ARTIFICIAL NEURAL NETWORK WITH EXPERIMENTAL DESIGN: A FAST TOOL FOR RISK ANALYSIS AND FORECASTING IN AN OIL RESERVOIR, A CASE STUDY

Nasser Alizadeh [*1] ✉ iD

[*1] Petro-Pars Oil and gas company, Tehran, Iran

ABSTRACT

Running simulation models is CPU intensive. In computing expensive tasks such as parameter screening, sensitivity and risk analysis (uncertainty analysis) and production optimization, it can be useful to establish a simple surrogate model (proxy model) that mimics the simulation model with regard to a specific target value (for example, total production) in order to reduce the computing time and to study the available uncertainties in the reservoir and their impacts on production.

Artificial Neural Networks (ANN) are one of the main tools used in machine learning. The quality of the ANN as a proxy model is dependent on how the experiments that were used to make and train it are designed. In particular, it is crucial to understand the input parameters such that their respective dependencies, correlations, and ranges are incorporated in the modelling. A combination of simulation runs should be set up that can be used to train the ANN. This task is usually referred to as the design of experiments (DOE) which gives the most informative data sets to train ANN.

In this study DOE was used to train the ANN in an oil reservoir under gas injection scenario and the trained ANN, in turn, was applied to create the production profiles which were further used for risk analysis.

The accuracy of the results obtained in this study indicates that ANN as a proxy model combined with DOE as a sampling method for training purpose is a fast and reliable tool that can replace the simulator. This dynamic proxy model can be used for risk analysis, production optimization and production forecasting of oil reservoirs under Enhanced Oil Recovery (EOR) methods.

1. INTRODUCTION

Petroleum reservoir management is a dynamic process that recognizes the uncertainties in the reservoir performance resulting from our inability to fully characterize reservoirs and flow processes. It seeks to mitigate the effects of these uncertainties by optimizing reservoir performance through a systematic application of integrated, multidisciplinary technologies. Proxy modelling is a technology that can be used for uncertainty analysis and optimization.

A simulation case is put together by a huge number of input parameters. A first step towards understanding the uncertainty in the simulated results is to use domain knowledge to filter out the input parameters that are most

likely to influence the results of interest. This process is called sensitivity analysis. The next step in assessing uncertainty is to find out the impacts of the most influential uncertain parameters on simulated results, which is called risk analysis. For uncertainty analysis we might have to run a huge number of cases depending on the number of input variables, which can be very time consuming and frustrating task. A proxy model which is an analytical model can be developed and trained to give the same results as simulator and then the simulator will be replaced by this proxy model and it can be used for sensitivity analysis, risk analysis, optimization and prediction.

The quality of the proxy model depends on training data. A systematic way of selecting the training data, design of experiments is used to make better proxy models. A combination of simulation runs should be set up that can be used to train the proxy model. This tax is usually referred to as the design of experiments (DOE).

In this study we have used Artificial Neural Network as proxy model, which was trained by DOE. We are aiming to demonstrate that ANN with DOE can be used as a fast tool for uncertainty analysis and production forecasting in an oil reservoir under EOR methods. In our case study the reservoir was producing by immiscible gas injection.

In primary production, a large amount of oil remains in reservoir. Immiscible gas injection is one of the most commonly used methods of enhanced oil recovery (EOR) that recovers a considerable amount of residual oil. Investigating the oil reservoirs, associated with various uncertain parameters, under different EOR methods, is a major issue in reservoir engineering.

In uncertainty analysis studies, numerous simulation runs, corresponding to different values of uncertain parameters within their ranges, must be carried out to identify the effect of uncertain parameters on system response. This is a time-consuming and expensive procedure. Alternatively, developing a trained ANN as a proxy model, which gives outputs close to the simulation results, is considered to be an appropriate technique.

ANN Training is a challenging task (Du et al. 2003), which must be carried out so that ANN can capture the whole ranges of uncertain parameters. Experimental Design method efficiently distributes simulation runs within ranges of uncertain parameters, and therefore minimizes the number of simulation runs required for training an ANN (Steppan et al., 1998).

ANN has already been used for prediction (Birang et al. 2007; He et al. 2001; Huang, Y. F et al. 2003) and Experimental Design has been used to set up proper input data (Kabir et al. 2002; White and Royer, 2003). A combination of ANN and Experimental Design (Jalali and Mohaghegh, 2009; Lechner et al. 2005; Reis, 2006) has been applied in dealing with uncertain systems.

In the past different papers have discussed the applications of ANN in simulation. But in this study we are applying ANN as a dynamic proxy model which can mimic the results of simulator at different simulation times very fast and accurately.

The purpose of this study is to demonstrate that a combination of ANN and Experimental Design can be used to study the effects of deterministic uncertain parameters and the risk associated with them on production in an oil reservoir under immiscible gas injection or other EOR methods. We are also aiming to demons

2. CONCEPT OF FEEDFORWARD BACKPROPAGATION NEURAL NETWORK

ANN are excellent tools for finding patterns which are far too complex or numerous for a human programmer to extract and teach the machine to recognize. The ANN can be used as a proxy model which should be trained to give the same outputs as numerical simulator.

Feedforward backpropagation neural network is the most commonly used type of ANN for engineering purposes (He et al., 2001). In order to train an ANN based on backpropagation algorithm, after initializing the weights in the first iteration, ANN produces the output data. The error between the ANN output data (a) and actual output data (t) is calculated as follows:

$$e = (1/N)\sum_{i=1}^{N}\{t(i) - a(i)\}^2$$

(1)

Where N is the number of output data. This error is then fed back through the ANN to adjust the weights based on the equation below:

$$w_{i,j}(k+1) = w_{i,j}(k) - g_{k \times} Ir \qquad (2)$$

Where $W_{i,j}$ represents the value of weight connecting neuron j to i, lr is a constant called learning rate, k indicates the number of iterations and g_k is defined as:

$$g_k = \delta e / \partial w_{i,j} \qquad (3)$$

Following training, this ANN can produce output data corresponding to specific input data without changing the values of weights (Demuth *and Beale, 2002;* Minakowski, 2008).

3. CONCEPT OF EXPERIMENTAL DESIGN

Artificial Neural Network should be trained in order to deliver the same results as the experimental/observed or simulated data. We need to know how to set up and select the combination of experiments that are used to train the proxy model. The task of making this selection is often referred to as the design of experiments (DOE).

The quality of a proxy model is obviously dependent on the simulation cases that were used to train it. If, as an example, no low values of critical water saturation were used in any of the simulation cases that were used to train the proxy, the proxy model cannot be used to estimate flow in a case with a low critical water value.

Selecting the design of experiment, means selecting the sampling method. There are two main sampling methods for designs:

- Random sampler. When using random sampler, the uncertain variables are sampled randomly from the distribution that has been assigned to each of the variables. Different combinations of the variables are selected every time the method is used.
- Deterministic sampler. A deterministic sampler selects the sample of each of the uncertain variables in a prescribed way. As an example, a deterministic sampler could make combinations of the input variables where only the smallest and the largest values are used. Every time the sampler is used, it picks exactly the same combinations of input variables.

In this study we have used deterministic sampler to pick the simulation cases which were used to design the ANN. The deterministic samplers require two or three levels (values of the variables) at which they sample the distribution specified for each uncertain variable. These levels are low, mid and high values of the variable. The value for a variable at each of those levels is set depending on the distribution range and type. For example, for a variable that has a uniform distribution assigned to it, those levels would simply be the minimum, the mean and the maximum value. If a value is represented by a normal distribution, then the value for which the cumulative density function (CDF) is 5% is used as the low value, the mean is used as the mid, and the value for which the CDF is 95% is used as high value.

There are two widely used deterministic sampler which were used in this study, inscribed central composite and full factorial samplers. The former is similar to Box-Behnken method in the way that the corners of the sampling space are not sampled. It has sample points in the centre of each of the faces and one sample point in the centre of sampling cube. In addition, it has sampling points on an inscribed hyper sphere. Thus, the accuracy in the central sampling space is good. This method has five levels, which are low value, mid value, high value, positive radius of the inscribed sphere and negative radius of the inscribed sphere (Steppan et al., 1998). The latter, a full factorial sampler was first introduced by John Bennet Lawes and Joseph Henry Gilbert in 19th century. It is an experiment whose design consists of two or more factors, each with discrete possible values or levels. In this study we have used three-level full factorial design. This method uses minimum, mean and maximum values for each variable and includes all possible combinations of variables.

4. MODEL DEVELOPMENT

4.1. MS FIELD SPECIFICATIONS

MS field, discovered in 1995, is a large anticlinal closure which extends over an area of 210 kilometers in southern part of Iran in the average water depth of 55 meters. Original water oil contact was established at depth of 1149 meters in the southern flank of the structure and gas oil contact is at depth of 776 meters. Gross thickness of the formation is approximately 182 meters. The formation is characterized by excellent porosity (average porosity is 27%), low water saturation (20%) and average permeability of 14 and 2 mD in horizontal and vertical directions, respectively. MS is a shared field between two neighbouring countries and the development plan for this oil field is as follows:

Injecting gas into the reservoir (in the gas cap) through one injector, which is completed in the first layer for the pressure maintenance and immiscible displacement of oil by gas. There are 12 oil producers in the reservoir which are completed in the layers 4 and 5. Fig. 1 illustrates the 3D simulation model of the reservoir and the locations of the producers and gas injector as well.
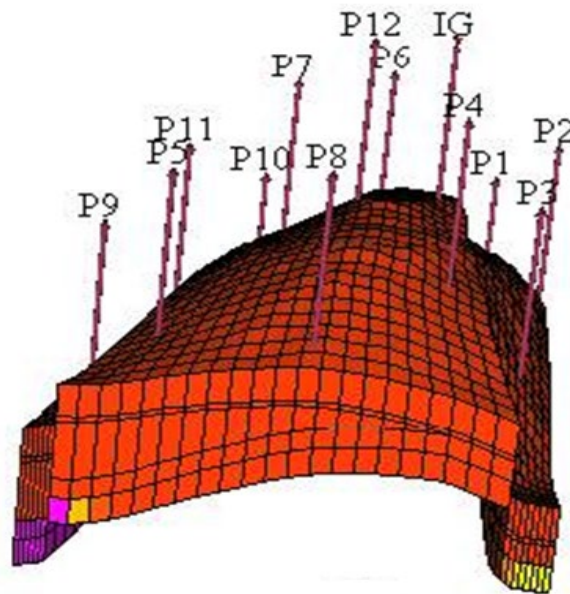

Figure 1: 3D Simulation model of MS field. P and IG represent the production and gas injection wells, respectively.

4.2. SENSITIVITY ANALYSIS

A simulation case is put together by a huge number of input parameters out of which some are uncertain, that means their values are not exactly known and they are ranging between a minimum and maximum (deterministic uncertain parameters). A first step towards understanding the uncertainty in the simulated results is to use domain knowledge to filter out the input parameters that are most likely to influence the results of interest. This process is called sensitivity analysis. In this study the 3D simulation model of the MS field was used for sensitivity study.

The pore volume multiplier (fluid in place of MS filed) was considered as uncertain parameter as MS is a shared field between two neighbouring countries. The other selected uncertain parameters and their ranges are summarized on table 1. In total there were 10 uncertain parameters selected for sensitivity analysis.

Table 1: Minimum, base and maximum values of uncertain parameters.

Uncertain Parameter	Minimum	Base	Maximum
pore volume multiplier (MULTPV)	0.5	1	2
transmissibilty multiplier in X direction (MULTX)	0.5	1	2
transmissibilty multiplier in Y direction (MULTY)	0.5	1	2
transmissibilty multiplier in Z direction (MULTZ)	0.5	1	2
fault transmissibility multiplier (MULTFLT)	0.5	1	2
bottom hole pressure of gas injector [psia] (BHPinj)	4000	5000	6000
surface gas injection rate [Mscf/day], Qinj	1000	5000	10000
aquifer productivity index [stb/day/psi] (PIa)	10	200	400
initial volume of water in aquifer [stb] (Va)	7.00E+07	7.00E+08	7.00E+09
critical gas saturation (SGCR)	0.01	0.03	0.06

In sensitivity runs one of the uncertain input parameters are changed at a time while all other parameters are kept fixed. This is done to try to determine which of the input parameters affect the results the most. Because only one variable is changed at a time, interactions between input variables are not measured. In this study the objective of sensitivity analysis was to find out which uncertain parameters were most influential on oil production and field pressure in MS field under immiscible gas injection scenario.

The results of sensitivity runs (Tornado plots) for 12 years of simulation time are given in Fig. 2. It can be observed that the most influential uncertain parameters on oil and gas production rates and field pressure are pore volume multiplier (MULTPV), transmissibility multiplier in Z direction (MULTZ), gas injection rate (Qinj) and critical gas saturation (SGCR). The longer the bar on tornado plot, the more influential is that parameter on response.
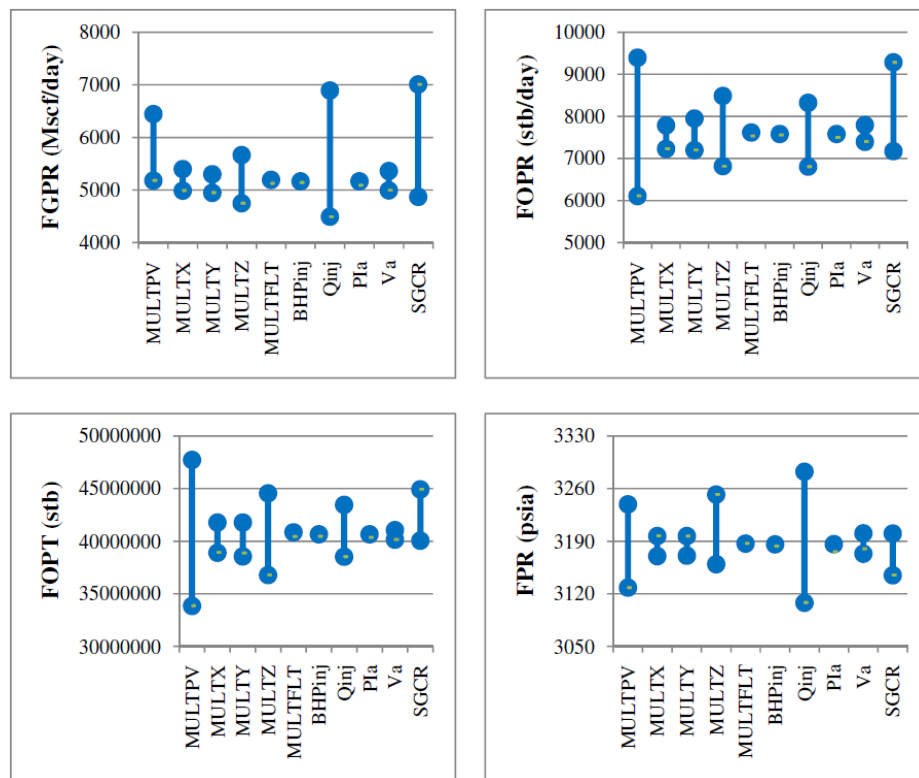

Figure 2: Tornado plots of field gas production rate (FGPR), field oil production rate (FOPR), cumulative oil production (FOPT) and field pressure (FPR) after 12 years of production. The longer the bar, the more influential is that variable on response.

4.3. TRAINING AND TESTING THE ANN

Having reduced the number of uncertain parameters by sensitivity analysis, the most influential uncertain parameters (pore volume multiplier, transmissibility multiplier in Z direction, surface gas injection rate and critical gas saturation) were used as input into ANN to train it. The ANN was trained in order to give the same results as numerical simulator.

Experimental Design was applied to obtain the most informative data sets to train ANN and prevent overtraining problem. Variable settings for the most influential uncertain parameters, based on three-level full factorial and inscribed central composite designs, have been shown on table 2.

Table 2: Three-level full factorial and inscribed central composite designs.

THREE –LEVEL FULL FACTORIAL DESIGN [a]														
Run	MULTPV	MULTZ	SGCR	Qinj	Run	MULTPV	MULTZ	SGCR	Qijn	RUN	MULTPV	MULTZ	SGCR	Qinj
1	-1	-1	-1	-1	28	-1	1	0	0	55	1	1	-1	0
2	-1	-1	-1	1	29	0	0	1	1	56	1	1	1	0
3	-1	-1	1	-1	30	0	0	-1	-1	57	0	-1	-1	-1
4	-1	-1	1	1	31	0	0	1	-1	58	0	-1	-1	1
5	-1	1	-1	-1	32	0	0	-1	1	59	0	-1	1	-1
6	-1	1	-1	1	33	1	0	0	1	60	0	-1	1	1
7	-1	1	1	-1	34	-1	0	0	-1	61	0	1	-1	-1
8	-1	1	1	1	35	1	0	0	-1	62	0	1	-1	1
9	1	-1	-1	-1	36	-1	0	0	1	63	0	1	1	-1
10	1	-1	-1	1	37	0	1	1	0	64	0	1	1	1
11	1	-1	1	-1	38	0	-1	-1	0	65	-1	0	-1	-1
12	1	-1	1	1	39	0	1	-1	0	66	-1	0	-1	1
13	1	1	-1	-1	40	0	-1	1	0	67	-1	0	1	-1
14	1	1	-1	1	41	1	0	1	0	68	-1	0	1	1
15	1	1	1	-1	42	-1	0	-1	0	69	1	0	-1	-1
16	1	1	1	1	43	1	0	-1	0	70	1	0	-1	1
17	1	0	0	0	44	-1	0	1	0	71	1	0	1	-1
18	-1	0	0	0	45	0	1	0	1	72	1	0	1	1
19	0	1	0	0	46	0	-1	0	-1	73	-1	-1	0	-1
20	0	-1	0	0	47	0	1	0	-1	74	-1	-1	0	1
21	0	0	1	0	48	0	-1	0	1	75	-1	1	0	-1
22	0	0	-1	0	49	-1	-1	-1	0	76	-1	1	0	1
23	0	0	0	1	50	-1	-1	1	0	77	1	-1	0	-1
24	0	0	0	-1	51	-1	1	-1	0	78	1	-1	0	1
25	1	1	0	0	52	-1	1	1	0	79	1	1	0	-1
26	-1	-1	0	0	53	1	-1	-1	0	80	1	1	0	1
27	1	-1	0	0	54	1	-1	1	0	81	0	0	0	0
INSCRIBED CENTRAL COMPOSITE DESIGN [b]														
Run	MULTPV	MULTZ	SGCR	Qinj	Run	MULTPV	MULTZ	SGCR	Qijn	RUN	MULTPV	MULTZ	SGCR	Qinj
1	-1	-1	-1	-1	11	1	-1	1	-1	21	0	0	2	0
2	-1	-1	-1	1	12	1	-1	1	1	22	0	0	-2	0
3	-1	-1	1	-1	13	1	1	-1	-1	23	0	0	0	2
4	-1	-1	1	1	14	1	1	-1	1	24	0	0	0	-2
5	-1	1	-1	-1	15	1	1	1	-1	25	0	0	0	0
6	-1	1	-1	1	16	1	1	1	1	26	0	0	0	0
7	-1	1	1	-1	17	2	0	0	0	27	0	0	0	0
8	-1	1	1	1	18	-2	0	0	0	28	0	0	0	0
9	1	-1	-1	-1	19	0	2	0	0					
10	1	-1	-1	1	20	0	-2	0	0					

a: -1, +1, 0 represent minimum, maximum and mean values, respectively.
b: +/-2, +/-1 and 0 represent star points, factorial points, and centre points

The number of required simulations runs based on three- level full factorial and inscribed central composite designs for 4 influential uncertain parameters were equal to 81 and 28, respectively. There were 12 identical runs in these 2 designs, therefore the total number of simulation runs would be 97.

Having assessed various configurations of the number of hidden layers and neurons in each layer, a feedforward backpropagation neural network illustrated in Fig. 3 was developed in order to get the best fit during training using the aforesaid 97 runs.
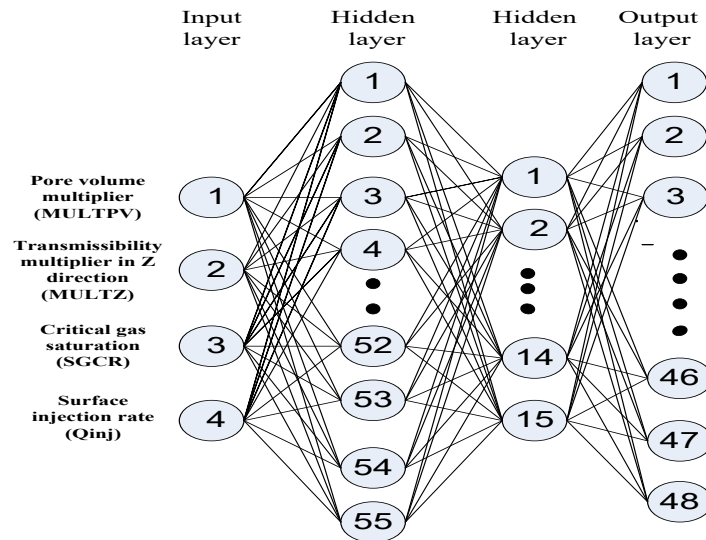
Figure 3: Neural network sketch.

Four neurons in the input layer receive the most influential uncertain parameters (MULTPV, MULTZ, Qinj and SGCR). The output layer consists of 4 categories. The first category contains 12 neurons producing FGPR during 12 years, so that each neuron is allotted to yield FGPR in the beginning of each year. The second, third and fourth categories are similar to the first category except that they are allotted for producing FOPR, FOPT, FPR respectively.

In training runs, the average deviation of outputs predicted by ANN from the ones calculated by simulator was under 0.7% and the maximum deviation was less than 11%. The results of training runs can be seen in Fig. 4.
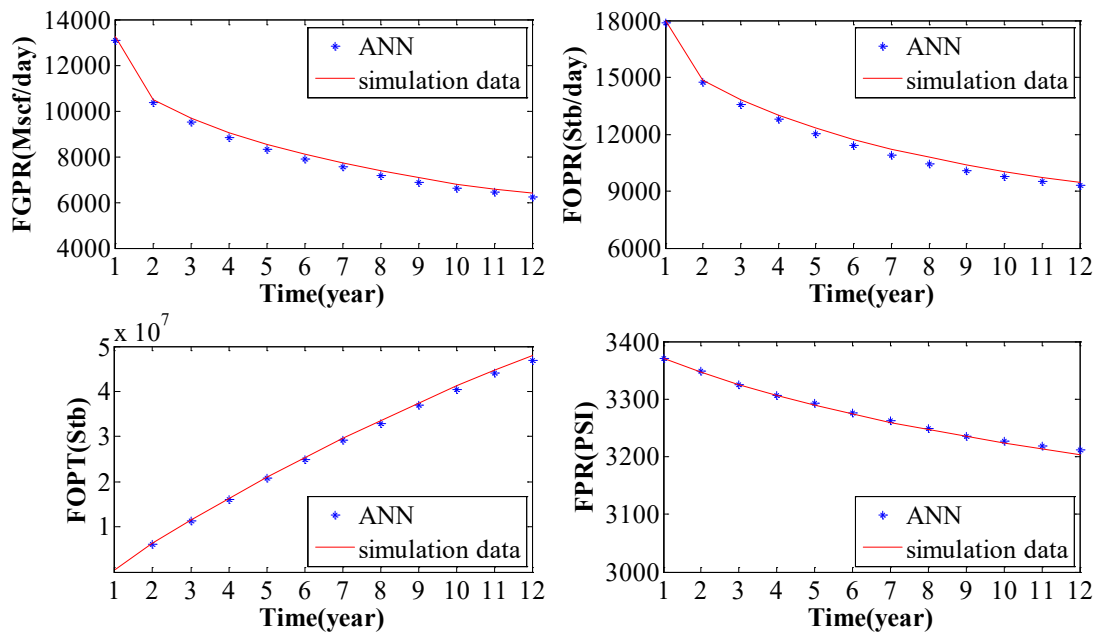

Figure 4: Comparison of ANN and simulation model results (training runs).

In order to ensure the accuracy of ANN, the trained ANN was tested with 10 additional runs, which have not been seen by the ANN during training. As an illustration, the deviation of outputs predicted by the ANN from the simulation results, corresponding to FGPR, FOPR, FOPT and FPR for one testing run have been shown in Fig. 5. During testing, the average deviation was under 2% and the maximum deviation was less than 8%.
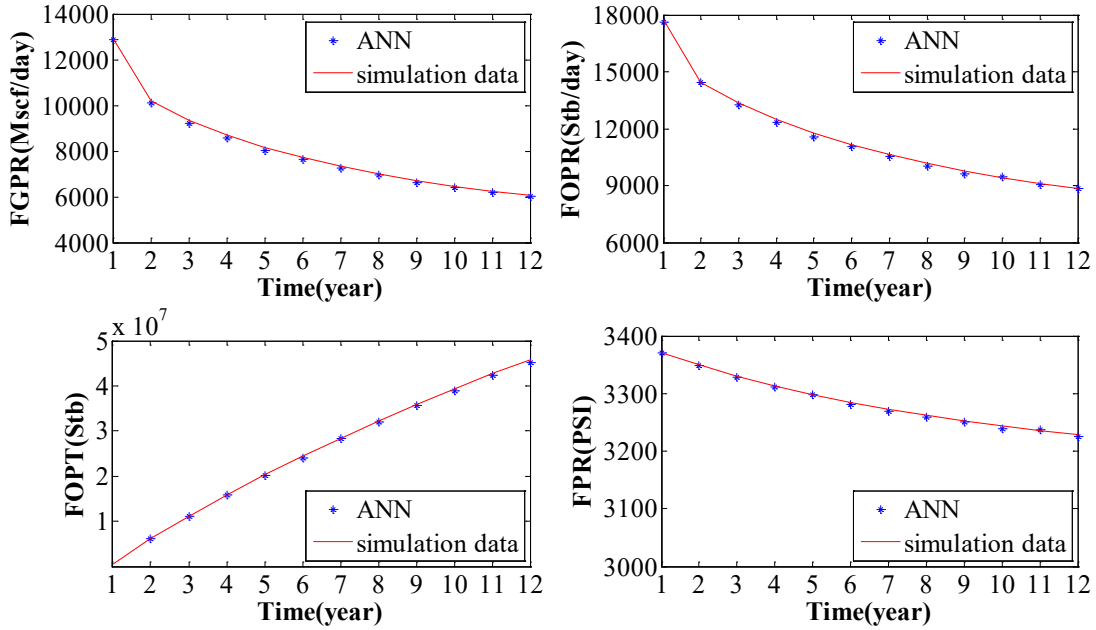
Figure 5: Comparison of ANN and simulation model results (testing run).

4.4. RISK ANALYSIS STUDY BY THE TRAINED ANN

Having trained the ANN, it was used for risk analysis in MS field under immiscible gas injection scenario to assess the effects of most influential uncertain parameters on oil production. Normally this task is run when the number of input variables has been reduced by use of sensitivity studies. The probability distribution function (PDF) of the response can be used to observe the risk or impacts of the most influential uncertain parameters on oil production.

The aim was to make runs with different combinations of the most influential uncertain parameters to assess interaction effects in addition to the influence of each single variable. Mont-Carlo sampler was used as sampling method. It is a stochastic method that samples the variables randomly from their input distribution. 20000 Mont-Carlo simulation runs were submitted. The results in form of a probability distribution at the end of simulation time (12 years after start of production) is shown in Fig. 6. It can be observed that due to the available uncertainty in the model the cumulation oil production of the MS field is ranging from 2.7×10^7 to 5.6×10^7 stb.
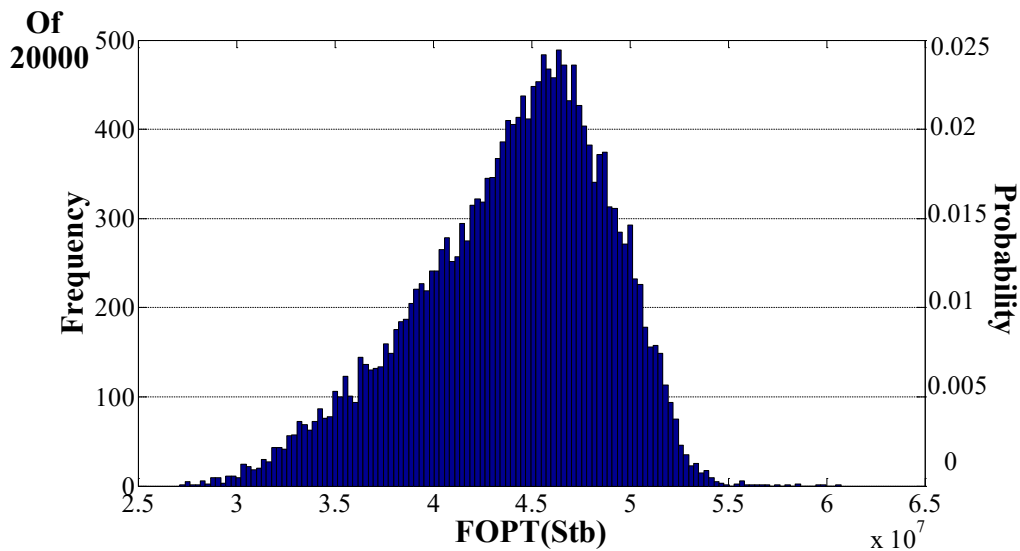

Figure 6: Probability distribution (PDF) of cumulative oil production (FOPT) after 12 years of production.

4.5. PREDICTION BY THE TRAINED ANN

Predicting the reservoir performance is one of the main objectives of developing a proxy model. In this study the trained ANN was used to predict the performance of the MS field under immiscible gas injection. The ANN was run for the following values of the most influential uncertain parameters:

Surface gas injection rate: 6000 Mscf/day.
Pore volume multiplier: 1 (base value)
Transmissibility multiplier in Z direction: 1 (base value)
Critical gas saturation: 0.03 (base value)

The predicted production profile and field pressure by ANN for the surface gas injection rate of 6000 Mscf/day have been compared with those predicted by numerical simulator. The field oil production rate, gas production rate, field pressure and cumulative oil production have been plotted in Fig. 7. It can be observed that there is a good agreement between the results of ANN and numerical simulator. That means the trained ANN can be used to predict the performance of the MS field under gas injection scenario.
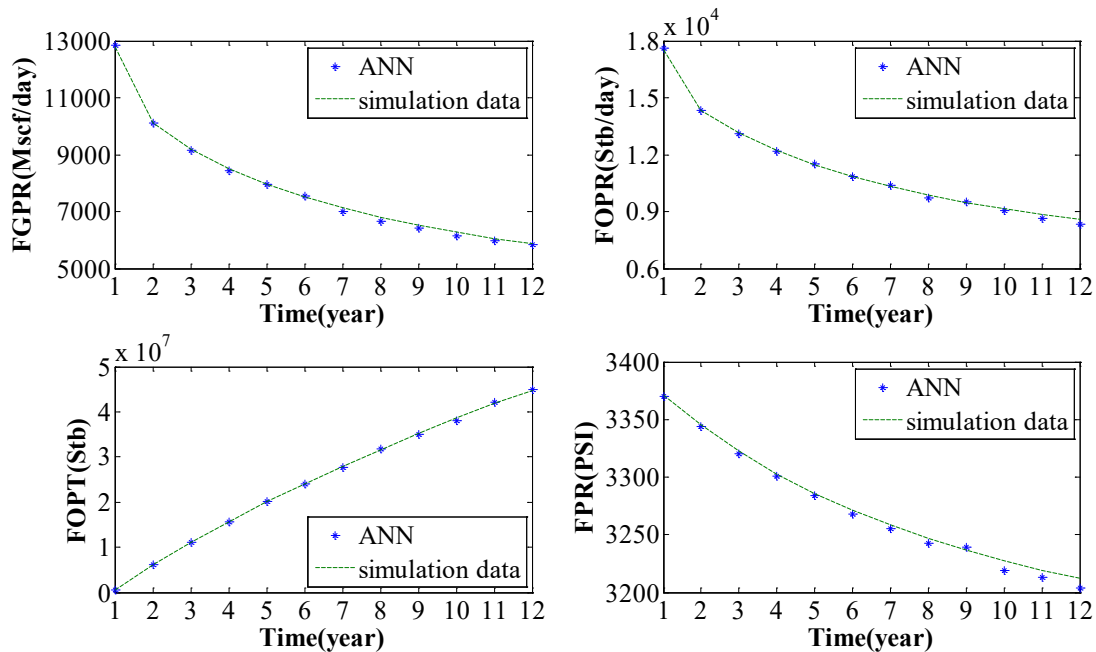

Figure 7: Comparison of the results predicted by ANN and simulation model corresponding to surface gas injection rate of 6000 [Mscf/day] and base values of pore volume multiplier (MULTPV), transmissibility multiplier in Z direction (MULTZ) and critical gas saturation (SGCR).

5. RESULTS AND DISCUSSION

In this study initially a 3D simulation model of an Iranian oil field was developed and history matched to perform sensitivity analysis in order to filter out the uncertain input parameters that were most likely to influence oil production in an Iranian field under immiscible gas injection. Sensitivity analysis indicated that pore volume multiplier, transmissibility multiplier in Z direction, gas injection rate and critical gas saturation were the most influential uncertain parameters on total oil production. Having identified the most influential uncertain parameters, they were used to train ANN. Experimental design was applied to obtain the most informative data sets to train ANN and the trained ANN, in turn, was applied for risk analysis to get the probabilistic production profile of the field. Mont-Carlo sampler was used as sampling method which generated 20000 randomly selected samples from the 4 selected influential uncertain parameters. Mont-Carlo is the recommended stochastic sampling method for risk analysis. The output of the risk analysis showed that cumulative oil production of the field was ranging from 2.7×10^7 to 5.6×10^7 stb. This dispersion of the result was due to available uncertainty in the model and in other words, it was

the impact of most influential uncertain parameters on total production. This dispersion should be minimized. In other words, the production should be optimized by minimizing the effects of available uncertainty in the model. This optimization is not covered in this study and will be presented in our next paper.

6. CONCLUSIONS

In this study we have applied ANN as a dynamic proxy model which can mimic the results of simulator at different simulation times very fast and accurately.

The quality of ANN is dependent on how the experiments that are used to make and train it are designed. The results of this study show that a combination of simulation runs which are set up by DOE are able to train ANN accurately. DOE gives the most informative data sets to train ANN.

Based on the results of this study it can be concluded that ANN as a dynamic proxy model, trained with DOE is a fast, accurate and cost-effective tool that can be used for risk analysis and production forecasting in an oil reservoir under EOR methods.

SOURCES OF FUNDING

CONFLICT OF INTEREST

The author have declared that no competing interests exist.

ACKNOWLEDGMENT

None.

REFERENCES

[1] Birang, Y., Dinarvand, N., Shariatpanahi, S. F., Edalat, M. 2007. Development of a new artificial-neural-network model for predicting minimum miscibility pressure in hydrocarbon gas injection. SPE 105407, Middle East Oil &Gas Show and Conference, Bahrain, March 11-14.

[2] Demuth, H., Beale, M. 2002. Neural network toolbox user's guide. The MathWorks, Inc.

[3] Du, Y., Weiss, W. W., Xu, J., Balch, R. S., Li, D. 2003. Obtain an optimum artificial neural network model for reservoir studies. SPE 84445, Annual Technical Conference and Exhibition, Colorado, October 5-8.

[4] He, Z., Yang, L., Yen, J., Wu, C. 2001. Neural-network approach to predict well performance using available field data. SPE 68801, Western Regional Meeting, March 26-30.

[5] Huang, Y. F., Huang, G. H., Dong, M. Z., Feng, G. M. 2003. Development of an artificial neural network model for predicting minimum miscibility pressure in CO2 flooding. Journal of petroleum science and engineering 37, 83-95.

[6] Jalali, J., Mohaghegh, S. D. 2009. Reservoir simulation and uncertainty analysis of enhanced CBM production using artificial neural networks. SPE 125959, Eastern Regional Meeting, West Virginia, September 23–25.

[7] Kabir, C. S., Chawathe, A., Jenkins, S. D., Olayomi, A. J., Aigbe, C., Faparusi, D. B. 2002. Developing new fields using probabilistic reservoir forecasting. SPE 77564, Annual Technical Conference and Exhibition, Texas, 29 September-2 October.

[8] Lechner, J. P., AG, O., Zangl, G. 2005. Treating uncertainties in reservoir performance prediction with neural networks. SPE 94357, Europec/EAGE Annual Conference, Spain, June 13-16.

[9] Parada Minakowski, C. H. 2008. An artificial neural network based tool-box for screening and designing improved oil recovery methods. Doctoral Thesis, Pennsylvania State University, USA.

[10] Reis, L. C. 2006. Risk analysis with History matching using experimental design or artificial neural networks. SPE 100255, Europec / EAGE Annual Conference and Exhibition, Austria, June 12-15.

[11] Steppan, D. D., Werner, J., Yeate, R. P. 1998. Essential Regression and Experimental Design for Chemists and Engineers.

[12] White, C. D., Royer, S. A. 2003. Experimental design as a framework for reservoir studies. SPE 79676, Reservoir Simulation Symposium, Texas, February 3-5.