
Petroleum Science and Technology

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lpet20

# Artificial Neural Network Modeling for the Prediction of Oil Production

S. Elmabrouk [a b] , E. Shirif [a] & R. Mayorga [a]

[a] Petroleum System Engineering Department , University of Regina , Regina , Saskatchewan , Canada

[b] Petroleum Engineering Department , University of Tripoli , Tripoli , Libya

Published online: 14 Mar 2014.

*Petroleum Science and Technology*, 32:1123–1130, 2014

ISSN: 1091-6466 print / 1532-2459 online
DOI: 10.1080/10916466.2011.605093

# Artificial Neural Network Modeling for the Prediction of Oil Production

S. Elmabrouk,[1,2] E. Shirif,[1] and R. Mayorga[1]
[1]*Petroleum System Engineering Department, University of Regina, Regina, Saskatchewan, Canada*
[2]*Petroleum Engineering Department, University of Tripoli, Tripoli, Libya*



Numerical simulations and decline curve analysis are classic tools used for predicting reservoir performance. Numerical simulations are a very complex tool, offering a nonunique solution with a high degree of uncertainty. Decline curve analysis does not take into account opening or closing intervals and variable injection rates. In this study, the authors designed a feedforward backpropagation neural network model as an alternative technique for predicting oil reservoir production performance. Real historical production data obtained from a Libyan oil field was used to train the network. This training network can serve as a practical reservoir production management tool.

*Keywords*: artificial neural network, oil prediction, reservoir management, reservoir performance production

## INTRODUCTION

The prediction of oil reservoir production performance has been an on-going challenge for engineers. It is an essential component to petroleum reservoir management. Traditionally, numerical simulations and decline curve analysis have been used to predict a reservoir's performance based on its current and past performance.

A reservoir simulation can predict hydrocarbon reservoir performance under various operating strategies. A set of reservoir parameter values are placed in the simulator to obtain a set of fluid flows. These are listed as a time-series over a specified period of time. This time-series is compared with historical data to evaluate the differences. If they are not in agreement, the input parameters are modified and the simulation is repeated until satisfactory agreement (history matching) is reached between the simulated flow outputs and historical production data. This history matching is very time consuming and offers a nonunique solution with a high degree of uncertainty.

Decline curve analysis fits the observed production rates of individual wells, groups of wells, or a whole reservoir, using mathematical functions to predict future production by extrapolating the declining function. The basis for decline curve analysis is to match past production performance with a model assuming production continues to follow the historical trend. Most decline curve analysis techniques are based on the empirical Arps exponential, hyperbolic, and harmonic equations (Arps, 1954). In some cases, production decline data do not follow a particular model but crossover the

Address correspondence to Saber Elmabrouk, Petroleum Engineering Department, University of Tripoli, Tripoli, Libya. E-mail: saber_elmabrouk@yahoo.com

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/lpet.

entire set of curves (Camacho and Raghavan, 1989). In addition, decline curve analysis does not take into account opening or closing intervals and variable water or gas injection rates.

In recent years, there has been a steady increase in the application of artificial neural networks (ANNs) in engineering. ANNs have been used to address some fundamental and specific problems that conventional computing problems have been unable to solve, especially when the data for engineering design, interpretations, and calculations are less than adequate. Lately, neural network was found to be a very suitable technology for solving problems in the petroleum and gas engineering. Most recently, Elshafei et al. (2011) presented two neural networks to predict the performance of the multi-stage gas/oil separation plants in crude oil production. The neural networks accept the initial and final pressures and temperatures of each stage and then try to utilize the oil composition information to predict the stage gas/oil ratio. Recently, Elmabrouk et al. (2010) established a neural network model that can map certain relationship that control previous production performance of the reservoir to predict current average reservoir pressure without shutting in the wells. On the other hand, the current literature, with respect to the prediction of oil reservoir production performance using artificial intelligence techniques, is sparse because it is a new area of research. In 2001, Zhong et al. developed a backpropagation neural network model to predict the production performance of oil wells with respect to spatial variation and time series. The network contains one hidden layer with 10 neurons, an input layer containing 11 neurons and one bias neuron. The 11 inputs include the wells number, the X and Y coordinates of the wells, the cumulative production at time $t$–2, the cumulative production at time $t$–1, the cumulative production at time $t$, the derivative of cumulative production at time $t$, the shut-in switch, the average distance to surrounding wells, average cumulative production of surrounding wells at time $t$, and the cumulative production days at time $t$. The output layer contains one neuron representing the cumulative production at time $t+1$. By comparing two cases, the results indicate an adequate capacity for short term predictions.

Nguyen and Chan (2005) used Decision Support System (DSS) to model production by incorporating curve fitting and ANN. DSS is a computer-based system designed to model data and qualify decisions. According to the results, Nguyen and Chan (2005) concluded no one model provides a consistently reliable prediction. This is due to a weakness in the DSS and assumptions based on expert opinion. A major weakness of the system is an assumption that the production rate is proportional to the initial production.

Da Silva et al. (2007) used four data-mining technologies (Specific Knowledge, General Knowledge, Optimal Search, and Gols) to build a time series model for predicting oil production in petroleum reservoirs. They studied several models to forecast production curves by applying historical data from Bonsucesso and Carmopolis Brazilian oil fields. The study shows the Optimal Search approach worked very well and the Gols approach gave reasonable results, while Specific Knowledge and General Knowledge gave poor results.

The objective of this study was to design a feedforward backpropagation neural network model as an alternative technique to predict oil reservoir production performance taking into account a number of production and injection wells in service and variable water injection rates. Historical production data from a Libyan oil field located in the Sirte Basin was used to train the network. The training network can serve as a practical and robust reservoir production management tool.

## NEURAL NETWORKS

The most popular ANN model is the multilayer perceptron (MLP) architecture trained using the feedforward backpropagation algorithm. The MLP architecture is composed of at least three layers of processing units interconnected via weighted connections. The first layer consists of the input 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 and the desired output. During the second phase, the computed output error propagates back through the network (error backward pass). During the third phase, connection weights are corrected by feeding the sum of the squared errors from 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 within a predetermined tolerance of the desired output (Ripley, 1996).

The selection of an optimum architecture of a model is a difficult task requiring a procedure of trial and error (Heaton, 2008). Thus, several networks with various (a) numbers of hidden units, (b) training algorithms, and (c) activation functions are attempted and the generalization error is estimated for each. The network with the minimum estimated generalization error is chosen.



## OIL RESERVOIR PRODUCTION DATA

An oil field in the Sirte Basin of Libya has 11 oil wells and 3 water injection wells. The field has been in production since September, 1978, when it had an initial average production of 3,700 BOPD, an initial average reservoir pressure of 2,960 psig, and a bubblepoint pressure of 2,710 psig. After seven months of production, the average reservoir pressure declined to 2,665 psig. This rapid decline in reservoir pressure indicates a limited aquifer size and very weak water influx. Pressure maintenance by water injection was initiated in April, 1982, when three wells were converted from oil production wells to injection wells and placed on the water injection system. As of December, 2009, the reservoir had produced a cumulative oil production of 38.5 MMSTB, and a cumulative water production of 19.75 MMBW with an average oil production rate of 2,316 BOPD and average water cut of 38%. Figure 1 illustrates the reservoir's oil production performance.

## DATA PRE-PROCESSING

Naturally, an accurate model cannot be obtained with insufficient data. Therefore, before training the network, the oil field production data are verified and pre-processed to avoid peculiar answers

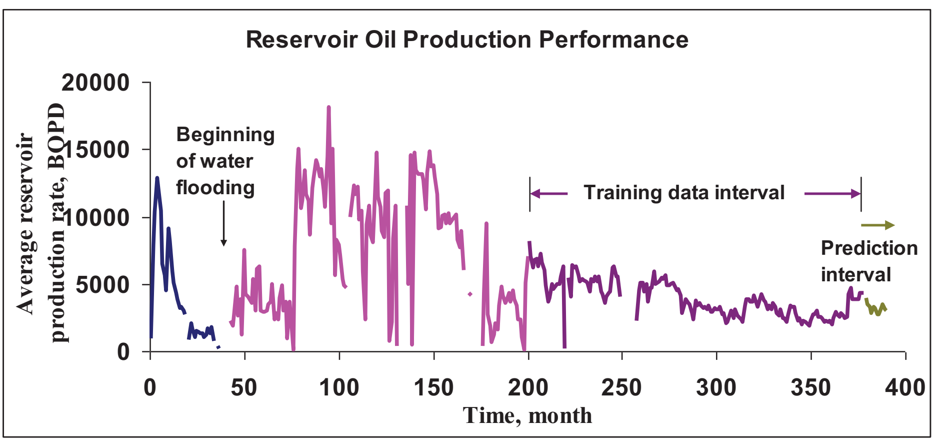


FIGURE 1 Reservoir oil production from September 1978 to July 2010.

from the trained network model. However, depending on the problem, there may be special features from the data that may be used to test its quality. One way to check the quality is to view graphical representations of the data in question, in hopes of selecting a reasonable subset while eliminating problematic portions. As presented in Figure 1, the oil field production data are time-dependent and incorporate atypical patterns. Such patterns lead to deterioration of the model's performance. Thus, primary production from its initial production in September, 1978, to March, 1982, in when the water injection began and the period from April, 1982, to January, 1995, were excluded, as illustrated in Figure 1. The remaining production data points were split into two sets before specifying the model architecture. The first set (from February 1995 to July 2009) was used to build the network model. The second set (from August 2009 to July 2010) was used for to predict the average reservoir oil production rate.

To avoid overfitting and improve generalization of the network model, the first dataset was subdivided randomly into training, validation, and testing sets. The training set (67%) is used to compute the gradient and update the network weights and biases. The validation set (16.5%) is used to check the network performance during the training process. Training can be stopped when the performance of the model on the validation dataset provides a minimum error. The testing set (16.5%) is used to fine-tune models. It is not used for training or validation, it used to indentify the optimum network architecture, to select the appropriate model and assess the performance (generalization) of a fully-specified model.



## THE NETWORK ARCHITECTURE

The key problem in our approach is determining how to select as few inputs as possible. In other words, our task is to find an effective set of inputs to capture the main features of our prediction problem. The proposed network was based on practical experience and the need to map relationships that control past oil reservoir production performance to predict production levels taking into account a number of production and injection wells in service and variable water injection rates. The chosen network has five inputs and one output. The inputs are the average reservoir oil production rate at time $t$, the average reservoir gas production rate at time $t$, the water injection rate at time $t+1$, the number of oil wells in production at time $t+1$, and the number of injection wells at time $t+1$. The output is the average reservoir oil production rate at time $t+1$. The selection of an optimum architecture of a network can be achieved using a trial-and-error approach. Accordingly, the best results were obtained by a network model consisting of two hidden layers, 13 nodes in the first layer and eight in the second layer. The nodes in the hidden and output layers are activated with a logistic function and subsequently trained by the Quasi-Newton algorithm which provides a backpropagation error with the lowest sum-squared-error.

## RESULTS AND DISCUSSION

Figure 2 shows a comparison between the network outputs of the oil reservoir rate versus the actual reservoir oil flow rate for each training data point. The average absolute error (AE) of the network output is 381.2 BOPD and the average absolute relative error (ARE) is 10.41%. The network provides results which are very close to the actual reservoir oil rates. This indicates excellent agreement between the actual and observed reservoir oil rates. The performance of the training, validation, and testing datasets are presented in Figures 3, 4, and 5, respectively, indicating the network describes the data very well. The statistical parameters used to measure the network's prediction capability are obtained from the training, validation and testing datasets and summarized in Table 1. The

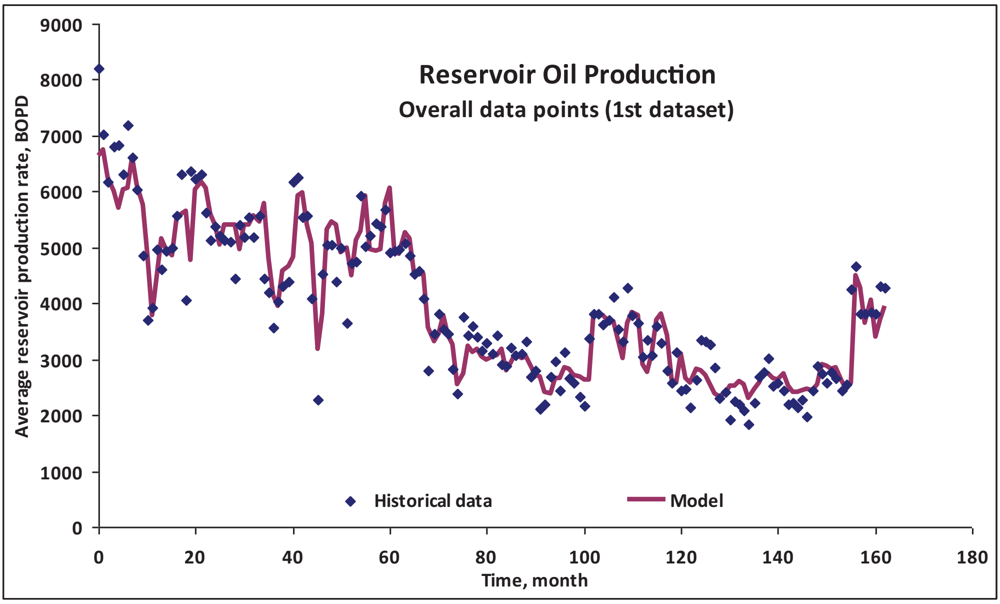


FIGURE 2 ANN model output versus the actual reservoir production (first dataset).

proposed network provides prediction values for oil reservoir rates with average AEs of 391 BOPD, 324 BOPD, and 392 BOPD for the training, validation and testing datasets, respectively. Average AREs of 10.3%, 9.2%, and 12.0% were obtained from the training, validation, and testing datasets, respectively.

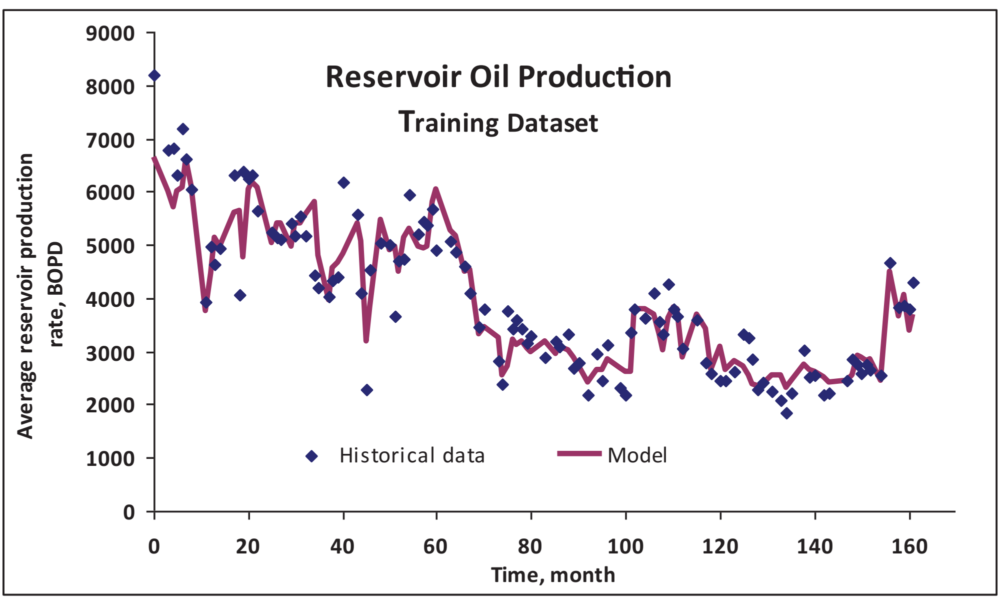


FIGURE 3 Performance of the ANN model on training dataset.

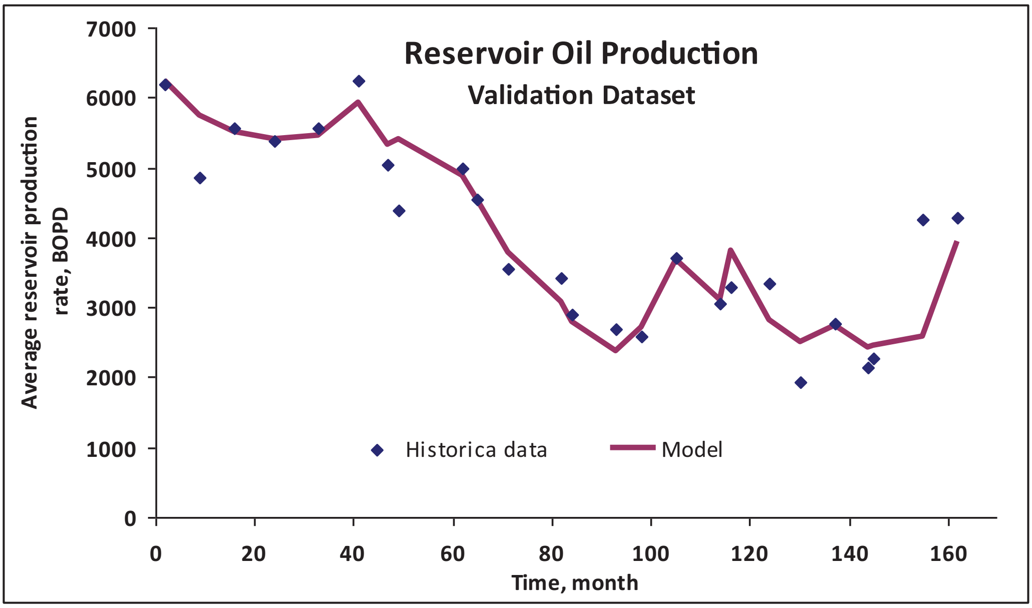


FIGURE 4 Performance of the ANN model on validation dataset.

To study the robustness and accuracy of our 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. This illustrated in Figure 6. The network provides reservoir oil rates with an average AE of 358 BOPD and average ARE of 11.97%, as illustrated in Table 1.

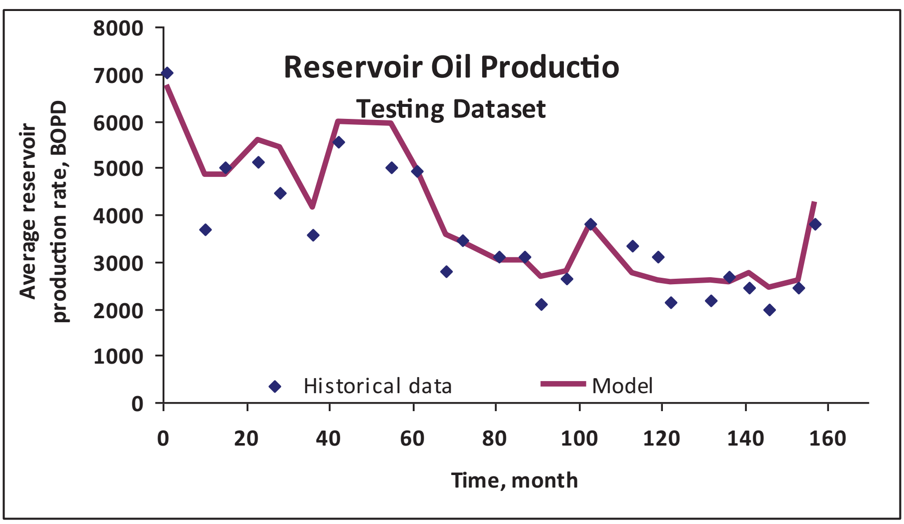


FIGURE 5 Performance of the ANN model on testing dataset.

TABLE 1
Statistical Analysis of Network Model Accuracy

| | *Target BOPD* | *Output BOPD* | *AE BOPD* | *ARE %* |
|---|---|---|---|---|
| Overall (First dataset) | | | | |
| Average | 3866 | 3909 | 381 | 10 |
| Standard deviation | 1315 | 1236 | 348 | 9 |
| Minimum | 1832 | 2310 | 1.42 | 0.02 |
| Maximum | 8195 | 6759 | 1680 | 39.6 |
| Training | | | | |
| Average | 3922 | 3920 | 391 | 10 |
| Standard deviation | 1347 | 1224 | 351 | 8 |
| Minimum | 1832 | 2310 | 1.42 | 0.02 |
| Maximum | 8195 | 6644 | 1602 | 39.7 |
| Validation | | | | |
| Average | 3926 | 3960 | 325 | 9 |
| Standard deviation | 1224 | 1260 | 374 | 10 |
| Minimum | 1931 | 2374 | 1.84 | 0.04 |
| Maximum | 6250 | 6251 | 1680 | 39.4 |
| Testing | | | | |
| Average | 3556 | 3811 | 392 | 12 |
| Standard deviation | 1210 | 1263 | 303 | 9 |
| Minimum | 1985 | 2465 | 9.6 | 0.25 |
| Maximum | 7038 | 6759 | 1163 | 31.5 |
| Prediction (Second dataset) | | | | |
| Average | 3177 | 3509 | 358 | 12 |
| Standard deviation | 356 | 256 | 233 | 8.3 |
| Minimum | 2676 | 3131 | 8.76 | 0.22 |
| Maximum | 3970 | 3961 | 666 | 24.3 |

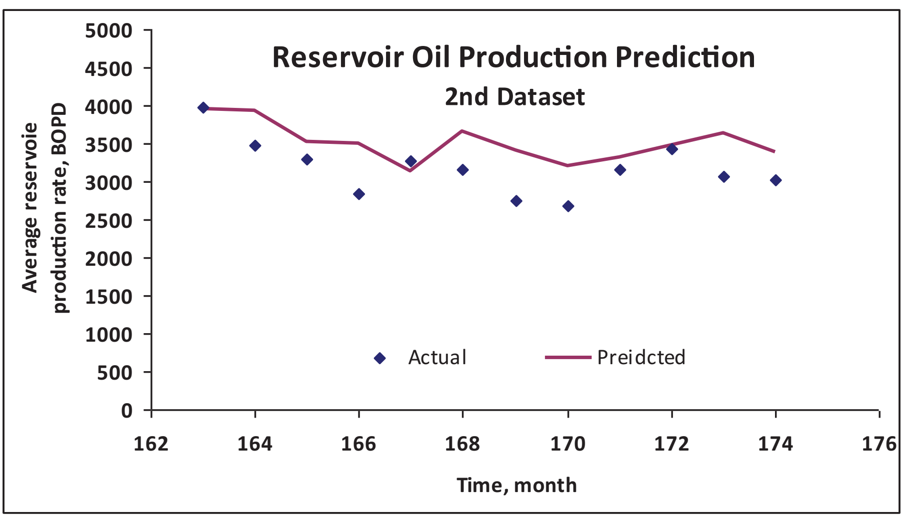


FIGURE 6 Prediction of average reservoir oil production (one year ahead from August 2009 to July 2010).

## CONCLUSION

This study provides a new, alternative method to predict oil reservoir production based on historical production data. It demonstrates the ANN technique can be used to predict oil reservoir performance, and can serve as a practical and robust tool with regard to reservoir management.

## FUNDING

The authors wholeheartedly extend their thanks and appreciation to the Faculty of Graduate Studies and Research (FGSR), University of Regina, for their generous financial support.

## REFERENCES

Arps, J. J. (1945). Analysis of decline curves. *Trans. AIME* 160:228–247.

Camacho, V., and Raghavan, R. (1989). Boundary-dominated flow in solution-gas drive reservoirs. *SPE Res. Eng.* November: 503–512.

Da Silva, L. C. F., Portella, R. C. M., and Emerick, A. A. (2007). Predictive data-mining technologies for oil-production prediction in petroleum reservoir. SPE 107371, *SPE Latin American and Caribbean Petroleum Engineering Conference*, Buenos Aires, Argentina, April 5–18.

Elmabrouk, S., Shirif, E., and Mayorga, R. (2010). A neural network approach to predict average reservoir pressure. *5th Technology of Oil and Gas Forum*, Tripoli, Libya, October 12–14.

Elshafei, M., Khoukhi, A., and Abdulraheem, A. (2011). Neural network aided design of oil production units. *IEEE Trans.* 10–13:638–641.

Heaton, J. (2008). *Introduction to neural networks for Java* (2nd ed.). St. Louis, MO: Heaton Research.

Nguyen, H. N., and Chan, C. W. (2005). Applications of data analysis techniques for oil production prediction. *J. Pet. Sci. Eng.* 37:153–169.

Ripley, B. D. (1996). *Pattern recognition and neural networks*. New York, NY: Cambridge University Press.

Zhong, H., Yang, L., Yen, J., and Wu, C. (2001). Neural network approach to predict well performance using available field data. SPE 68801, *SPE Western Regional Meeting*, Bakersfield, CA, March 26–30.