PETROVIETNAM JOURNAL
Volume 6/2019, p. 48 - 57
ISSN-0866-854X

An applied machine learning approach to production forecast for basement formation - Bach Ho field

Tran Dang Tu[1], Nguyen The Duc[1], Le Quang Duyen[2], Pham Truong Giang[1], Le Vu Quan[1], Le Quoc Trung[1], Tran Xuan Quy[1], Pham Chi Duc[1]
[1]Vietnam Petroleum Institute
[2]Ha Noi University of Mining and Geology
Email: tutd@vpi.pvn.vn

Summary

Oil production forecast is a major challenge in the oil and gas industry. Simulation model and prediction results play an important role in field operation and management. Currently, production forecast problems are resolved mainly by using pure traditional prediction methods. Generally, production forecast by dynamic simulations does not provide reliable results in case where a lot of uncertain parameters remain when the dynamic model is constructed.

In fact, in Vietnam, the dynamic models of fractured reservoirs give unreliable results and differ with actual performance. It is a challenge to build and design reasonable production plans for fractured granite reservoirs in Vietnam. In order to replace the disadvantages of simulation model by different methods, a growing trend of research in the world is constructing predictive tools by using machine learning algorithms.

The paper introduces the applicability of machine learning through the artificial neural network to predict oil production for basement formation - Bach Ho field. The research results show that Artificial Neural Network (ANN) model has improved the ability to predict production with high accuracy.

Key words: Artificial Neural Network, machine learning, oil production, reservoir management, Bach Ho field.

1. Introduction

Machine learning (ML) is the scientific study of algorithms and statistical models that computer systems use to effectively perform a specific task without using explicit instructions, relying on models and inference instead. It is seen as a subset of artificial intelligence. Machine learning algorithms build a mathematical model of sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to perform the task [2, 3]. Machine learning is closely related to computational statistics, which focuses on making predictions using computers. The study of mathematical optimisation delivers methods, theory and application domains to the field of machine learning. Data mining is a field of study within machine learning and focuses on exploratory data analysis through unsupervised learning. In its application across business problems, machine learning is also referred to as predictive analytics [4].

Date of receipt: 20/2/2019. Date of review and editing: 20/2 - 1/4/2019.
Date of approval: 3/6/2019.

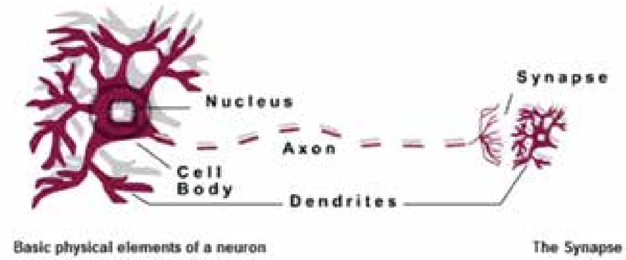

Figure 1. *Basic physical elements of a biological neuron [1].*

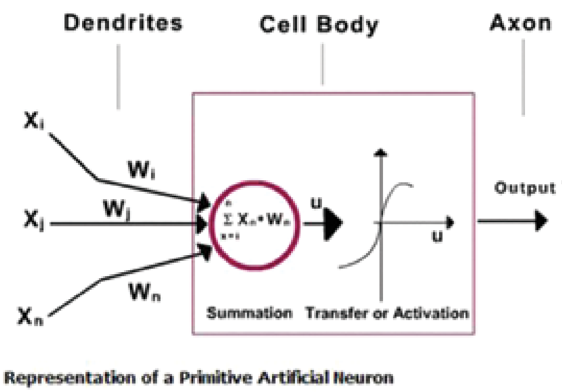

Figure 2. *Representation of neuron in ANN [1].*

One of the most popular machine learning methods - ANN is employed for this purpose. In computer science, ANN is formed of computer architecture, inspired by biological neural networks (the central nervous systems of animals, particularly, the brain) and used to estimate or approximate functions that can depend on a large number of inputs and are generally unknown. ANN is generally presented as systems of interconnected "neurons" which can compute values from inputs and are capable of machine learning or pattern recognition, thanks to their adaptive nature. Figures 1 and 2 show the basic biological neuron structure and representation of artificial neuron.

2. Neural networks

The most popular ANN model is the multi-layer perceptron (MLP) architecture trained using the feedforward backpropagation algorithm. The MLP architecture is composed of at least three layers vector and the last layer consists of the output vector. The intermediate layers, called hidden layers, represent neural pathways and modify the input data through several weighted connections.

There are three major phases to network training with backpropagation. During the initial phase, the input vector is presented to a network, which is activated via the forward pass. This generates a difference (error) between the input of the network (error backward pass). During the output layer back, through the hidden layers, to the input layer. This process is repeated until the connection weights produce an output which is with a predetermined tolerance of the desired output [2].

The selection of an optimum architecture of a model is a difficult task requiring a procedure of trial and error [5]. Thus, several networks with various numbers of hidden units, training algorithms, and activation functions are attempted and the generalisation error is estimated for each. The network with the minimum estimated generalisation error is chosen.

3. Production data of basement formation of Bach Ho field

The basement formation of Bach Ho field has produced commercially since 1988. Based on the well test results of wells 2, 401, 401, and 417, which were the first exploration, appraisal, and production wells, the initial reservoir pressure was 417atm at 3,650m TVDSS. In the first production stage, reservoir drive mechanisms were rock or compaction drive and solution gas without water drive and water injection supply. After several years of production the reservoir pressure decreased significantly to 280atm. Pressure maintenance by water injection was initiated in 1993 when a few production wells were converted to injection wells and connected with the water injection system. As of May

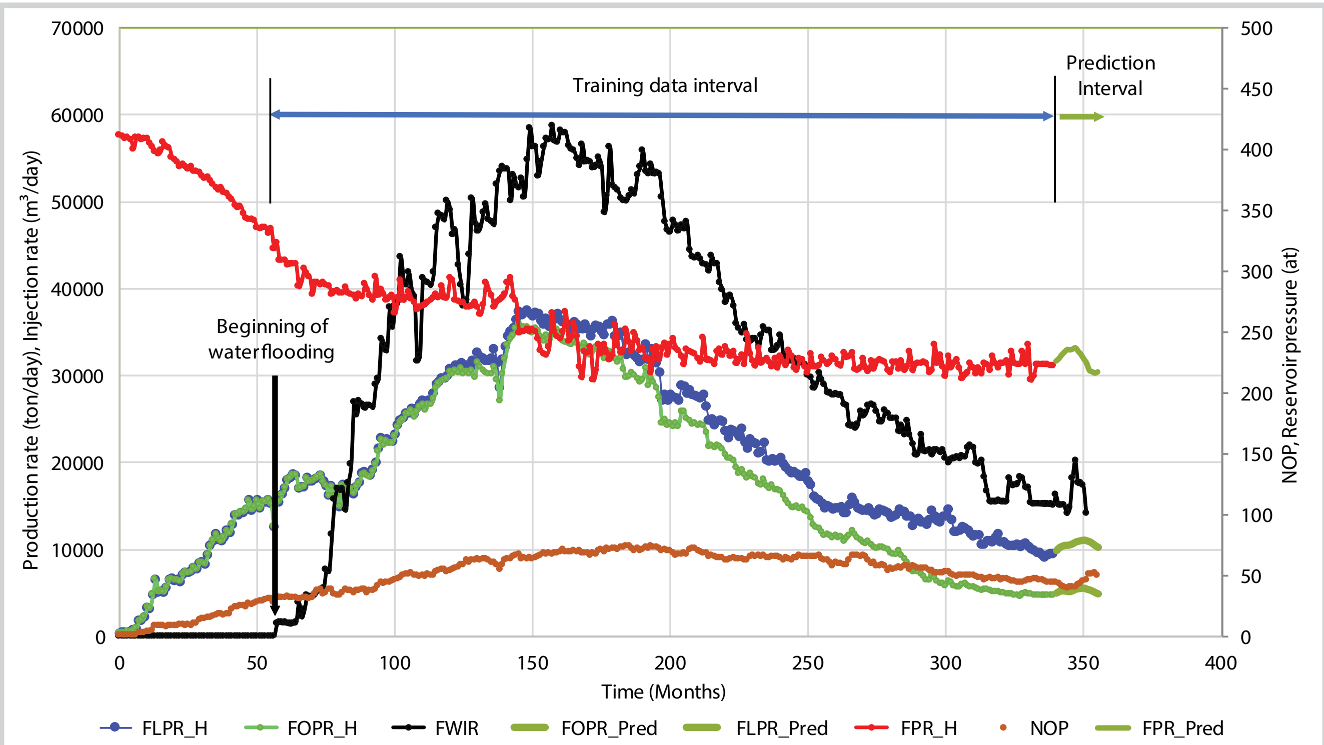

Figure 3. *Reservoir oil production from September 1988 to May 2018.*

2018, the reservoir had achieved a cumulative oil production of 180 million tons, accounting for 86% of Vietsovpetro's total oil production, with an average oil rate of 6,000 tons per day, and an average water cut of 60% [6].

4. Network architecture

Neuron architecture is composed of five inputs and three outputs. The inputs are the average field oil production rate (FOPR) at time t, the average field liquid production rate (FLPR) at time t, the average reservoir pressure (FPR) at time t, the average water injection rate (FWIR) at time t+1 and the number of production wells (NP) at time t+1. The outputs are the average field oil production rate (FOPR) at time t+1, the average field liquid production rate (FLPR) at time t+1, the average reservoir pressure (FPR) at time t+1. The selection of an optimum neuron network architecture can be achieved using a trial and error approach. Figure 3 shows the oil production rate from September 1988 to May 2018.

4.1. Short-term production prediction

- Data pre-processing

Normally, an accurate network model can be achieved without adequate data. Therefore, before training model, production data have to guarantee high reliability to avoid peculiar answers from trained network model. However, depending on the problem, there may be special features from the data that are able to test its quality. One way to check the quality is to view the graphical representations of the data in question, in the hope of selecting a reasonable subset while eliminating problematic portions. As presented in Figure 3, the oil field production rate is time-dependent and was split into two sets. The first set (from May 1993 to December 2016) used 284 data months to build the network model. The second set (from January 2017 to May 2018) used 15 data months to predict the average oil production rate, liquid production rate, and reservoir pressure.

To avoid overfitting or underfitting results and improve the generalisation of the network model, the first set was subdivided randomly into three parts: training, validation, and testing. The training set used 190 data months (67%) to compute the gradient and update the network weights and biases. The validation set used 47 data months (16.5%) to evaluate the quality of the training process. Training can be stopped when the performance of the model on the validation dataset provides a minimum error. The testing set used 47 data months (16.5%) to fine-tune the network model. It is not applied for training and validation process, only used to identify optimum network architecture, to select a suitable network model and assess their performance.

- ANN network architecture

The best results were obtained from ANN model consisting of 2 hidden layers and 50 neurons for each one. The node in the hidden and output layers is activated through Sigmoid function and trained by the Backpropagation Neural Network algorithm (BPNN).

4.2. Long-term production prediction

- Data pre-processing

The first set used 236 data months (from May 1993 to December 2012) to build a network model. The second set used 60 data months (from January 2013 to December 2017) to predict oil production rate, liquid production rate, and reservoir pressure. The training set used 160 data months (67%) to calculate gradient and update the network weights and biases. The validation set used 38 data months (16.5%) to evaluate the quality of the training process. Training can be stopped when the performance of the model on the validation dataset provides a minimum error. The testing set used 38 data months (16.5%) to fine-tune the network model. It is not applied for training and validation process, only used to identify optimum network architecture, to select a suitable network model and assess their performance.

- ANN network architecture

The best results were obtained from ANN model consisting of 1 hidden layer and 60 neurons for each one. The node in the hidden and output layers is activated through Sigmoid function and trained by the Backpropagation Neural Network algorithm (BPNN)

5. Assessing and comparing the production prediction results of ANN model and those of the dynamic simulation model

5.1. Evaluating short-term production prediction results from the ANN model

The statistic method is used to assess the accuracy of the ANN model in the training, validation, and testing processes (Table 1) through the average absolute error (AE) and average relative error (ARE) of three parameters: oil production rate, liquid production rate, and reservoir pressure:

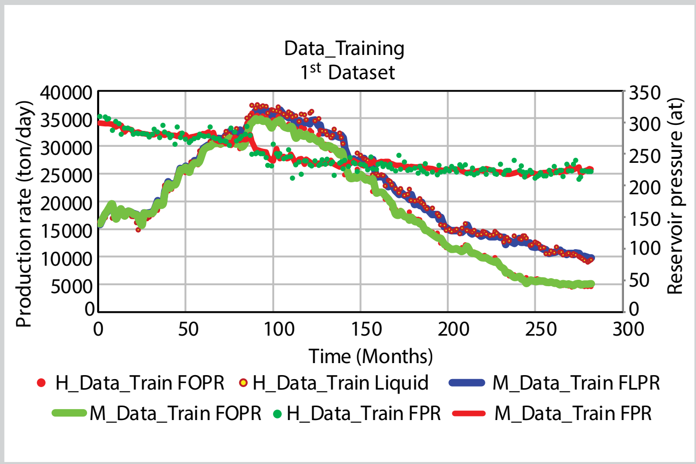

Figure 4. *Performance of short-term prediction training set.*

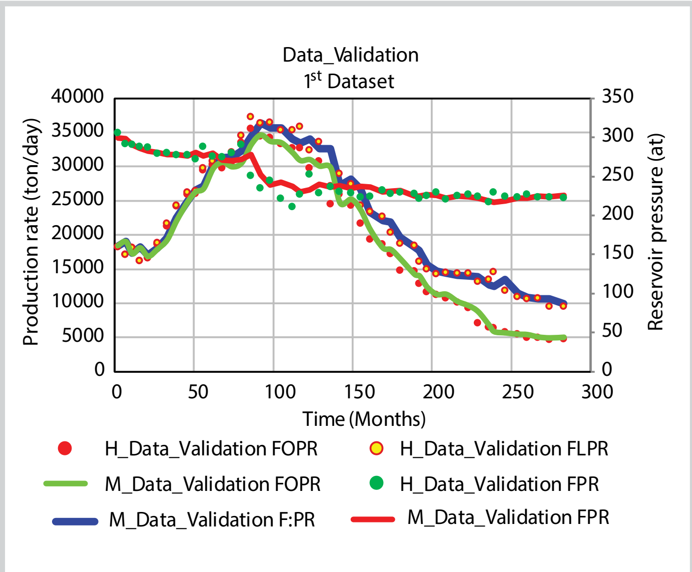

Figure 5. *Performance of short-term prediction validation set.*

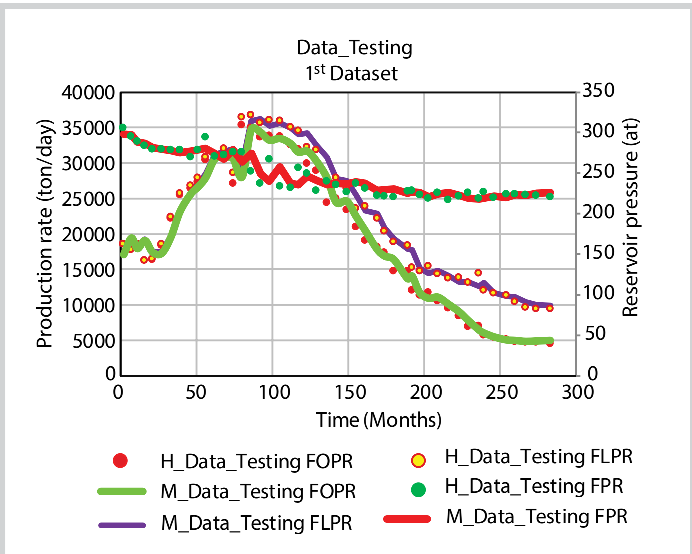

Figure 6. *Performance of short-term prediction testing set.*

- Training set:

+ AE: 526 tons/day, 637 tons/day, 6at;

+ ARE: 3.11%, 3.13%, 2.47%;

- Validation set:

+ AE: 998 tons/day, 1112 tons/day, 6.67at;

+ ARE: 5.51%, 5.26%, 2.76%;

- Testing set:

+ AE: 1157 tons/day, 1165 tons/day, 6.12at;

+ ARE: 6.46%, 5.54%, 2.5%.

The errors are in the allowable limit. The results of training, validation and testing processes are described in Figures 4, 5 and 6.

To study the robustness and accuracy of the network approach, with respect to predicting oil reservoir production, the second dataset was used to predict the reservoir oil production. The predicted reservoir oil rate values agree with the historical values indicating the training network can serve as a practical robust reservoir production management tool (Figure 7). The network provides reservoir oil rates with an average AE of 255 tons/day and average ARE of 4.82%, as illustrated in Table 1.

5.2. Evaluating long-term production prediction results from the ANN model

The statistic method is used to assess the accuracy of ANN model in the training, validation, and testing processes (Table 1) through the average absolute error (AE) and average relative error (ARE) of three parameters: oil production rate, liquid production rate, and reservoir pressure:

- Training set:

+ AE: 553 tons/day, 644 tons/day, 5.25at;

+ ARE: 2.79%, 2.78%, 2.1%;

- Validation set:

+ AE: 1001 tons/day, 1025 tons/day, 6.34at;

+ ARE: 4.91%, 4.4%, 2.52%;

- Testing set:

+ AE: 1215 tons/day, 1261 tons/day, 7.69at;

+ ARE: 5.6%, 5.43%, 3.13%.

Table 1. Statistical analysis of network model accuracy for short-term production prediction

		FOPR H	FOPR ANN	AE1	ARE1 (%)	FLPR H	FLPR ANN	AE2	ARE2 (%)	FPR H	FPR ANN	AE3	ARE3 (%)
DATABASE I	TRAINING												
	Average	19523	19421	526	3.11	22410	22273	637	3.13	245	245	6.00	2.47
	Standard deviation	10034	9914	485	2.79	8815	8656	522	2.63	26	25	5.37	2.23
	Minimum	4521	4765	4	0.01	9081	9719	10	0.04	210	216	0.08	0.03
	Maximum	35959	34902	2496	17.84	37452	36707	2765	14.81	309	298	33.11	15.62
	VALIDATION												
	Average	19469	19558	998	5.51	22289	22270	1112	5.26	245	247	6.67	2.76
	Standard deviation	9973	9649	995	4.92	8772	8411	1020	4.45	28	26	7.08	3.02
	Minimum	4669	4901	15	0.23	9438	9902	20	0.14	211	216	0.16	0.06
	Maximum	35478	34648	5407	23.20	37204	36382	5480	20.22	305	299	27.86	12.69
	TESTING												
	Average	19380	19523	1157	6.46	22216	22250	1165	5.54	246	247	6.12	2.50
	Standard deviation	10015	9794	1281	4.98	8765	8531	1218	4.72	27	25	6.47	2.59
	Minimum	4525	4900	6	0.11	9464	9947	65	0.19	217	218	0.02	0.01
	Maximum	35367	34801	7271	20.56	36778	36267	6883	18.88	306	298	26.70	10.18
DATABASE II	PREDICTION												
	Average	5280	5277	254.50	4.82	11095	10615	573	5.00	220	228	10.38	4.83
	Standard deviation	350	173	111.78	2.13	726	312	504	4.13	7	7	8.68	4.15
	Minimum	4692	4951	30.54	0.56	9464	9947	66	0.63	206	216	0.35	0.15
	Maximum	5848	5538	434.59	8.71	12097	11041	1637	13.54	234	237	24.78	11.83

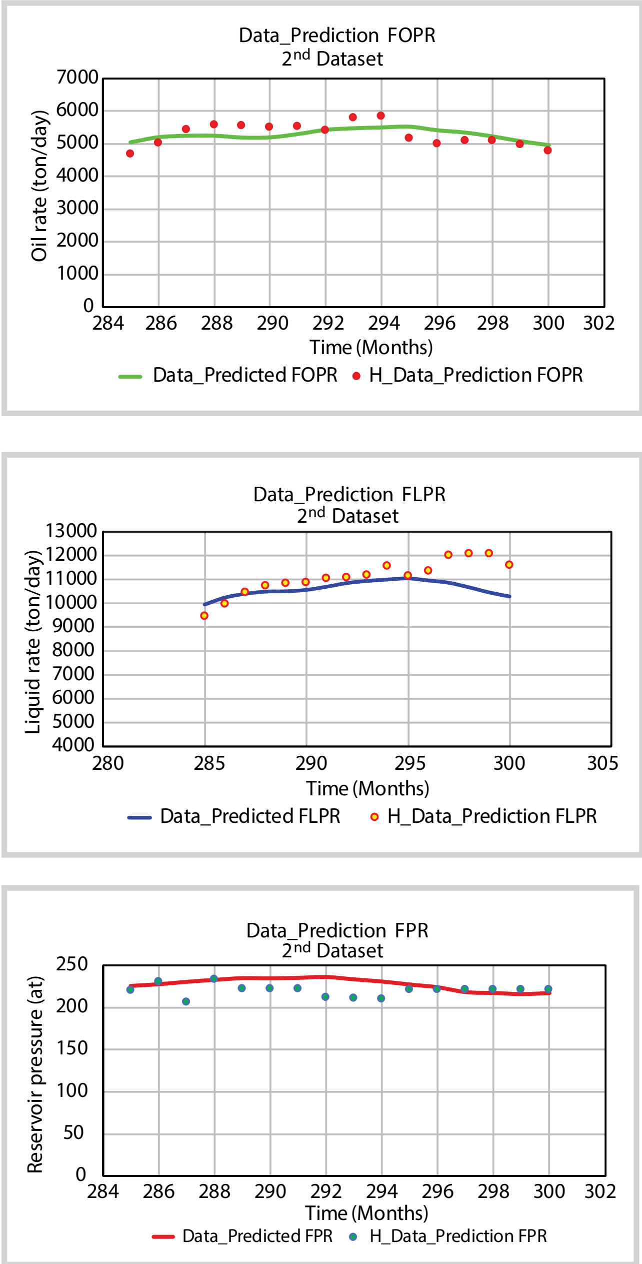

Figure 7. *Prediction of average oil production rate, liquid production rate and reservoir pressure (from January 2017 to April 2018).*

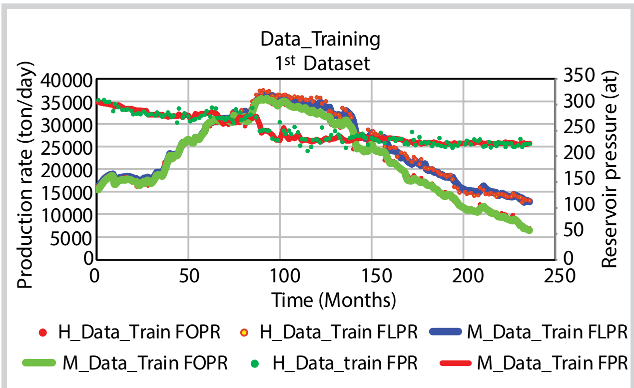

Figure 8. *Performance of long-term prediction training set.*

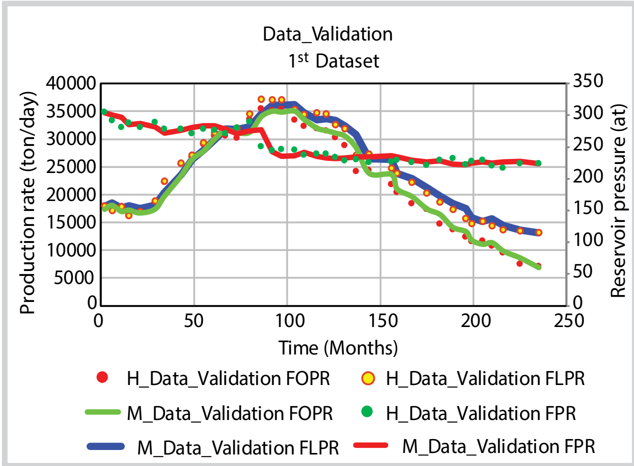

Figure 9. *Performance of long-term prediction validation set.*

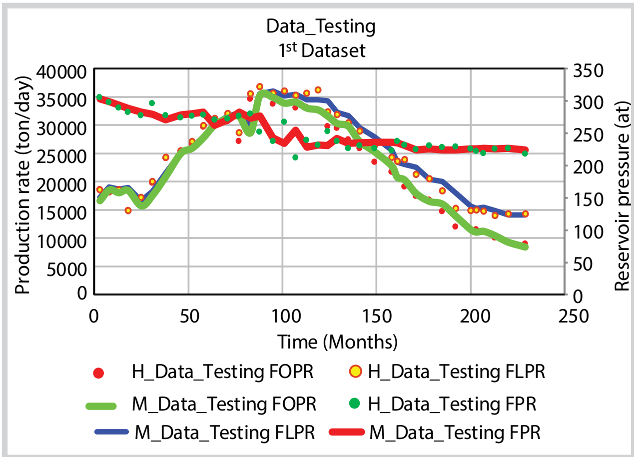

Figure 10. *Performance of long-term prediction testing set.*

The errors are in the allowable limit. The results of training, validation and testing processes are described in Figures 8, 9, and 10.

To study the robustness and accuracy of the network approach, the second dataset was used to predict the reservoir oil production. The predicted reservoir oil rate values agree with the historical values indicating the training network can serve as a practical robust reservoir production management tool (Figure 11). The network provides reservoir oil rates with an average AE of 698 tons/day and average ARE of 12.61%, as illustrated in Table 2.

5.3. Comparing the production prediction results of ANN model and dynamic simulation model results in the short term and in the long term

- Comparing the results of short-term production prediction and those of long-term production prediction

From Figures 12 and 13, it is obvious that short-term oil production prediction of ANN model (284 data

Table 2. *Statistical analysis of network model accuracy for long-term production prediction*

		FOPR H	FOPR ANN	AE1	ARE1 (%)	FLPR_H	FLPR ANN	AE2	ARE2 (%)	FPR H	PR ANN	AE3	ARE3 (%)
DATASET I	**TRAINING**												
	Average	22342	22302	553	2.79	24599	24498	644	2.78	250	249	5.25	2.10
	Standard deviation	8365	8571	478	2.45	7820	7475	551	2.45	27	26	5.41	2.14
	Minimum	6864	6498	6	0.06	12650	12748	9	0.04	211	223	0.02	0.01
	Maximum	35959	35569	3941	16.11	37452	36606	3693	13.62	309	304	34.37	12.94
	VALIDATION												
	Average	22430	22633	1001	4.91	24633	24668	1025	4.40	250	252	6.34	2.52
	Standard deviation	8729	8666	844	3.97	8149	7644	794	3.45	27	27	5.01	1.95
	Minimum	7083	6852	58	0.18	13167	13137	12	0.08	217	222	0.24	0.11
	Maximum	22430	22633	1001	4.91	24633	24668	1025	4.40	250	252	6.34	2.52
	TESTING												
	Average	22563	22753	1215	5.60	24779	24796	1261	5.43	250	251	7.69	3.13
	Standard deviation	8236	8300	1307	5.68	7804	7367	1216	5.53	27	26	9.02	3.89
	Minimum	8935	8453	30	0.27	14021	14080	95	0.38	212	224	0.25	0.11
	Maximum	34899	35255	5887	23.91	36778	36052	5963	27.26	306	303	42.64	20.11
DATASET II	**PREDICTION**												
	Average	5405	5238	697.84	12.61	11302	11899	1254	11.44	222	241	19.60	8.94
	Standard deviation	630	863	559.05	10.08	1369	638	761	7.29	7	15	16.88	7.90
	Minimum	4521	4419	16.73	0.26	9081	11281	9	0.08	206	224	0.00	0.00
	Maximum	7031	7697	2628.28	52.67	14553	13755	2758	25.71	240	278	66.92	31.73

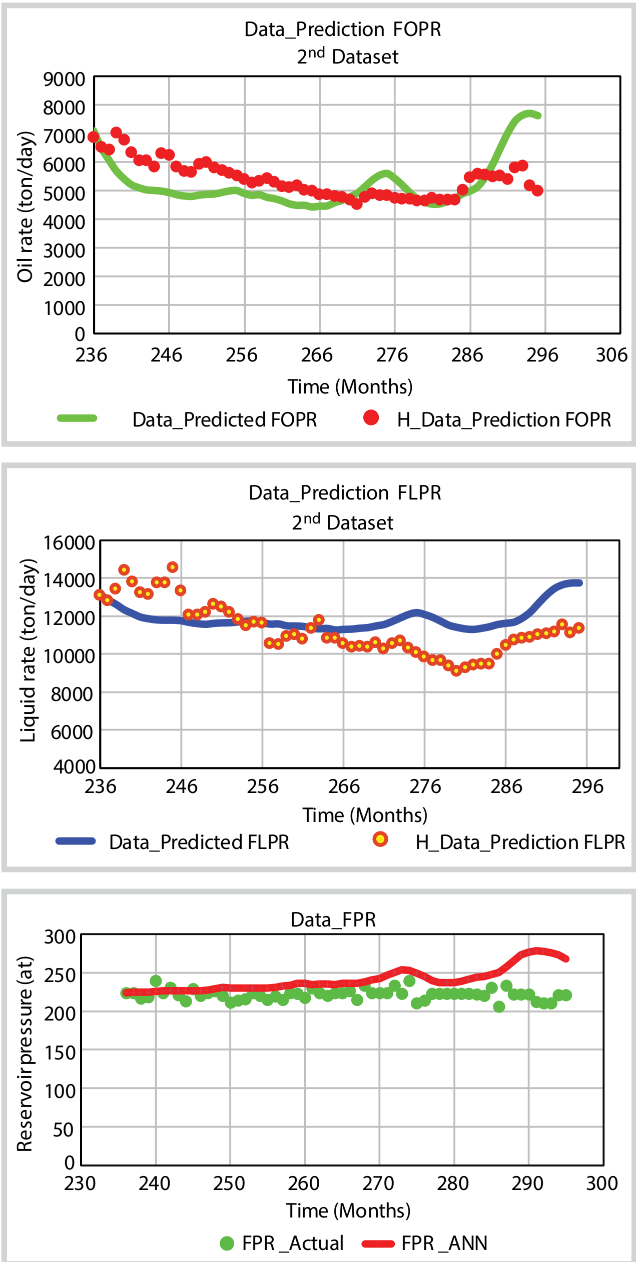

Figure 11. Prediction of average oil production rate, liquid production rate and reservoir pressure (from January 2013 to December 2017).

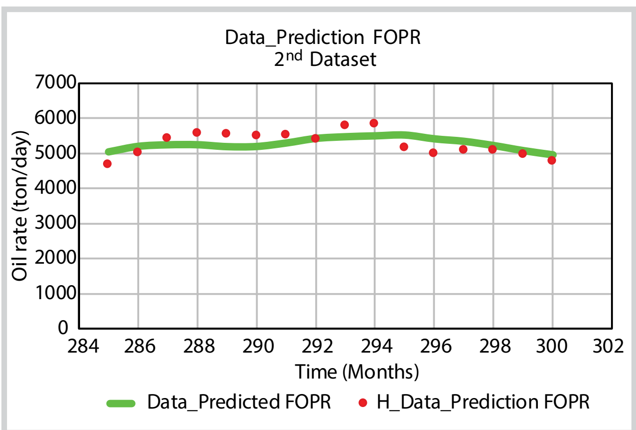

Figure 12. Short-term oil production rate prediction.

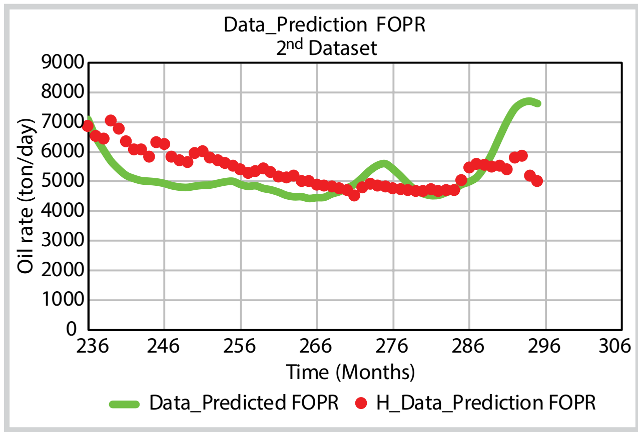

Figure 13. Long-term oil production rate prediction.

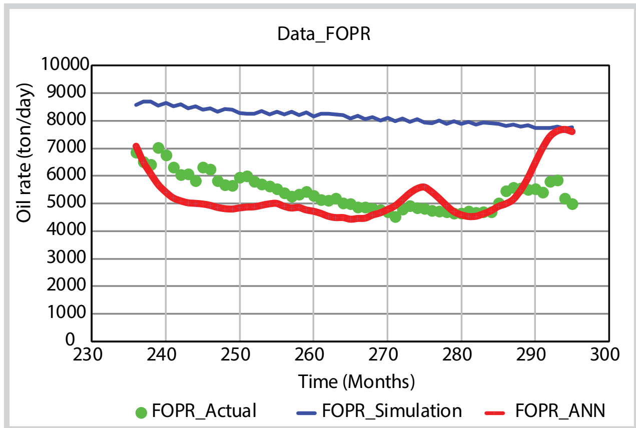

Figure 14. The results of oil production rate prediction.

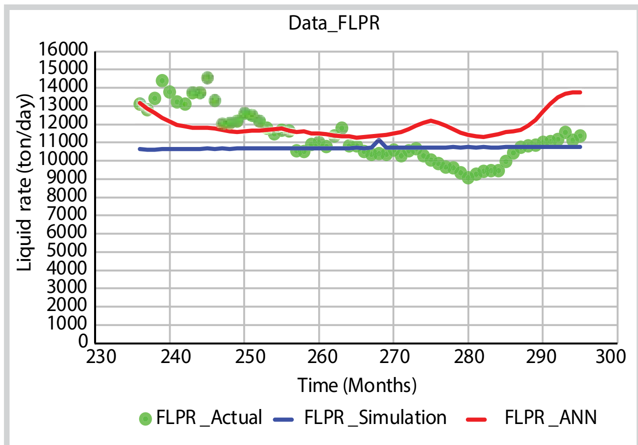

Figure 15. The results of liquid production rate prediction.

months) gives better results than long-term oil production prediction of ANN model (236 data months).

- Comparing the results of long-term production prediction of ANN model and dynamic simulation model results

- In comparison with traditional prediction method, the artificial neural network method can learn arithmetic problems which the in-out relationship is non-linear with

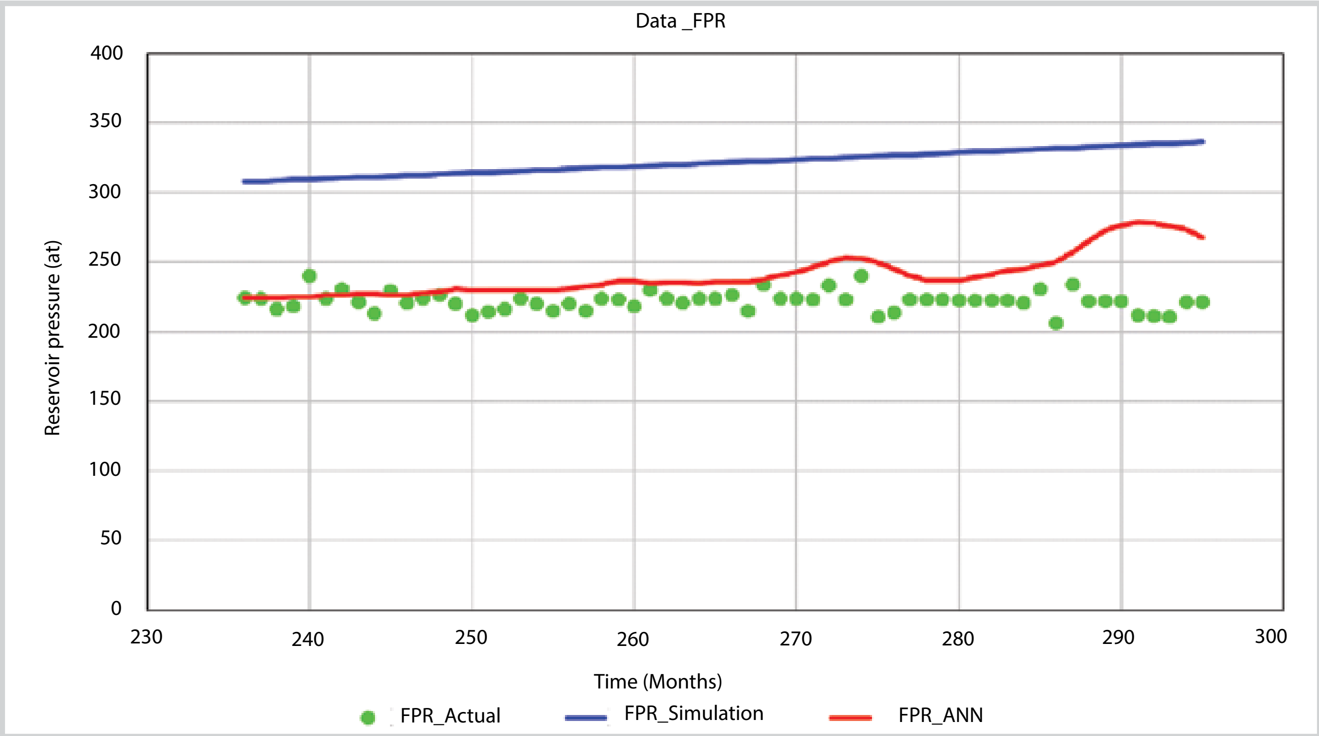

Figure 16. *The results of reservoir pressure prediction.*

high accurate prediction, corresponding to production input data. The ANN method forecasts input data without being based on the subjective experiment of professors. After the training process, the ANN model will actively determine weights for each input parameter and their relationship. Accordingly, the results of the ANN model are more trustworthy than the traditional prediction method.

- In the training set, the network will regulate input parameters to satisfy mean squared error value, ANN's convergence ability depends on original arguments, so many sensitivity scenarios must be run to choose the best original arguments. On the other hand, the training in complicated networks becomes more difficult than shallow and narrow networks, in which case the optimisation is more likely to converge to some useless local optima. Ideally, we would like to design a model of reasonable complexity but powerful representation for the data we feed into it. Moreover, to avoid overfitting the model, the size of the training data has also to be considered in the designing. Therefore, taking all these concerns into account and after several trials on the validation dataset, one layer of the hidden layer with the proper number of neurons fits the best.

- Figures 14, 15 and 16 show comparison between the results of long-term production prediction of the ANN model and those of the dynamic simulation model.

Observation: From the experimental results, the performance of the two models was assessed, showing:

- Dynamic simulation model:

The model is controlled by a constant value of liquid production rate (LPR) in the period from January 2013 to December 2017. As a matter of fact, LPR did not remain stable due to production operation (well shut-in, facility maintenance, weather conditions…). The precise assumption of production qualification depends on experiment, simulator's subjective and field development plan. On the other hand, the simulation oil production rate has high deviation compared with history at starting forecasting date, the model does not obtain reliability and neither does it capture geological complication, rock property distribution, fracture network, and hydrodynamic connectivity of granite basement. Until now, there is not a granite basement simulation method that is accurate, reliable and widely recognised.

- ANN model:

The parameters: oil production rate, liquid production rate, and reservoir pressure are very close to the actual data, the trend of results and actual production match closer than the dynamic model. Nevertheless, the confinement of the ANN model only applies to predict short term.

6. Conclusion and recommendation

This research work aims to present a new approach to predict oil production rate based on the historical production data. The results of methodology show prediction problem generalisation ability on the ANN model, become an effective implement to resolve variable problems in operation and management field production techniques. The ANN model has many features: data learning possibility, adaptation, a decision with deficiency or noise data, which are a significant advantage compared with numerical simulation.

- ANN application will be more effective when the first stages such as training samples, extracting characteristics and pre-processing are well done;

- ANN model postulates more time to train and adjust network argument;

- As for the future work, other particularly different algorithms and input data effected to production prediction such as well bottom hole pressure, choke size, and gas lift rate will be integrated.

Acknowledgement

This work is supported by the Ministry of Industry and Trade of Vietnam, project No. CNKK.001/19.

References

1. Q.Cao, R.Banerjee, S.Gupta, J.Li, W.Zhou, B.Jeyachandra. *Data driven production forecasting using machine learning*. SPE Argentina Exploration and Production of Unconventional Resources Symposium, Buenos Aires, Argentina. 1 - 3 June, 2016.

2. Brian D.Ripley. *Pattern recognition and neural networks*. Cambridge University Press. 1996.

3. John R.Koza, Forrest H.Bennett, David Andre, Martin A.Keane. *Automated design of both the topology and sizing of analog electrical circuits using genetic programming*. Springer. 1996: p. 151 - 170.

4. Jerome H.Friedman. *Data mining and statistics: What's the connection*. Computing Science and Statistics. 1997.

5. Jeff Heaton. *Introduction to neural networks for java (2nd edition)*. Heaton Research. 2008.

6. Trần Văn Hồi, Nguyễn Văn Đức, Phạm Xuân Sơn. *Tìm kiếm thăm dò và phát triển dầu trong đá móng mỏ Bạch Hổ: Tư liệu, sự kiện và bài học kinh nghiệm*. Vietsovpetro. 2018: trang 7 - 20.

7. Yunan Li, Yifu Han. *Decline curve analysis for production forecasting based on machine learning. SPE Symposium: Production Enhancement and Cost Optimisation, Kuala Lumpur, Malaysia*. 7 - 8 November, 2017.

8. A.Mirzaei Paiaman, S.Salavati. *The application of artificial neural networks for the prediction of oil production flow rate. Energy Sources, Part A: Recovery, utilization, and environmental effects*. 2012; 34(19): p. 1834 - 1843.

Table 3. *Lan Do GWC estimation*

Depth (m TVDss)	Bottom hole depth (m TVDss)		Water-gas contact (m TVDss)		
	LT-1P	LT-2P	Initial	February 2018	At G_p = 0.39 trillion ft^3
1,080					
1,090					
1,100	1,101				
1,110					
1,120					1,124
1,130		1,138			
1,140					
1,150				1,153	
1,160					
1,170			1,170		
1,180					

Table 4. *Prediction of water influx into the Lan Do wells*

Well	Depth (m TVDss)	BRV at well depth (million m^3)	Water influx volume at the time of flooding		Cumulative production (trillion ft^3)	Time of starting to flood
			million m^3	million barrels		
1P	1,101	140	132	832	0.62	N/A
2P	1,138	370	78	490	0.38	August 2020

LD-2P are 1,101m TVDss and 1,138m TVDss respectively and GWC in February 2018 is 1,153m TVDss, the water will be present in well 2P before the end of production (Table 3).

Prediction for water influx is calculated with the assumption that the average annual production of Lan Do is 1,642 billion m^3 of gas and 0.02 million barrels of condensate to the end of field life. At the end of production, when cumulative production G_p = 0.39 trillion ft^3, well LD-2P will be flooded with water. The results of predicting time and cumulative production at the time wells start to flood are presented in Table 4.

3. Conclusion

The prediction of water influx for Lan Do field is calculated based on the reservoir and fluid parameters, the reservoir pressure prediction and calculations based on the material balance method with the average annual production of 1,642 billion m^3 of gas and 0.02 million barrels of condensate. From the results, the gas-water contact will move upwards and the earliest flooded well is LD-2P (in August 2020). Therefore, it is necessary to consider adjusting the production with reasonable rate of LD-2P to slow the water produced time and prolong the time of production of Lan Do field.

Reference

1. Viện Dầu khí Việt Nam. *Nghiên cứu phục vụ quản lý khai thác các mỏ dầu khí do nhà thầu nước ngoài điều hành tại Việt Nam*. 2017.

2. Hoang Vu Nam. *Block 06.1 PSC monthly production report*. Rosneft Vietnam. 2018.

3. Nguyen Tuan Anh. *Block 06-1 production quarterly review with PVN*. Rosneft Vietnam. 2018.

4. B.C.Craft, M.F.Hawkins. *Applied petroleum reservoir engineering (2nd edition)*. 1991.

5. BP Oil Company. *Block 06-1 Lan Tay - Lan Do development plan*. 2008.