TECHNICAL UNIVERSITY OF CLUJ-NAPOCA

ACTA TECHNICA NAPOCENSIS

Series: Applied Mathematics, Mechanics, and Engineering
Vol. 65, Issue III, September, 2022

USING THE ARTIFICIAL NEURAL NETWORK TO APPROXIMATE THE GAS-LIFT PERFORMANCE CURVE

Mariea MARCU

Abstract: *In the paper, artificial neural networks and a regression equation were used to approximate the gas-lift performance curve in order to compare their performances. Also, a sensitivity study of the artificial neural network performances to the variation of its geometric parameters and the activation function was carried out. The data sets used to build the gas-lift performance curve have different characteristics identified by the number of data points and the presence or absence of the outliers. In all cases, the performances of the artificial neural network were better than those of the regression equation. However, if the data set contains many outliers, the artificial neural network, although it has smaller errors, tends to build an abnormal curve.*

Keywords: *gas-lift performance curve, nodal analysis, artificial neural network, outliers, regression equation, activation function*

1. INTRODUCTION

The gas-lift is a production system that is applied, generally, after a naturally flowing well ceases to produce because the reservoir energy decrease and cannot bring these fluids to the surface. The gas-lift system uses the energy of the gas injected from the surface to lift the fluid from the oil reservoir to the surface.

Each gas-lift well has a particular response meaning the liquid flow rate produced to the gas injection flow rate. The function liquid flow rate=f(gas injection flow rate) ($Q_L = f(Q_G)$) defines the gas-lift performance curve of a gas-lift well. This curve is the input of the optimization process of the gas injection flow rate for each gas-lift well, taking into account that the gas-lift production system is effective when it is applied in many wells. Consequently, the gas lift performance curve is very important and it is necessary to build it as accurately as possible.

In the specialized literature, many works describe different methods to build the gas-lift performance curve, such as regression and artificial intelligence [2][5][6][7].

2. GAZ-LIFT PERFORMANCE CURVE

The gas-lift performance curve is built on the basis of discrete data sets obtained from the measurement in the field or from the simulation based on the nodal analysis. The nodal analysis involves defining the components of the production system, as well as choosing a node somewhere in the system. In this way, the production system is divided into two parts: the inflow part which is represented by the fluid flow through the reservoir, and the outflow part which is represented by the upward two-phase fluid flow through the tubing string.

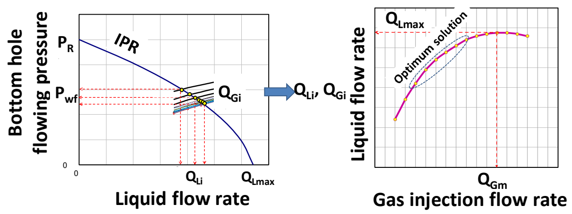
The performances of the two-part are characterized by the IPR (Inflow Performance Relationships) curves, respectively by the OPR (Outflow Performance Relationships) curves.

IPR curves are built on the basis of well test data (bottom hole pressure and liquid flow rate) and using a calibrated correlation. To build the OPR curves, it is necessary to know much data such as well trajectory, tubing string characteristics, PVT properties of reservoir fluids and gas injected flow rate, etc., and to calibrate an upward two-phase flow correlation.

Received: 30.07.22; **Similarities**: 11.08.22: **Reviewed**: 31.08./22.08.22: **Accepted**:15.09.22.

The solution of the flow equations in the node can be obtained graphically or analytically.

In figures 1 and 2 we show the graphical solution.



a. b.

Fig. 1. a. Nodal analysis; b. Gas lift performance curve.

As we show in figure1 a., the operating points are determined at the intersection between OPR and IPR curves. Further, the gas injection rate and liquid flow rate values at the operating points are used to build the gas lift performance curve(fig.1.b.).

If we analyze all the steps necessary to obtain the data set for a simulated gas-lift performance curve, we identify many errors due to uncertainties that can appear at each step. However, the nodal analysis provides a large number of data points necessary to build the gas lift performance curve.

3. ARTIFICIAL NEURAL NETWORK USED TO APPROXIMATE THE GAS-LIFT PERFORMANCE CURVE

To solve specific problems in the oil and gas industry, different artificial intelligence techniques were applied [1][3].

Sometimes these techniques were combined to obtain faster and better results. For example, Artificial Neural Network was combined with a Genetic Algorithm to provide the optimum solution in the case of gas injection allocation for several gas-lift wells [4][7]. Here, the Artificial Neural Network(ANN) was used as a proxy model and the Genetic Algorithm was used to solve the nonlinear optimization problems in order to provide the optimum solution.

The ANN is the simplest and the most used. Over time, different types of ANN have been developed, the most used being the Multilayer Perceptron(MLP).

The ANN was inspired by human brain behavior. The structure of an artificial neural network consists of multiple nodes named neurons which are placed on several layers as the input layer, hidden layers, and output layer.

The number of input data can vary between 1 (fig.2) and n. Thus, the input layer can have one to n neurons. Also, the number of output data, respectively the number of neurons per output layer can vary between 1(fig.2) and k .

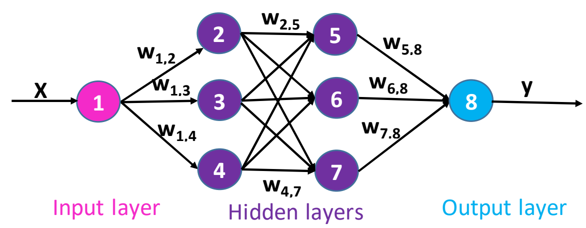


Fig. 2. Architecture of a simple ANN with two hidden layers and a single neuron on input, respectively output layer.

The input data are loaded by the input layer. Each neuron is connected by links with different weights, $w_{i,j}$ to the other neurons (fig.2). In this way, the neurons interact with each other.

Therefore, each neuron takes its input data, performs a simple operation (y=$\sum w_{ij}\, x_j$) on this data, and then uses an activation function to obtain its output which is passed to other neurons on the next layer.

The solver optimizes the weights in the process of neural network training in order to produce accurate output data.

4. RESULTS AND DISCUSSION

In our paper, we want to show the performance of an ANN in approximating the gas-lift performance curve based on five data sets with some characteristics.

The other objective of this study is to determine how the architecture and the activation function influence the performances of the ANN.

Therefore, we use a feed-forward neural network as a regressor for a function with a single independent variable $Q_L = f(Q_G)$. The architecture of the ANN is similar to the one shown in fig.2.

To compare the performances of the ANN in gas lift performance curve fitting, we also use the common regression methods. From these, we select the regression equation (1) developed by Namdar[5], which is the newest and leads to better results than the old regression equations.

$$Q_L = a + b \cdot Q_G + c \cdot {Q_G}^{0.7} + d \cdot \ln(Q_G + 0.9) + e \cdot \exp\left(-{Q_G}^{0.6}\right) \quad (1)$$

where Q_L , Q_G are oil flow rate, respectively gas-injection flow rate and a to e are the coefficients determined by regression.

To compare the fitting performances of the ANN and the regression equation of Namdar[5] we use the statistical parameters like root mean squared errors(RMSE) and squared Pearson correlation coefficient, R^2 given by the following equations:

$$RMSE = \sqrt{\frac{\sum(x_i - x_{im})^2}{n}} \quad (2)$$

$$R^2 = \frac{\left[\sum_{i=1}^{n}(x_{im} - \overline{x_m})(x_i - \bar{x})\right]^2}{\sum_{i=1}^{n}(x_{im} - \overline{x_m})^2 \sum_{i=1}^{n}(x_i - \bar{x})^2} \quad (3)$$

where n is the number of the data points (liquid flow rate, Q_{Li}, gas injection flow rate, Q_{Gi});

x_{im} —measured values (liquid flow rate);

x_i – predicted value (liquid flow rate);

$\overline{x_m}$ —average values of measured liquid flow rate $\overline{x_m} = \sum_{i=1}^{n} x_{im}/n$;

$\bar{x}$ —average values of predicted liquid flow rate, $\bar{x} = \sum_{i=1}^{n} x_i/n$.

The data used in our study are shown in figure 3 and were provided by Behjoomanesh et al[2].

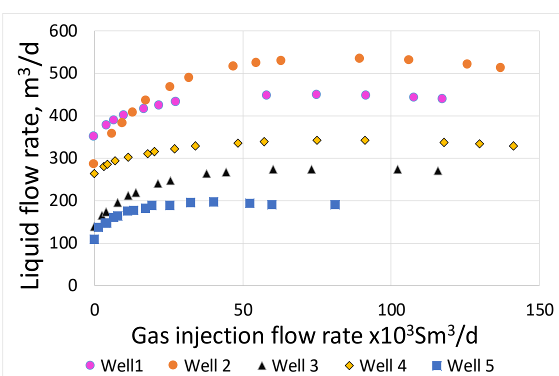


Fig.3. The data sets for five gas-lift wells.

We chose these data sets because they have different numbers of data points and some data points are slightly outliers (especially well 5).

In these real conditions, we want to know what regression method has the better performance.

To achieve all the objectives of our study we develop a program in Python to calculate the regression coefficients of the Namdar equation [5], build the ANN, and calculate statistical parameters in the case of two fitting methods.

For each well, we will consider different working scenarios where the number of hidden layers varies between 1 and 12.

Also, the number of neurons on each layer is considered constant as 5, 10, 50, and 100 neurons.

We use different types of activation functions (logistic sigmoid function, hyperbolic tangent function, and rectified linear unit function (ReLU)) and LBFGS as solver to find what function leads to the best results.

LBFGS(Limited–memory Broyden–Fletcher–Goldfarb–Shanno algorithm) is a quasi-Newton optimizer that adjusts the weights of the links between the neurons to minimize the errors during the training process.

We select LBFGS as the solver because the data sets are small, converge faster, and perform better [8].

The equations and the ranges of the activation functions used in our study are shown in table 1.

Table 1

Activation functions.

Function	**Equation**	**Range**
Sigmoid function	$\sigma(x) = \frac{1}{1+e^{-x}}$	(0,1
Hyperbolic tangent	$f(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$	(-1,1)
ReLU	$f(x) = \begin{cases} 0 \text{ for } x < 0 \\ x \text{ for } x \geq 0 \end{cases}$	[0,∞)

To compare the performances of the activation function shown in table 1, we consider the data set for each well, an ANN with 4 hidden layers which contain the constant numbers of neurons per layer.

The results of the calculations in terms of average RMSE and average R^2 for all numbers of neurons/ layers considered and for all the five wells are shown in table 2.

Table 2

Average RMSE and R^2 for fitting the gas-lift performance curve using an ANN with different architecture and activation functions for Well 1 to Well 5.

Activation function	Average RMSE	Average R^2
Well 1		
Activation function	**Average RMSE**	**Average R^2**
Logistic sigmoid	1,11351	0,99862
Tanh	0,42116	0,99981
ReLU	3,27776	0,98095
Well 2		
Activation function	**Average RMSE**	**Average R^2**
Logistic sigmoid	1,58843	0,99956
Tanh	1,27843	0,99971
ReLU	2,11835	0,99828
Well 3		
Activation function	**Average RMSE**	**Average R^2**
Logistic sigmoid	1,39967	0,99901
Tanh	1,10069	0,99938
ReLU	2,91165	0,99187
Well 4		
Activation function	**Average RMSE**	**Average R^2**
Logistic sigmoid	0,66103	0,99920
Tanh	0,56289	0,99942
ReLU	1,25260	0,99533
Well 5		
Activation function	**Average RMSE**	**Average R^2**
Logistic sigmoid	1,64002	0,99555
Tanh	1,59164	0,99581
ReLU	6,39792	0,86267

From table 2 we observe that the hyperbolic tangent function leads to the smallest errors in the fitting of the gas-lift performance curve.

The largest errors, as expected, were recorded for the ReLU function which is a linear function. However, during the simulations, the ReLU function leads to smaller errors than the logistic sigmoid function and even the hyperbolic tangent function when the number of neurons per layer is more than 50.

Consequently, for the sensitivity study of ANN performances with respect to its architecture, we consider the hyperbolic tangent as the activation function.

To find a more performant architecture of the ANN for each well, we consider different working scenarios, where the number of hidden layers varies between 1 to 12, and the number of neurons constant on each layer has the values: 5; 10; 50; 100. The RMSE was calculated for each working scenario and the results of calculus are shown graphically in figure 4.

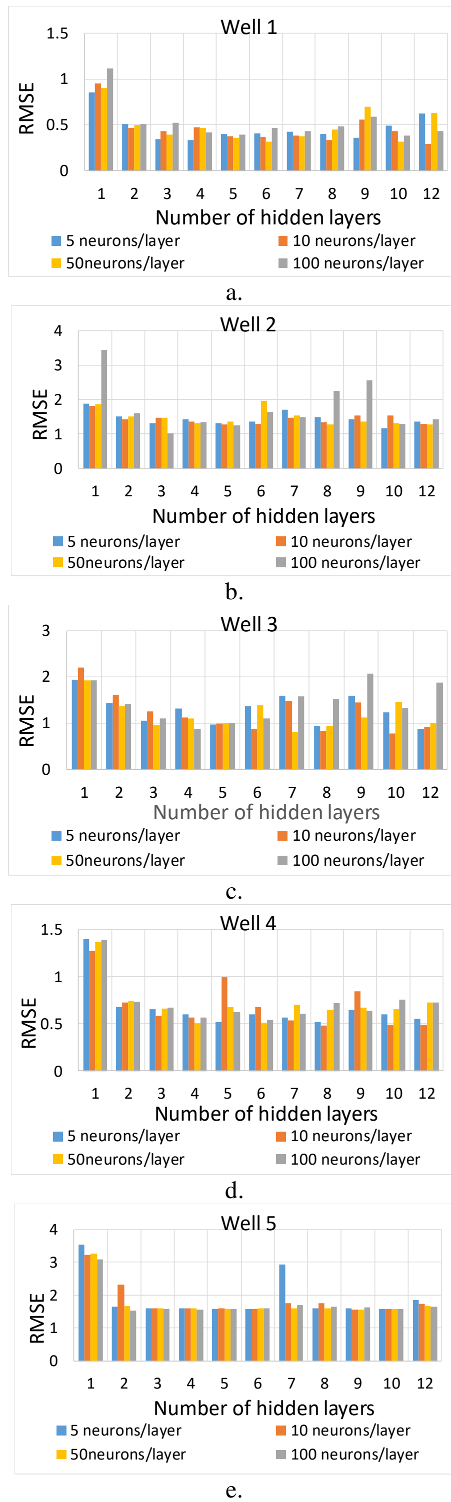


Fig.4. RMSE vs. the number of hidden layers and neurons/layer for: a. Well 1, b. Well 2, c.Well 3, d. Well 4, e. Well

From figure 4 we observe that the ANN with one hidden layer has the worst performance in all the cases studied. Alternatively, even for 9 hidden layers (Well 1 to Well 4) or 7 hidden layers (Well 5) the fitting performances of the ANN are lower because of the overfitting of the training data. If we analyze the data set of Well 5(with many outliers) and the results from figure 4 we observe that the performances of the ANN are approximately the same for more than one hidden layer and more than 5neurons/layer. In the case of Well 3, we identify three outliers and we observe an accentuated variation of the ANN performances with the number of layers and neurons/layer. Alternatively, Well 1 and Well 2 haven't the outliers, and well 4 has one outlier. In this case, the variation of RMSE is not very accentuated for more than two hidden layers. The architecture of the ANN with the best performances for each well is shown in table 3. The calculus for all the cases takes into account a tolerance of 10^{-8}. Also, in table 4 we show the fitting results with the Namdar equation [5].

Table 3

Minimum RMSE and maximum R^2 in the fitting of gas-lift performance curve with ANN having a different architecture for the five wells.

Well	ANN architecture	RMSE	R^2
1	12 layers 10 neurons/layer	0,29617	0,99990
2	5 layers 10 neurons/layer	1.28249	0.99971
3	7 layers 50 neurons/layer	0,81130	0,99967
4	5 layers 5 neurons/layer	0,52155	0,99950
5	2 layers 100 neurons/layer	1,52838	0,99613

Table 4

RMSE and R^2 in the fitting of gas-lift performance curve with the Namdar regression equation for the five wells.

Well	RMSE	R^2
1	0,31055	0,99989
2	1,92638	0,99934
3	1,71111	0,99852
4	0,71611	0,99907
5	2.60772	0.98874

The Namdar equation [5] has a slightly smaller performance in the fitting of the gas-lift performance curve than ANN(tables 3 and 4). Also, for relatively smooth data sets (fewer outliers) the performances of the Namdar equation [5] and ANN in fitting the gas-lift performance curve are closed (Well1, Well 2, and Well4).

For Well5 the difference between RMSE for ANN (2 hidden layers and 100 neurons/layer) and the Namdar equation [5] is big (table 3).

However, if we plot the fitted gas-lift performance curves with the two methods we see that ANN, in this case, builds an abnormal curve because of the overfitting of the training data set (fig.5). Therefore, a simple comparison between the values of RMSE is not enough to evaluate the fitting performances.

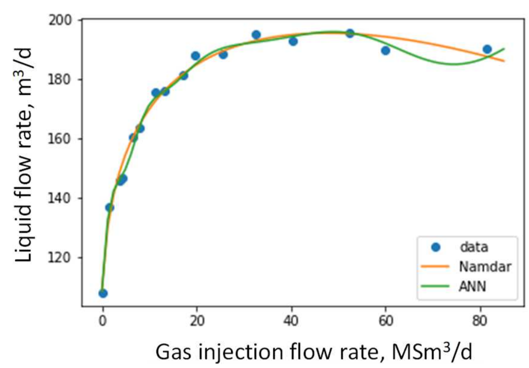


Fig.5 Data set for Well 5, gas-lift performance curve fitted with ANN (2L, 100 N/L) and the Namdar equation.

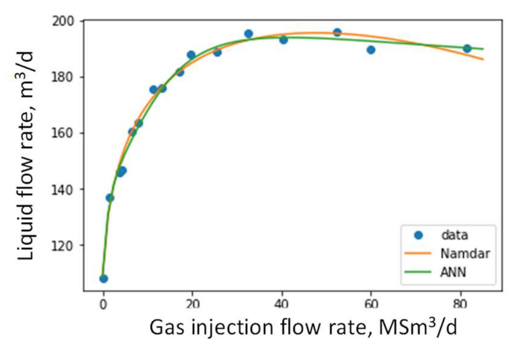


Fig.6 Data set for Well 5, gas-lift performance curve fitted with ANN (5L, 50 N/L) and the Namdar equation.

In figure 6, the gas-lift performance curve fitted with the ANN (5 hidden layers, 50 neurons/layer, and tolerance 10^{-4}) is better than that from figure 5. RMSE, in this case, is 2,17815 being lower than that obtained from the gas-lift performance curve fitting with the Namdar equation.

5. CONCLUSION

Unlike other papers, the present paper presents an analysis of the results obtained when the ANN and a regression equation are applied to build the gas-lift performance curve based on the data sets which contain the outliers. It was also studied how the activation function, the tolerance, and the network architecture influence the fitting performances. In all the studied cases the ANN has better performances than the regression equation of Namdar. Also, it results that the best activation function is hyperbolic tangent.

We found that the number of hidden layers must be more than 2 and the number of neurons/layer can vary between 5 and less than 100. However, in the case of many layers and neurons/layers, the results can be poor. In the cases of the data sets with outliers, the risk is to obtain an abnormal gas-lift performance curve. Therefore, in this case, the architecture of the ANN and the tolerance must be chosen such that do not lead to overfitting of the training data set.

6. REFERENCES

[1] Alkinani, H.H, Al-Hameedi, A.T.T., Dunn-Norman, Sh.., Flori. E.R., Alsaba, M.T., Amer, A.S., *Application of Artificial Neural Networks in the Petroleum Industry. A Review*, Paper Number SPE-195072-MS presented at the SPE Middle East Oil and Gas Show and Conference, Manama, Bahrain, 2019.

[2] Behjoomanesh,M., Keyhani, M., Ehsan Ganji-Azad, Izadmehr, M., Riahi, S., *Assessment of total oil production in gas lift process of wells using Box-Behnken design of experiments in comparison with traditional approach*. Journal of Natural Gas Science and Engineering, Vol. 27, pp.1455-1461, 2015.

[3] Li, H., Yu, H., Cao, N., Tian, H., Cheng, S., *Applications of Artificial Intelligence in Oil and Gas Development*, Archives of Computational Methods in Engineering, Vol. 28, pp.937–949, 2021.

[4] Money, C., G., Adewumi, A.O., Obolo., M.O., *Oil Well Characterization and Artificial Gas Lift Optimization using Neural Network combined with Genetic Algorithm*, Discrete Dynamics in Nature and Society, Hindawi Publishing Corporation, Article ID 289239, 10 pp., 2014.

[5] Namdar, H., *Developing an Improved approach to solving a new gas lift optimization problem*, Journal of Petroleum Exploration Production and Technology, No.9, pp.2965-2978, 2019.

[6] Rashid, K., Bailey, W., Couet, B., *A Survey of Methods for Gas lift Optimization*, Modeling and Simulation in Engineering, ID 516807, 16 pp, 2012.

[7] Tavakoli, R., Daryasafar, A, Keyhani, M., Behjoomanesh, M., *Optimization of Gas Lift Allocation using Different Models*, Recent Adv. Petrochem. Sci., Vol.1, No.2, pp.23-29, 2017.

[8] https://scikit-learn.org

.

Utilizarea reţelelor neuronale artificiale la aproximarea curbei de performanţă a unei sonde în gaz-lift

Rezumat. În lucrare, au fost utilizate rețelele neuronale artificiale și o ecuație de regresie pentru a aproxima curba de performanță a sondei în gaz-lift în vederea comparării performanțelor acestora. De asemenea, a fost realizat un studiu de sensibilitate a performanțelor rețelei neuronale artificiale la variația parametrilor geometrici ai acesteia și a funcției de activare. Seturile de date utilizate pentru construirea curbei de performanță a sondei în gaz-lift au caracteristici diferite identificate prin numărul de puncte de date şi prezenţa sau absenţa punctelor de date anormale. În toate cazurile, performanțele rețelei neurale au fost mai bune decât cele ale ecuației de regresie. Cu toate acestea, dacă setul de date conține multe date anormale, rețeaua neuronală artificială, deși are erori mai mici, tinde să construiască o curbă anormală.

MARCU Mariea, Associate Professor, Ph.D. Eng. Petroleum and Gas University of Ploiesti, Department of Well Drilling, Extraction and Transport of Hydrocarbons, Email: mmarcu@upg-ploiesti.ro, + 40 - 244.573 171