Journal of Mining and Earth Sciences

Website: http://jmes.humg.edu.vn

Application of artificial neural network for predicting production flow rates of gaslift oil wells

Hung Tien Nguyen [1,*], Duong Hong Vu [1], Toan Huu To [1], Nhung Thi Nguyen [2]

[1] *Hanoi University of Mining and Geology, Hanoi, Vietnam*
[2] *Petrovietnam Domestic Exploration Production Operating Company Limited, Ho Chi Minh City, Vietnam*

ARTICLE INFO

Article history:
Received 28th Jan. 2022
Revised 04th May 2022
Accepted 04th June 2022

Keywords:
ANN,
Enhence oil recovery,
Gas lift,
Oil production flow rate.

ABSTRACT

In petroleum industry, the prediction of oil production flow rate plays an important role in tracking the good performance as well as maintaining production flow rate. In addition, a flow rate modelling with high accuracy will be useful in optimizing production properties to achieve the expected flow rate, enhance oil recovery factor and ensure economic efficiency. However, the oil production flow rate is traditionally predicted by theoretical or empirical models. The theoretical model usually gives predicted results with a wide variation of error, this model also requires a lot of input data that might be time-consuming and costly. The empirical models are often limited by the volume of data set used to construct the model, therefore predicted values from the applications of these models in practical condition are not highly accurate. In this research, the authors propose the use of an artificial neural network (ANN) to establish a better relationship between production properties and oil production flow rate and predict oil production flow rate. Using production data of 5 wells which use continuous gas lift method in X oil field, Vietnam, an ANN system was developed by using back-propagation algorithm and tansig function to predict production flow rate from the above data set. This ANN system is called a back-propagation neural network (BPNN). In comparison with the oil production flow rate data collected from these studied continuous gas lift oil wells, the predicted results from the constructed ANN achieved a very high correlation coefficient (98%) and low root mean square error (33.41 bbl/d). Therefore, the developed ANN models can serve as a practical and robust tool for oilfield prediction of production flow rate.

**Corresponding author*
E - mail: nguyentienhung.dk@humg.edu.vn
DOI: 10.46326/JMES.2021.63(3).10

1. Introduction

Determination of oil production flow rate, where direct rate measurement is not feasible, is a challenge to every petroleum engineer, especially in late production period wells produced by continuous flow gas lift. It is important and valuable to predict the production flow rate of these wells. The prediction can be used for tracking, and monitoring the wells performance, therefore it is useful in flow rate maintenance. In addition, flow rate modelling with high accuracy will be useful in predicting production properties to adjust and maintain flow rate, enhance oil recovery factor and ensure economic efficiency.

The production flow rate was traditionally predicted by using theory or empirical models. However, theory models require a lot of data as input sources and some of them are not usually collected in practical production since it is time - consuming and costly.

A theory of correlation between production flow rate and other production properties was first introduced by Tangren in 1949 (Tangren et al., 1949). In this model, the fluid is multi-phase flow and the ratio of gas to liquid is less than 1. From the theory correlation of Tangren, empirical correlations were then developed by Gilbert (1954), Baxendell (1958) and Ros (1960).

Gilbert used a 268 data set that included well head pressure, size of the choke, ratio of gas to liquid and critical flow liquid rate to generate an empirical equation as follows (Gilbert, 1954).

$$Q_l = \frac{P_{wh} D_{64}^{b}}{aGLR^{c}} \quad (1)$$

Where: P_{wh} - wellhead pressure (psia); D_{64} – the size of the choke (1/64 inch); GLR – the ratio of gas to liquid (Standard cubic feet/Stock tank barrel (SCF/STB)); Q_l - critical-flow liquid rate (Stock tank barrel per day (STB/day)); a, b and c - the main coefficients evaluated based on sufficient data is available for certain reservoir with a = 435, b = 1.89, c = 0.546.

The empirical equation (1) was then improved for a critical rate calculation by Baxendell (1958) and for a new correlation to derive oil and gas mass under critical flow condition by Ros (1960). A new set of coefficient values a = 0.2618, b = 1.88, c = 0.65 for Gilbert's equation was developed by Achong (1961). Poettmann and Beck (1963) the rearranged model from Ros (1960) to a diagram based on field unit. Also, from Ros (1960), instead of using gas liquid ratio Ashford (1974) established an equation for critical multiphase flow in choke valves by using mixture density. Similarly, Al-Towailib and Al-Marhoun developed a new equation for choke critical flow (Al-Towailib and Al-Marhoun, 1994). Based on extensive field data, Al-Attar and Abdul Majeed proposed new equations for different chokes' bean settings (Al-Attar and Abdul-Majeed, 1988).

All above empirical models were limited by the area of data used in the researches, each model only presents a specific research field. Therefore, these models were not widely used since the results from their applications in other fields did not usually show a good accuracy.

To solve the limitations of theory and empirical, an artificial neural network (ANN) can be used as a good replacement method for connecting and modelling the complicated relationship between production flow rate and other related production properties. Some ANN models for the prediction of production flow rate have shown their high accuracy (Khamehchi et al., 2009; Mirzaei-Paiaman and Salavati, 2012; Gorjaei et al., 2015; Rashida et al., 2019; Azim, 2020; Khan et al., 2020; Ahmed et al., 2021; George, 2021; Barjouei et al., 2021). These models have shown their advantages in comparison with other traditional methods. However, the ANN method has not been applied for oil production wells using the continuous gas lift method. This paper presents the application of the ANN method for the prediction of the production flow rate of several continuous gas lift oil wells.

2. Data processing (Noise removal)

During the training process, abnormal data can be considered as noise since it can affect the accuracy and generality of ANN. Thus, raw data set needs to be cleaned before it can be used as training data set, the data is first corrected and the noise is removed by using an algorithm to identify Z-score (Tripathy et al., 2013). Z-score is identified by equation (2):

$$z = | X_i - X_{mean} | / SD \quad (2)$$

Where: X_{mean} - the arithmetic average of the values, X_i - a value and SD is the standard deflection of value. From Tripathy et al. (2013), the Z-score is used to identify noise in a data set as follow:

- If z < 2, the value is suitable to use;
- If 2 < z < 3, the value could affect the overall result of training;
- If z > 3, the value is noise and needs to be removed.

3. Data analysis

The selections of input data for training ANN is an important step that affects the accuracy of the model. In order to decide on input data, the authors analyze the correlations between production rate and other production properties from 05 production oil wells (01, 02, 03, 04 and 05) in X oil field, Vietnam.

Since R^2 is a statistical measurement that represents the relationship between two variances, a close to 1 of R^2 value indicates a strong correlation between production flow rate and one of the other production properties. The results from Figure 1 indicate that all production data except BS&W does not correlate well with the oil production flow rate, thus they can be used as input data for the ANN training process.

3.1. Data normalized

Since data from the data set is in different scales and units which could affect the efficiency

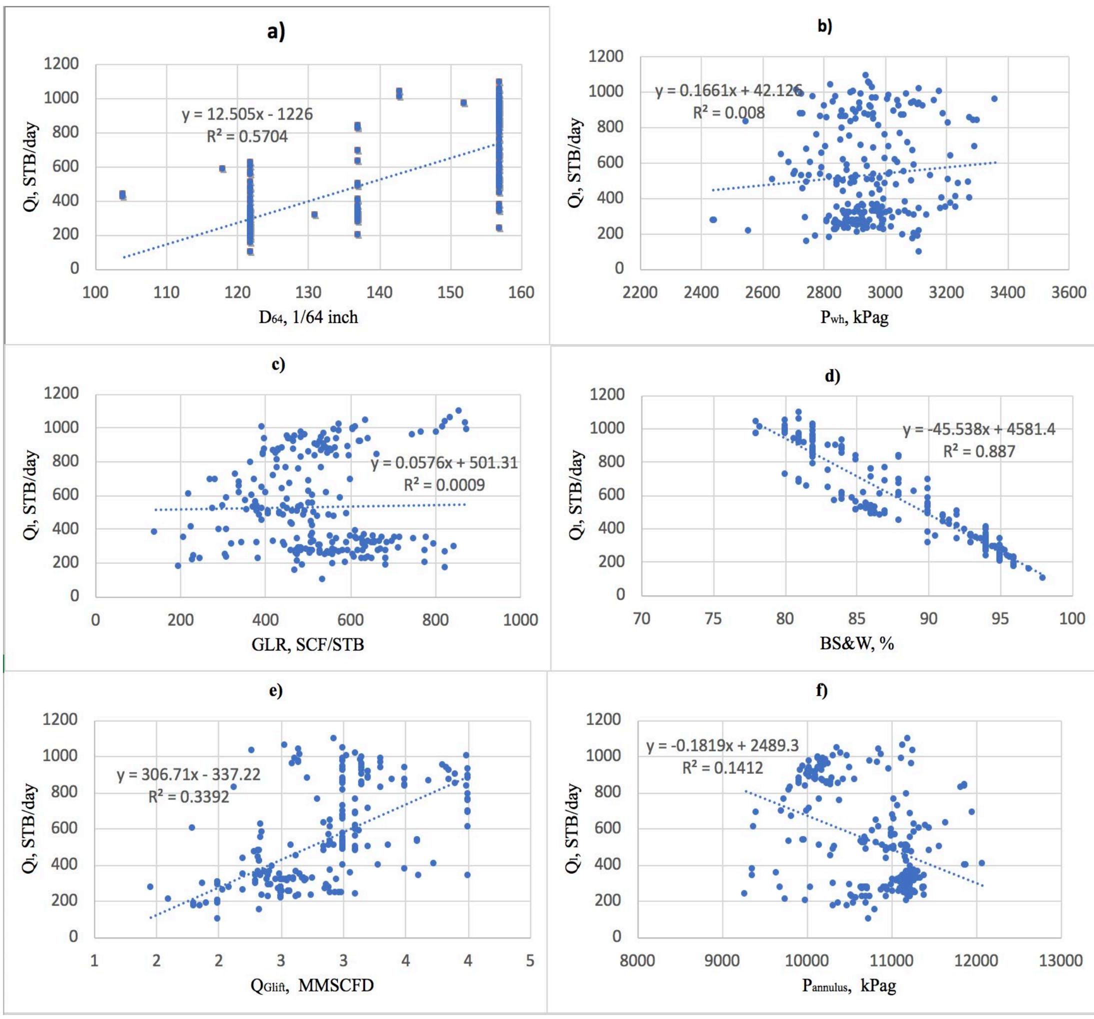

Figure 1. Correlations between production flow rate (Q_l) and a) choke size; b) wellhead pressure (P_{wh}); c) ratio of gas to liquid (GLR); d) Free water, sediment and emulsion (BS&W); e) injected gas lift rate Q_{glift}; f) injected gas lift pressure ($P_{annulus}$).

of the algorithm, practical time and the accuracy of the ANN model, the normalization of the data set is necessary to give all data in the same scale from 0 to 1. The normalized value is calculated as follows:

$$X_{normalize} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (3)$$

Where: $X_{normalize}$ - normalized value; X - the value of data; X_{min} - the minimum value of data; X_{max} is the maximum value of data.

3.2. ANN model construction for the prediction of oil production flow rate

An artificial neural network is a computational model that mimics the way a biological neural signals to another. An ANN system is comprised of a number of neurons that are linked to others for processing information (Mohaghegh, 2000). ANN is inspired by the human brain, an ANN system is improved through a training process, it is also able to memorize experiences and apply the memorized experiences to predict unseen data. An ANN system is usually constructed based on following the assumptions:

- Information is processed at various elements which are called neural;
- The signal is transmitted from a neural to its connected neural;
- Each connection is multiplied by the transmitted signal;
- The output of each neuron is computed by a non-linear function of the sum of its inputs.

Since ANN is able to reproduce and model non-linear process, it has found the application in many disciplines such as information technology, biology, management, economy, and medicine… where the relationships of related factors are non-linear.

In this research, ANN is applied to predict production flow rate of 5 wells that use the continuous gas lift method in X oil field, Vietnam. An ANN system is developed by using the back-propagation algorithm (Mohaghegh, 2000) and the tansig function to predict the production flow rate from the above data set. This ANN system is called a back-propagation neural network (BPNN). The used data set was divided into two folders (Table 1). The first folder includes 231 samples that were collected at studied wells from 2019 to 2020, all samples include data on the production flow rate (Q_1), choke size (D_{64}), wellhead pressure (P_{wh}), the ratio of gas to liquid (GLR), free water, sediment and emulsion (BS&W), injected gas lift rate (Q_{Glift}), injected gas lift pressure ($P_{annulus}$) (Rashida et al., 2019; Azim, 2020; George, 2021; Barjouej et al., 2021). In this first folder, 70% of the samples are used for the training process, 15% of the samples are used for the testing process and the rest 15% are used for validation. The second folder includes 33 samples that were collected at well 04 and 05 in 2021. This second folder was used for verifying the ability of production flow rate prediction of ANN.

In the training process, ANN learn by processing examples, each of which contains known input and output layers. The input layer includes D_{64} (1/64 inch), Pwh (kPag), GLR (SCF/STB), BS&W (%),Q_{Glift} (MMSCFd), P_{Glift} (kPag) while output layer includes Q_l (STB/day). The input data is processed to give output results which are then compared with the target results in output layers for identifying the difference after every cycle. This difference is called error value, the error value then propagates back to neurals in output and hidden layers. The network then adjusts its weighted associations using the error value. The propagation is processed in a number of times until the error value reaches to an allowance minimum value or the number of cycle is equal to a defined number. After a sufficient number, the neural network produces an output that is similar to the target output. This is known as supervised learning. In this research, the authors have constructed and evaluated the back propagation neural network which has the structure as shown in Figure 2 in which the input layer is comprised of 6 neural (D_{64}, Pwh, GLR, BS&W, Q_{Glift}, P_{Glift}) and the output layer has one neural (Q_l). A cost function used in the hidden layer is tansig and this for output layer is pureline.

The difference between various network models is the number of neural in hidden layer usually varies from 4 to 12. The number of neural in hidden layer needs to be carefully selected and it needs to ensure the predicted results of ANN have a good correlation with output samples.

Table 1. Data of 5 studied wells.

Production Properties		First folder of data (2019-2020)	Second folder of data (2021)
Number of samples		231	33
Q_l, STB/day	Minimum value	42.62	198.24
	Maximum value	1156.89	880.16
	Mean value	515.54	549.46
	Standard error	317.77	226.04
D_{64}, 64th of an inch	Minimum value	52	137
	Maximum value	157	157
	Mean value	133.9	149.94
	Standard error	20.43	9.7
P_{wh}, kPag	Minimum value	2295.1	2757
	Maximum value	5887	3074
	Mean value	2968.13	2905.88
	Standard error	284.29	79.33
GLR, SCF/STB	Minimum value	106.75	417.58
	Maximum value	2008.8	890.66
	Mean value	561.33	615.23
	Standard error	219.44	144.68
BS&W, %	Minimum value	75.5	82
	Maximum value	99	96
	Mean value	89.36	89.23
	Standard error	6.01	4.44
Q_{Glift}, MMSCFd	Minimum value	0.45	1.8
	Maximum value	5	4.0
	Mean value	2.71	3.27
	Standard error	0.68	0.94
P_{Glift}, kPag	Minimum value	8247	9865
	Maximum value	12085	11283
	Mean value	10619.91	10630.9
	Standard error	836.45	486.83

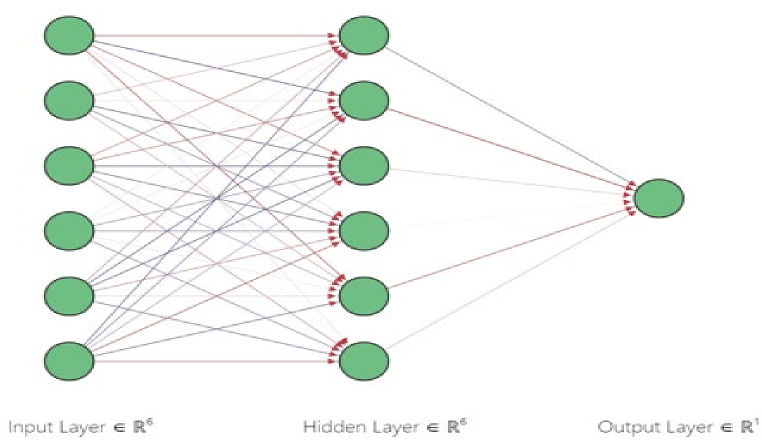

Figure 2. Structure of artifical neural network (ANN).

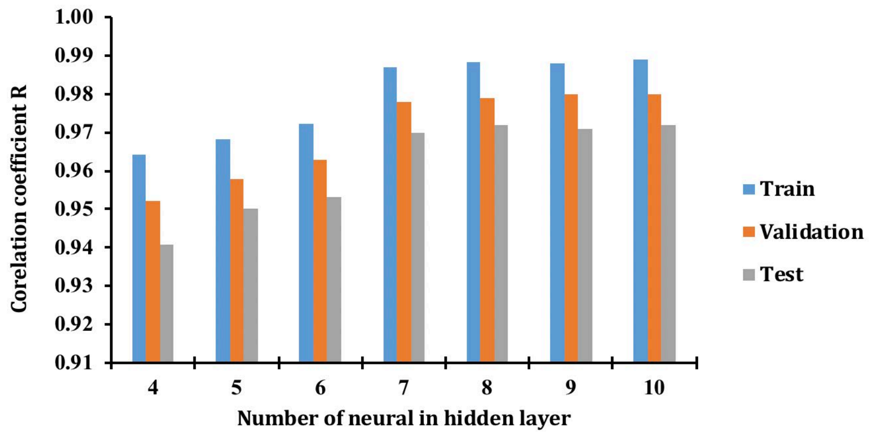

Figure 3. Results of correlation coefficient with different number of neural in the hidden layer.

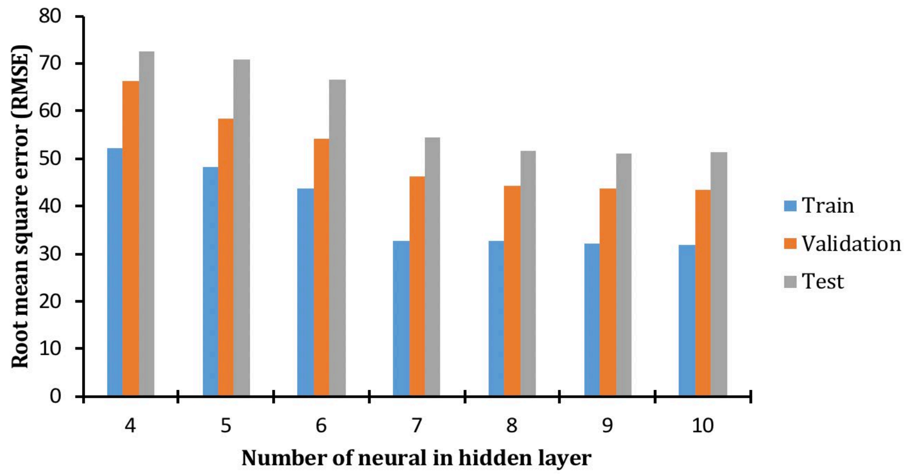

Figure 4. Results of root mean square error (RMSE) with different number of neural in the hidden layer.

Table 2. Summary the results of correlation coefficient and root mean square error (RMSE) from different ANN models.

Number of neural in hidden layer	Correlation coefficient R			Root mean square error RMSE		
	Train	Validation	Test	Train	Validation	Test
4	0.964	0.952	0.941	52.326	66.400	72.654
5	0.968	0.958	0.950	48.230	58.327	70.896
6	0.972	0.963	0.953	43.865	54.037	66.587
7	0.987	0.978	0.970	32.733	46.195	54.453
8	0.988	0.979	0.972	32.686	44.320	51.674
9	0.988	0.980	0.971	32.233	43.643	50.980
10	0.989	0.980	0.972	31.864	43.320	51.354

It also needs to avoid overfitting due to too many neural being used.

The results from various models with a different number of neural in the hidden layer are shown in Figures 3 and 4. From the comparisons of correlation coefficient number (R^2) and root mean square error (RMSE) (Table 2) between different models, the author has decided to use the model with 7 neurals in the hidden layer as the ANN model for predicting production flow of studied wells.

4. Results and Discussions

To evaluate the accuracy of the ANN model, the authors also used the multivariate regression method, as shown in equation (5) (Ghorbani et al, 2019), to calculate the oil production flow rate from input data that was used for the ANN training process then compared the results generated by two methods:

$$Q_l = a_1 D_{64} + a_2 P_{wh} + a_3 GLR + a_4 BS\&W + a_5 Q_{Glift} + a_6 P_{Annulus} + b \quad (5)$$

Where: $a_1, a_2, a_3, a_4, a_5, a_6$, and b - empirical factors whose values are shown in Table 3.

Predicted results from two different models were then compared with actual oil production flow rate to evaluate the accuracy of the model as shown in Figures 5 and 6. The results of the regression coefficient (R^2) and root mean square error (RMSE) are shown in Table 4.

From Table 4, it can be seen clearly that the results of the predicted production flow rate from ANN model are much more accurate than Multivariate Regression models with R^2 value is 0.9854 and RMSE is 33.41 bbl/d.

Lastly, to evaluate the potential application of ANN model constructed in this research, the authors use the model to predict oil production flow rate of oil wells in 2021 by using the second folder of data including 33 samples that were collected at well 01 and 05 in 2021. The results in Figures 7 and 8 clearly show good correlations between predicted and actual data in both wells. The curves in the two figures show very similar

Table 3. Coefficients of equations to determine Q (Multivariate regression method).

Parameter	Coefficients
Intercept (b)	3972.836
a_1	-46.435
a_2	-1.411
a_3	2.206
a_4	0.265
a_5	47.308
a_6	-0.030

Table 4. Model performance comparison.

Model	Root mean square error RMSE	Regression Coefficient R^2
ANN	33.41	0.9854
Multivariate Regression	68.76	0.9353

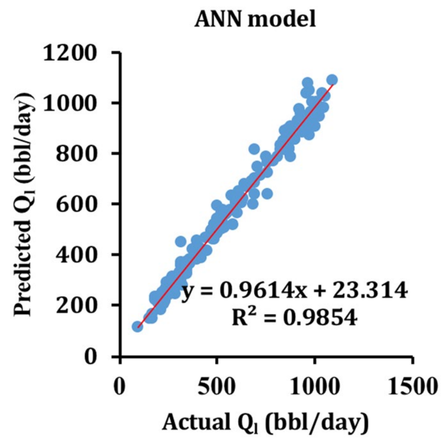

Figure 5. Cross plot of oil flow rate for ANN model (Predicted Q_l).

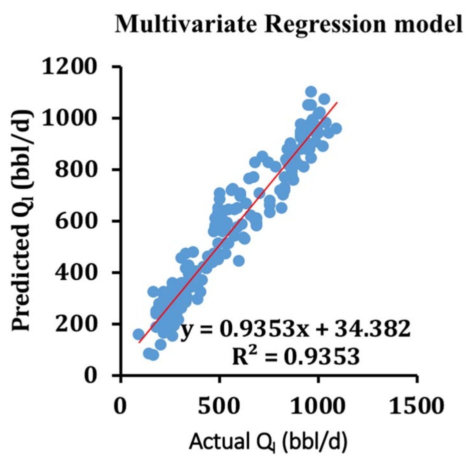

Figure 6. Cross plot of oil flow rate for Multivariate Regression model (Predicted Q_l).

trends even for the 2021 production period in which the data was not used as input data to train the ANN model.

5. Conclusion and recommendations

From the results of this research, the authors summarize conclusions and recommendations as follow:

- The ANN model using 7 neurals in the hidden layer and back-propagation is considered as the best model to predict oil production flow rate in 05 studied continuous gas lift flow wells. The predicted results are much more accurate than the multivariate regression model.

- The ANN model also shows a potential future application for the prediction of the oil production flow rate of other oil wells.

- To improve the accuracy of ANN model, further data set from previous years as well as updated data is needed for further training.

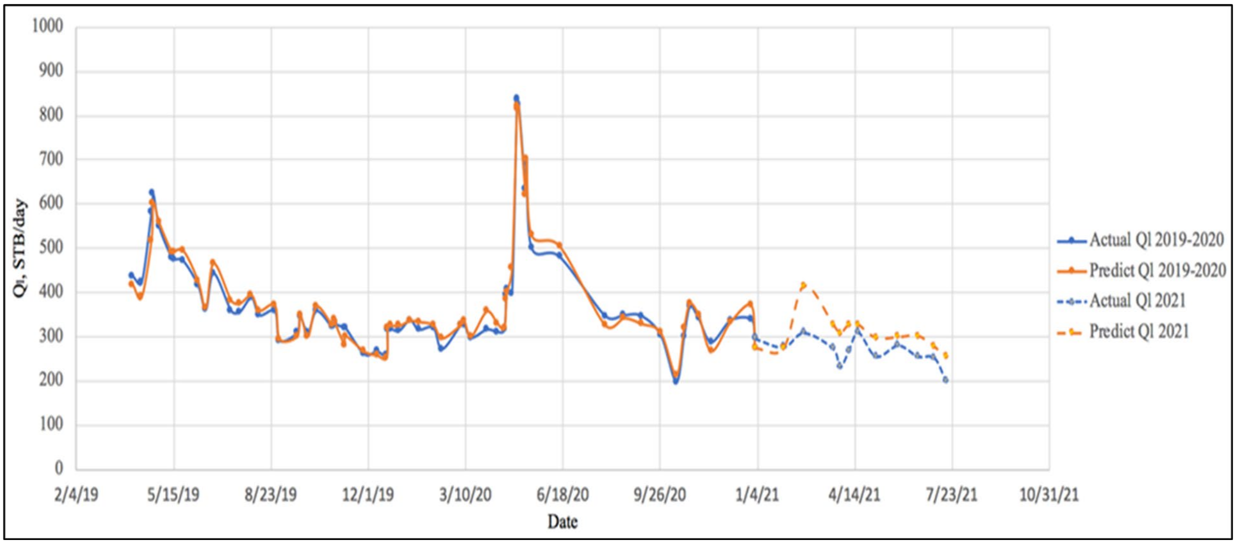

Figure 7. Comparing predicted production flow rate by the ANN model and actual data of well 01 in 2 production period (2019-2020 and 2021).

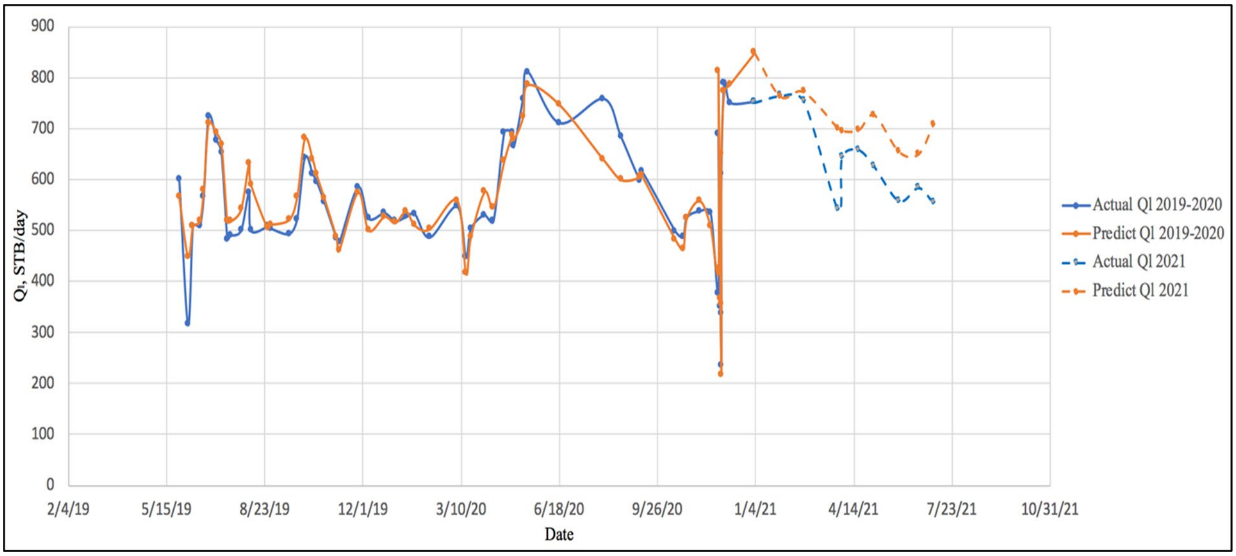

Figure 8. Comparing predicted production flow rate by the ANN model and actual data of well 05 in 2 production period (2019-2020 and 2021).

Acknowledgments

This research is funded by the Hanoi University of Mining and Geology under grant number T21-25.

Contribution authors

Hung Tien Nguyen contributes to the idea, data acquisition and analysis, and writes the manuscript; Duong Hong Vu contributes to edit the writing; Toan Huu To and Nhung Thi Nguyen contribute to collect the data.

References

Achong, I. (1961). *Revised Bean Performance Formula for Lake Maracaibo Wells.* Internal Report, Shell Oil Co. Houston, TX.

Elgibaly, A. A., Ghareeb, M., Kamel, S., & El-Bassiouny, M. E. S. (2021). Prediction of gas-lift performance using neural network analysis. *AIMS Energy*, *9*(2), 355-379.

Al-Attar, H. H. and Abdul-Majeed, G. H. (1988). Revised bean performance equation for East Baghdad oil wells. *SPE Production Engineering*, *3*(1), 127-131.

Al-Towailib, A. I., Al-Marhoun, M. A. (1994). A new correlation for two-phase flow through chokes. *Journal of Canadian Petroleum Technology, 33*(5). https://doi.org/10.2118/94-05-03

Ashford, F.E. (1974). An evaluation of critical multiphase flow performance through wellhead chokes. *Journal of Petroleum Technology*, *26*(8), 843-850.

Azim, R.A. (2020). Prediction of multiphase flow rate for arti cially owing wells using rigorous articial neural network technique. *Flow Measurement and Instrumentation*, *76*, 101835.

Barjouei, H.S., Ghorbani, H., Mohamadian, N., Wood, D.A., Davoodi, S., Moghadasi, J. and Saberi, H. (2021). Prediction performance advantages of deep machine learning algorithms for two-phase flow rates through wellhead chokes. *Journal of Petroleum Exploration and Production*, *11*(3), 1233-1261.

Baxendell, P.B. (1958). Producing wells on casing flow-an analysis of flowing pressure gradients. *Transactions of the AIME*, *213*(1), 202-206.

Gilbert W.E. (1954). *Flowing and gas-lift well performance.* In Drilling and Production practice, American Petroleum Institute. API-801-30H.

George, A. (2021). Predicting Oil Production Flow Rate Using Artificial Neural Networks-The Volve Field Case. *In SPE Nigeria Annual International Conference and Exhibition.* OnePetro. Lagos, Nigeria. https://doi.org/10.2118/208258-MS

Gorjaei, R. G., Songolzadeh, R., Torkaman, M., Safari, M., & Zargar, G. (2015). A novel PSO-LSSVM model for predicting liquid rate of two phase flow through wellhead chokes. *Journal of Natural Gas Science and Engineering*, *24*, 228-237.

Ghorbani, H., Wood, D.A., Moghadasi, J., Choubineh, A., Abdizadeh, P., & Mohamadian, N. (2019). Predicting liquid flow-rate performance through wellhead chokes with genetic and solver optimizers: an oil field case study. *Journal of Petroleum Exploration and Production Technology*, *9*(2), 1355-1373. https://doi.org/10.1007/s13202-018-0532-6.

Khamehchi, E., Rashidi, F., & Rasouli, H. (2009). Prediction of gas lift parameters using artificial neural network. *Enhanced Oil Recovery–Iranian Chemical Engineering Journal (Special Issue)*, *8*(43).

Khan, M.R., Tariq, Z. and Abdulraheem, A. (2020). Application of artificial intelligence to estimate oil flow rate in gas-lift wells. *Natural Resources Research*, *29*(6), 4017-4029. doi:10.1007/s11053-020-09675-7

Mirzaei-Paiaman, A., & Salavati, S. (2013). A new empirical correlation for sonic simultaneous flow of oil and gas through wellhead chokes for Persian oil fields. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, *35*(9), 817-825.

Mohaghegh, S. (2000). Virtual-intelligence applications in petroleum engineering: Part 1 -

Artificial neural networks. *Journal of Petroleum Technology, 52*(9), 64-73.

Poettmann, F. H., & Beck, R. L. (1963). New charts developed to predict gas-liquid flow through chokes. *World Oil*, *184*(3), 95-100.

Rashid, S., Ghamartale, A., Abbasi, J., Darvish, H., & Tatar, A. (2019). Prediction of critical multiphase flow through chokes by using a rigorous artificial neural network method. *Flow Measurement and Instrumentation*, *69*, 101579.

Ros, N. C. J. (1960). An analysis of critical simultaneous gas/liquid flow through a restriction and its application to flowmetering. *Applied Scientific Research, 9*(1), 374-388. http://doi.org/10.1016/j.flowmeasinst.2020.101835

Tangren, R. F., Dodge, C. H., & Seifert, H. S. (1949). Compressibility effects in two-phase flow. *Journal of Applied physics, 20*(7), 637-645.

Tripathy, S. S., Saxena, R. K., & Gupta, P. K. (2013). Comparison of statistical methods for outlier detection in proficiency testing data on analysis of lead in aqueous solution. *American Journal of Theoretical and Applied Statistics*, *2*(6), 233-242.