text{P}}$ are the measured flow rate and predicted flow rate, respectively. $\overline{\text{Flow}}^{\text{M}}$ represents the average of the measured flow rate data assortment.

Represent of how close a fitted line or developed model is to data points could be figure out by the Mean Squared Error (MSE) criteria. As know in statistical analysis, high precise and accurate model has an infinitesimal Mean Squared Error while estimated value is closed to relevant measured or actual value. Mean square error (MSE) is represented by the following expression [87,88]:

$$\text{MSE} = \frac{\sum_{i=1}^{n}(\text{Flow}_i^{\text{M}} - \text{Flow}_i^{\text{P}})^2}{n} \tag{A2}$$

where $n$ is the number of samples.

## References

[1] H.O. McLeod Jr., M.A. Pashen, 'Well-completion audits to evaluate gravel-packing procedures, SPE Drilling & Completion 12 (4) (1997) 228–237.
[2] S.M. Tariq, L.C. Ayestaran, Analyses and applications of pressure, flow rate, and temperature measurements during a perforating run, SPE Production Engineering 6 (1) (1991) 83–92.
[3] B. Ruslan, F.T. Brian, A.V. Amian, Calculation of Pressures for Foams in Well Completion Processes, SPE Annual Technical Conference and Exhibition, 6–9 October 1996, Denver, Colorado USA, 1996.
[4] B.D. Poe Jr., W.K. Atwood, J. Kohring, K. Brook, Characterization of Reservoir Properties Using Production Logs, International Oil & Gas Conference and Exhibition in China, 5–7, December, 2006, Beijing, China, 2006.
[5] B.D. Poe Jr., W.K. Atwood, J. Kohring, K. Brook, in: Evaluation of Commingled Reservoir Properties Using Production Logs, First International Oil Conference and Exhibition in Mexico, 31 August–2 September, 2006, Cancun, Mexico (2006).
[6] Ferreira, A., Vidal, J., Abi-Ramia, R., Jouti, I., Flores, S., Cañas, J., Rey, F., Landinez, G., Maizeret, P., Villanueva, G., Cabanilla, J., Alves, R., 2012. Innovative Work Scheme: Real Time Analysis During Well Test Operations, SPE Intelligent Energy International, 27–29 March 2012, Utrecht, The Netherlands.
[7] K.M. Sundaram, N. Kumar, V. Nangia, V. Desphande, S. Nahar, R. Jackson, Estimating Well Deliverability in Deepwater Wells With Interval Pressure Transient Testing (IPTT): A Case Study, SPE Oil and Gas India Conference and Exhibition, March 28–30 2012, Mumbai, India, 2012.
[8] Al Yateem, S.K., Al Amri, N., 2012. Moving toward Intelligent Field Applications: MPFM for Production Rate Testing and Beyond, SPE Intelligent Energy International, 27–29 March 2012, Utrecht, The Netherlands.
[9] P.K. Tellapaneni, S. Shukla, R.R. Jackson, H. Dixit, A New Approach for Wireline Formation Testers (WFT) Interval Pressure Transient Test Design, North Africa Technical Conference and Exhibition, 20–22 February 2012, Cairo, Egypt, 2012.
[10] G. Olivares, C. Escalona, E. Gimenez, Production Monitoring Using Artificial Intelligence, SPE Intelligent Energy International, 27–29 March 2012, Utrecht, The Netherlands, 2012.
[11] T. Weldu, S. Ghedan, O. Al-Farisi, in: Hybrid Artificial Intelligence and Conventional Empirical Approach for improved Prediction of Log-Derived Permeability of Heterogeneous Carbonate Reservoir, SPE Production and Operations Conference and Exhibition, 8–10 June 2010, Tunis, Tunisia (2010).
[12] Grujic O., Mohaghegh, S.D., Bromhal, G., 2010. Fast Track Reservoir Modeling of Shale Formations in the Appalachian Basin. Application to Lower Huron Shale in Eastern Kentucky, SPE Eastern Regional Meeting, 12–14 October 2010, Morgantown, West Virginia, USA.
[13] B.S. Mollaiy, An imperialist competitive algorithm artificial neural network method to predict oil flow rate of the wells, International Journal of Computer Applications 10 (2011) 47–50.
[14] R. Rallo, J. Ferre-Gin, A. Arenas, F. Giralt, Neural virtual sensor for the inferential prediction of product quality from process variables, Computers and Chemical Engineering 26 (2002) 1735–1754.