An Artificial Neural Network for Predicting Rate of Penetration in AL-Khasib Formation – Ahdeb Oil Field

Zaher Jabbar Attwan AL Zirej*, Hassan Abdul Hadi
Petroleum Department, College of Engineering, University of Baghdad, Baghdad, Iraq

Received: 17/7/2019 Accepted: 21/9/2019

Abstract
The main objective of this study is to develop a rate of penetration (ROP) model for Khasib formation in Ahdeb oil field and determine the drilling parameters controlling the prediction of ROP values by using artificial neural network (ANN).

An Interactive Petrophysical software was used to convert the raw dataset of transit time (LAS Readings) from parts of meter-to-meter reading with depth. The IBM SPSS statistics software version 22 was used to create an interconnection between the drilling variables and the rate of penetration, detection of outliers of input parameters, and regression modeling. While a JMP Version 11 software from SAS Institute Inc. was used for artificial neural modeling.

The proposed artificial neural network method depends on obtaining the input data from drilling mud logging data and wireline logging data. The data then analyzes it to create an interconnection between the drilling variables and the rate of penetration.

The proposed ANN model consists of an input layer, hidden layer and outputs layer, while it applies the tangent function (TanH) as a learning and training algorithm in the hidden layer. Finally, the predicted values of ROP are compared with the measured values. The proposed ANN model is more efficient than the multiple regression analysis in predicting ROP. The obtained coefficient of determination ($R^2$) values using the ANN technique are 0.93 and 0.91 for training and validation sets, respectively. This study presents a new model for predicting ROP values in comparison with other conventional drilling measurements.

Keywords: Artificial Neural Network, Rate of Penetration, Drilling Average, ROP Models.
Introduction

Most of the state oil companies sign drilling contracts that involve the production of very large amounts of oil and in order to minimize the total cost of well constructions and drilling, it is necessary to increase the drilling rate. Drilling rate is a key parameter in drilling optimization due to its role in the reduction of drilling operations costs [1]. Prediction of rate of penetration is very important to improve the drilling performance.

There are many conventional mathematical, direct and indirect, models to estimate and calculate the drilling rate and determine the rock mechanical properties. The drilling optimization is still a very big challenge in oil and gas industry, because of the large number of uncontrollable factors such as type of formation lithology, bottom hole temperature, formation compressive strength, and corrosive gases during the drilling of the formations [2,3]. The main aim of this study is to develop an empirical model for predicting the rate of penetration by using the artificial neural network for Khasib formation in Ahdeb oil field.

Artificial neural networks (ANNs) are information-processing procedures of large numbers of connected nodes and information in an attempt to solve the complicated non-linear relationships with high accuracy [4]. The ANN is an intelligent method which can update itself by an iteration style and depending on the provided database. One of the properties of the ANN is that there is no need for any static functions that requires a complete group of data; it is applying the correlation between the input and output information for the induction of the missing information [5].

Many experts introduced their studies on using the ANNs to predict and estimate the rate of penetration with different oil fields and cases. Bilgesu et al. [6] introduced a new approach and methodology for the prediction of ROP values at a drill site by using the drilling recorded data and the neural networks. They concluded that if the drilling rate falls below the expected values, a new bit can be selected based on the network predictions. Dashevskiy et al. [7] proposed a model using the neural networks and drilling variables (WOB, RPM) with the aim to obtain the optimum ROP values and down-hole diagnostics. Fonseca et al. [8] used the Auto-Regressive with extra Input Signals (ARX) Neural Networks model and proposed a model for ROP calculation in data from seven oil offshore field wells. With this methodology, they achieved results of high coefficient of determination ranged from 0.888 to 0.988 for the testing sets.

Akin [9] introduced a new approach for diamond bits drilling operations or hard formation through calculating the optimum rate of penetration, weight on bit, and rotary per minute by using the ANNs.

Moran et al. [5] provided a programmed ANN model for sophisticated ROP estimation and total drilling time when the well planning needs to change in wellbore size, formation drilled and total depth. The
model had flexibility of the computer software to analyze more information for the estimation and prediction of ROP based on the past experience and drilling information from offset wells.

Elkatany, [10] proposed a neural network model for the prediction of the rate of penetration with high accuracy using Self-Adaptive Differential Evolution Artificial Neural Network (SaDE-ANN). The obtained results showed that the ROP has an intense relation with the other drilling variables (WOB, RPM and Horse Power) and a reasonable relation with the unconfined compressive strength.

Ahmed et al. [11] proposed a method for predicting the rate of penetration in shale formation using fuzzy logic system based on five drilling parameters (WOB, RPM, ROP, Torque and flow rate) and five drilling fluid properties (mud weight, plastic viscosity, marsh funnel viscosity, yield point and solid content %). They proved that the fuzzy logic technique can be used effectively to predict the ROP with high performance, while the results showed a coefficient of correlation of \( R = 0.97 \) and an average absolute percentage error of AAPE = 7.3.

Li et al. [12] proposed a new method for the prediction of ROP ahead of the bit through real-time updated machine learning models.

Yuswandari et al. [13] applied an ANN model using data from a geothermal field in Indonesia to predict the rate of penetration. They used a multiple regression for each parameter in the data set by normalizing the importance technique to select the input variables, which have high impact on rate of penetration. They concluded that the final model can provide somewhat a picture of rate of penetration in nearby wells and can be improved by using more data from another well in the training set.

In this study, the proposed ANN predictive ROP model is an empirical model with high accuracy and high performance, which is provided with the weights between the input and hidden layers and hidden and output layers as well as the biases for hidden and output layers.

An artificial neuron is a simple element of a neural network, which consist of major components including, input data, weights, an activation function and output values. Each input parameter is multiplied by adjustable weights as shown in Eq. 1. Then the adjusted inputs are summed in a field called the local receptive field which enters through an activation function that executes a non-linear process on the information output and transmits it into the predicted output. Most of the activation functions are non-linear [14].

\[
y_i = \sum_{i=1}^{n} W_i X_i
\]

where yi: The summation of multiplying weights by the neuron values in the previous layer, Wi: weight value, Xi: input value.

There is a hidden layer(s) between the input and output layer(s) [15]. In the hidden layer, the signals received from the input neurons are processed and then transformed to the output layer. In addition, there is a bias neuron in the hidden layer which is connected to all neurons in the next layer but none in the previous layer [16]. Biases will be summed with the weighted inputs and the result is the net input, so that Eq. (1) becomes:

\[
y_i = \sum_{i=1}^{n} W_i X_i + b_i
\]

The net input (yi) will pass through an activation or transfer function to generate the neuron output[17].

1.1 Artificial Neural Network Architecture

Multi-Layer percpetron (MLP) is one of the best ANN structures which is widely used because of its ability of modeling a complex relationship between variables, which gives results with a high accuracy [18]. Multi-layer perceptron has an input layer, one or more hidden layers and an output layer [19]. The input layer only distributes the input elements. The hidden layer is between the input and output layers and the hidden neurons make the weighted sum by using the activation function. The output layer as well as the hidden layer make the weighted sum by applying the activation function.

1.2 Artificial Neural Network Learning Algorithm – Back Propagation

Learning algorithm is used to minimize the total error by updating the weights. Back-propagation algorithm, a supervised neural network, is a type of networks designed to solve the problems of classification through the multi-layer neural networks instead of the perceptron networks which deal with single layer neural networks.
This type is widely used in the pattern recognition and consists of a multi-layer neural network, which can be a training in a prompt and repetitive form until reaching the optimum level of the network performance according to the variables of weights and biases; this type was developed by Paul Werbos in 1974 and discovered by Rumelhart and Parker. To test whether an improvement is achieved through the performance of the network, two error-based metrics including the coefficient of determination ($R^2$) and the root mean square error (RMSE) are founded for the proposed model. This algorithm works by propagating the input values forward across the network, then propagating them from the last layer to the first layer and adjusting the connection weights and biases so that the overall error between the predicted and target outputs is minimized, ideally reaching zero [20].

**Data Analysis and Methodology**

**Drilling Mud Logging and Selection of Wire Line Logging Data Variable**

Khasib formation is one of the main pay zones in Ahdeb oil field that is located between Nomania town and Kut town of Wasit Province, 140 km southeast of Baghdad, in the fluvial plain between the Euphrates and Tigris rivers. The upper part is mainly consisted of light grey to grey limestone and the lower part consists mainly of light brown limestone [21]. It is important to determine the most significant variables that will affect the output function. Accordingly, different input parameters were selected for optimizing the rate of penetration.

Using multi-layer perceptron (input, hidden and output) layers and ANN analyses for ROP prediction, the network was trained by the back propagation method. The tangent function (Tanh) was used as an activation function and the data were randomly divided as 70% for training the network and 30% for the validation (Hold back Validation).

After analyzing the neural network, the best determination coefficient ($R^2$) value was 0.93 for the training set and 0.91 for the validation set. The results for each group of input parameters are shown in Table-1, which were obtained by establishing more than ninety attempts of training the neural networks with different selections of input parameters.

**Neural Network Analysis of Raw Data**

A neural network analysis was applied for the Khasib formation dataset. The raw dataset was obtained from one well in Ahdeb oil field (AD1-5-1H) for Khasib formation (raw dataset). The raw dataset consisted of independent variables (depth, WOB, RPM, TORQUE, pump pressure, wave travel time) and one dependent variable (ROP)[22].

The neural network consists of multilayer perceptron (input layer, hidden layer, and output layer). The input layer contained independent variables (depth, WOB, RPM, Torque, flow rate, pump pressure, and wave travel time), whereas the output layer contained the dependent variable ROP. The hidden layer consisted of one layer that contained seven neurons and applied TanH as an activation function. The training set is the section that rates the ROP model parameters, while the validation set is the section that validates the predictive strength of the ROP model. The validation method is the Hold Back method, which randomly divided the raw data into the training and validation sets, where the used validation proportion was 0.3333 of the original data.

**Table 1- Prediction of the Rate of Penetration with Different Selections of input Variables.**

| Input Variable Selection | Determination Coefficient, ($R^2$) | Training | Validation |
|--------------------------|----------------------------------|----------|------------|
| DEPTH, WOB, RPM, Q, TRQ, SPM, WTT | 0.9337634 | 0.9055535 |
| DEPTH, WOB, RPM, Q, TRQ, SPM | 0.8757587 | 0.8288593 |
| DEPTH, WOB, RPM, Q, TRQ, WTT | 0.917128 | 0.832934 |
| DEPTH, WOB, RPM, Q, SPM, WTT | 0.8251977 | 0.6983541 |
| DEPTH, WOB, RPM, TRQ, SPM, WTT | 0.9230936 | 0.895346 |
| DEPTH, WOB, Q, TRQ, SPM, WTT | 0.9110697 | 0.9032697 |
| WOB, RPM, Q, TRQ, SPM, WTT | 0.8935787 | 0.8609942 |
| WOB, RPM | 0.6001305 | 0.5062489 |
| WOB, TRQ | 0.7332279 | 0.6978608 |
| WOB, Q | 0.3962398 | 0.2500142 |
| WOB, SPM | 0.3443269 | 0.3285626 |
Actual versus Predicted Plot of Raw Data

Figures (1, 2) demonstrate the cross plots of actual versus predicted ROP that were obtained from the neural network analysis for the training and validation sets, respectively.

|        | Actual  | Predicted |
|--------|---------|-----------|
| WOB, WTT | 0.6058865 | 0.59355   |
| Q      | 0.0397367 | 0.0390452 |
| RPM    | 0.597532  | 0.4070751 |
| TRQ    | 0.7051982 | 0.6949183 |
| WOB    | 0.3425569 | 0.2948616 |
| SPM    | 0.0211111 | 0.0052643 |
| WTT    | 0.5165625 | 0.2989639 |

**Figure 1**- ROP measured Vs. ROP Predicted (Training Set).

**Figure 2**- ROP measured Vs. ROP Predicted (Validation Set).
ROP Model Performance Analysis of Raw Data

The results of ANN analysis of the data showed R² values of 0.8013983 and 0.5578696 for the training and the validation sets, respectively. In addition, the values of RMSE were 1.5842187 and 2.8684367 for the training and validation sets, respectively.

The results from training and validation sets showed that the values of R² are low and need to be improved by excluding the outliers. We used the box plot method to determine the outliers and noise in the data set and raw data of Khasib formation.

Table 2 shows the number of outliers for each variable according to the box plot method.

Table 2- Drilling Parameters Outliers for Khasib Formation

| Variable      | Outliers Number |
|---------------|-----------------|
| Depth         | 0               |
| ROP           | 27              |
| WOB           | 26              |
| RPM           | 15              |
| TORQUE        | 4               |
| Pump Pressure | 16              |
| Flow Rate     | 20              |
| Travel Time   | 19              |

Multiple Regression Analysis for Data without Outliers

Figure 5 shows the actual ROP vs. predicted ROP by using the multiple regression analysis of Khasib Formation data after excluding the outliers of the input variables.

The analysis resulted in an R² value of 0.85 and an RMSE value of 0.6037.

Eq. (3) shows the ROP model for Khasib Formation data without outliers by using the Multiple Regression Analysis:

\[
ROP = 11.3856720311665 + 0.0032497969915059 \times Depth + \\
- 0.0151962837563356 \times WOB + \\
0.00853766181451426 \times RPM + \\
- 0.767212724589397 \times Torque + \\
0.0424244667317726 \times SPM + \\
- 0.00397946264878767 \times Flow\ rate + \\
- 0.109527264225027 \times Wave\ Travel\ Time \quad (3)
\]
Neural Network Analyzing for Data without Outliers

From the results of box plots and the outliers, both independent and dependent showed many outliers values, which resulted in a negative effect on the results of $R^2$ R square and RMSE and thus on the prediction capacity of the developed ROP model of Khasib formation. Because of the exclusion of all the outlier values from the raw dataset, the ANN technique was applied again on the same data to obtain the best predictive model of the rate of penetration.

Figure 4 - The architecture of ANN Constructed for ROP prediction without outliers.

Actual versus Predicted ROP Plot

Figures-(5, 6) show the actual vs. predicted ROP in the training and validation sets for the results obtained from the ANN analysis of Khasib formation dataset after excluding the outliers.

Figure 5 - ROP Measured vs. ROP Predicted (Training Set)
2.4.2 ROP Model Performance Analysis

The results of ANN analysis of the data after excluding the outliers demonstrated R² values of 0.93 and 0.91 for the training and validation sets, respectively. The results also showed RMSE values of 0.4872786 and 0.469345 for the training and validation sets, respectively.

From the results of the training and validation sets which were obtained after excluding the outliers and noise from the raw data, we achieved a good improvement for the R square and RMSE values and, thus, a good ROP predictive values for Khasib formation.

3. Khasib Formation Predictive ANN Model

The TanH(x) function is:

\[ f(x) = \frac{2}{1+e^{-x}} - 1 \]  \hspace{1cm} (4)

All the input variables will be normalized in the range of -1, 1 before substituting in the Eq. (6), since we use TanH as an activation function.

The normalization equation is as follows:

\[ X_n = 2 * \left( \frac{X - X_{min}}{X_{max} - X_{min}} \right) - 1 \]  \hspace{1cm} (5)

where: \( X_n \): normalized input parameter, \( X \): actual input parameter, \( X_{max}, X_{min} \): maximum, minimum limits of the input parameters.

\[ \text{(ROP)}_n = \sum_{l=1}^{n} W_{2l} * \left( \frac{2}{1+e^{-2(W_{1l1} \cdot \text{depth} + W_{1l2} \cdot \text{WOB} + W_{1l3} \cdot \text{RPM} + W_{1l4} \cdot TRQ + W_{1l5} \cdot Q + W_{1l6} \cdot SPM + W_{1l7} \cdot WFT + b_{1l})}} - 1 \right) + b_2 \]  \hspace{1cm} (6)

where \( \text{(ROP)} \): normalized rate of penetration, \( W_{1l} \): input hidden weights, \( W_{2l} \): hidden -output weights, \( b_{1l} \): bias-hidden, \( b_2 \): bias- output.

Now we make the de-normalization by using Eq. (5) for the values of \( \text{(ROP)}_{\text{normalized}} \) obtained from Eq. (6) to get the real values of ROP.

Eq. (6) shows the ROP model for Khasib Formation by using the artificial neural network. Table-3 shows the weights and biases for Khasib formation ANN-ROP predictive model, which was estimated following the artificial neural analysis of the data after excluding the outliers. The weights are between the input and hidden layers and between the hidden and the output layers. The biases are for the hidden and output layers.
Table 3- Weights and Biases of Khasib Formation ANN Model

| Hidden Node (i) | Input-Hidden Weight (W₁) | Hidden-Outputs Weight | Bias |
|-----------------|---------------------------|------------------------|------|
|                 | Depth | WOB  | RPM  | TRQ  | SPM  | Q    | WTT  | W₂₁ | Hidden | Output |
| 1               | -     | 0.009283 | 0.017861 | 1.622583 | 0.188827 | -     | 0.016608 | 0.020854 | 1.734180 | 37.01366 | 8.510084 |
| 2               | -     | 0.0002761 | 0.004058 | 0.165878 | 0.609676 | 0.039113 | -     | 0.027996 | 0.027245 | 5.227166 | 8.879242 |
| 3               | -     | 0.001428 | 0.011533 | 0.130465 | 0.800555 | 0.115457 | -     | 0.024081 | 0.024081 | 4.063256 | 132.8024 |
| 4               | -     | 0.003097 | 0.0043698 | 0.148919 | 0.080069 | 0.071155 | -     | 0.098288 | 0.098288 | 4.563577 | 37.78632 |
| 5               | -     | 0.007439 | 0.025550 | 0.048242 | 0.1252109 | 0.056690 | -     | 0.063428 | 0.063428 | 5.648295 | 22.69718 |
| 6               | -     | 0.011392 | 0.030842 | 0.519942 | 0.724565 | 0.070455 | -     | 0.080672 | 0.080672 | 0.315732 | 246.4887 |

4. Mathematical Validation of the ANN - ROP Predictive Model Equation for Khasib Formation

To validate the developed ANN – ROP predictive model, the values of ROP were calculated based on the artificial neural network predictive model for Khasib formation datasets by using Eq. (6).

Figure -7 shows the actual and predicted ROP values obtained from the ANN model along a depth of 2626 - 3585 meter. The proposed predictive model achieved a high accuracy between the actual ROP and the calculated ROP.

The predicted ROP showed a symmetric distribution with depth and followed the trend of the actual ROP, except for the data point at the depth of 2644 m because of using a high WOB while drilling. We can confirm that the drilling and wireline parameters used in the model were physically appropriate and that the ANN predictive ROP model can represent the relationships between the parameters involved in ROP modelling.
Figure-7 Actual ROP and ANN-ROP Values along Depth
Conclusions
The proposed ANN model for Khasib formation-Ahdeb oil field, based on the high performance of the coefficient of determination (R²), gives a good prediction capacity of ROP values in comparison with the actual values. The multiple regression analysis gives good results for predicting ROP when it is compared with the measured ROP, but with lower efficiency than the predicted values of the ANN model.
1. Acquisition and analysis of inputs data is one of the most important steps in ANN analysis method for predicting the ROP.
2. The selected number of neurons or nodes in the hidden layer is an important step in developing ROP-predictive models by using ANN. If the number of neurons in the hidden layer is high, the ANN network tends to memorize the data and if the number of hidden neurons is small, the ANN predictive ROP is poor. The increasing of the neurons in the hidden layer does not necessarily improve the predicting of ROP and the best method to determine the number of neurons in the hidden layer for Khasib formation is by trial and error.
3. The validation of the proposed ANN model equation by using the Microsoft Excel program gives a good matching between the actual and predicted ROP values.

Acknowledgements
The authors would like to introduce special thanks, gratitude, and appreciation to Dr. Farqad Hadi for his tremendous academic support and sharing his experiences about the artificial neural network modeling.

References
1. K. H. Mnati, K.H. 2015. “Rate of Penetration Model With Aid Mud Logging and Wire Logging Data FOR Iraqi Oil Field.
2. Bataee, M. Mohseni, S. Engineering, P. and Faculty, A. 2011. “SPE 140029 Application of Artificial Intelligent Systems in ROP Optimization: a Case Study in Shadegan Oil Field. According to the field data, there are several methods to reduce the drilling cost of other wells. One Considering the geology and rock mec.”.
3. Shi, X., Liu, G., Gong, X., Zhang, J., Wang, J. and Zhang, H. 2016. “An Efficient Approach for Real-Time Prediction of Rate of Penetration in Offshore Drilling.” vol. 2016.
4. Ali, B.G.P.J. K. 1994. “Neural Networks: A New Tool for the Petroleum Industry? J.K. Ali, British Gas PLC COW19M.”.
5. Moran, D., Ibrahim, H., Purwanto, A., International, S., Osmond, J. and Corporation, H. 2010. “IADC / SPE 132010 Sophisticated ROP Prediction Technologies Based on Neural Network Delivers Accurate Drill Time Results,” pp. 1–9.
6. Bilgesu, A.H.I. and Altims, S. 1997. “A New Approach for the Prediction of Rate of Penetration (ROP) Values.pdf.”.
7. Dashevskiy, D., Dubinsky, V., Macpherson, J.D. and Inteq, B.H. 1999. “SPE 56442 Application of Neural Networks for Predictive Control in Drilling Dynamics.”.
8. Fonseca, T.C., Mendes, J.R.P., Serapião, A.B.S. and Guilherme, I.R. 2006. “Application of ARX neural networks to model the rate of penetration of petroleum wells drilling Application of Arx Neural Networks to Model The Rate Of Penetration of Petroleum Wells Drilling,” no. November 2015.
9. Akin, S. and Karpuz, C. 2018. “Estimating Drilling Parameters for Diamond Bit Drilling Operations Using Estimating Drilling Parameters for Diamond Bit Drilling,” vol. 3641, no. May 2014, 2008.
10. Elkatatny, S. 2018. “Rate of Penetration Prediction Using Self-Adaptive Differential Evolution-Artificial Neural Network,” pp. 1–12, 2018.
11. Abdulmalek, S., Elkatatny, S., Ali, A.Z., Mahmoud, M. and Abdulraheem, A. 2019. “Rate of penetration prediction in shale formation using fuzzy logic,” Int. Pet. Technol. Conf. 2019, IPTC.
12. Li, Y. and Samuel, R. 2019. “Prediction of penetration rate ahead of the bit through real-time updated machine learning models,” SPE/IADC Drill. Conf. Proc., vol. 2019-March.
13. Yuswandari, A., Prayoga, A. and Purba, D. 2019. “Rate of Penetration (ROP) Prediction Using Artificial Neural Network to Predict ROP for Nearby Well in a Geothermal Field,” *Proc. 44th Work. Geotherm. Reserv. Eng. Stanford Univ. Stanford, California, Febr. 11-13, 2019*, pp. 1–5.

14. Hagan, B.M.P., MT. and Demuth, M.T. 1996. *Publishing Company, USA, Neural network design.*

15. Al-rashidi, A.A. 2011. “Application of Neural Networks to Evaluate Factors Affecting Drilling Performance Abdulrahman AL-Basman.”

16. Asadi, A. 2017. “Application of Artificial Neural Networks in Prediction of Uniaxial Compressive Strength of Rocks Using Well Logs and Drilling Data,” *Procedia Eng.*, 191: 279–286.

17. Lv, C. 2017. “Levenberg-marquardt backpropagation training of multilayer neural networks for state estimation of a safety-critical cyber-physical system,” *IEEE Trans. Ind. Informatics*, 14(8): 3436–3446.

18. Sorouch, H., Oii, W., Me, T. and Tokhmechi, B. 2010. “SPE 130047 A Neural Network Approach for Real Time Evaluation of Sandstone Strength,” 4543(June): 14–17, 2010.

19. Esmaeili, A., Elahifar, B., Fruhwirth, R.K., Thonhauser, T.D.E. and Engineering, D. 2012. “SPE 163330 ROP Modeling Using Neural Network and Drill String Vibration Data.”

20. Sapna, S. 2012. “Backpropagation Learning A Lgorithm Based On Levenberg Marquardt Algorithm,” pp. 393–398.

21. Midland Oil Company, “Final Well Report,” Baghdad, Iraq, 2010.

22. Midland Oil Company, “Mud Logging Report,” Baghdad, Iraq, 2010.