

Development of drilling trip time model for southern Iranian oil fields: using artificial neural networks and multiple linear regression approaches

Omid Safaee Ardekani · Seyed Reza Shadizadeh

Received: 3 February 2011 / Accepted: 16 June 2013 / Published online: 28 June 2013
© The Author(s) 2013. This article is published with open access at Springerlink.com

Abstract One of the main goals of drilling venture is the minimum drilling cost. The minimum cost for every drilling interval depended upon the trip time. As general rule for any area where the trip time is not given, 1 h for 1,000 ft is used in calculation of cost per foot of drilling. In this study, an attempt is made to develop a model for estimating drilling trip time in the southern Iranian oil fields. For this purpose, drilling data from the drilling daily reports of the drilled wells in three southern Iranian oil fields were gathered. In this work, both an artificial neural networks (ANN) model and a multiple linear regression model have been developed for estimating drilling trip time. The results indicate that the ANN model predicts trip time more accurately than the multiple linear regression model. However, the multiple linear regression model is more usable.

Keywords Trip time · Drilling cost · Multiple regression analysis · Artificial neural networks · Iranian oil fields

List of symbols

C_{bi}	Bit cost, \$
C_{di}, C_r	Drilling cost and rig cost, \$/ft
ΔD_i	Formation interval drilled, ft
D_{in}	Starting bit depth, m
T_{ci}, T_{ti}	Connection time and trip time, h
T_{di}	Drilling time, h

Introduction

Making a trip refers to the process of removing the drill string from the hole to change a portion of the downhole assembly and then lowering the drill string back to the hole bottom. A trip is made usually to change a dull bit (Bourgoyne et al. 1991). The estimation of drilling trip time has a great significance in drilling engineering. In the following, some of its applications are mentioned.

Bit selection

Although there is no exact scientific theoretical approach for proper selection of drill bits, the following useful methods may provide a close estimate to the best bit for the given formation interval to be drilled: thorough evaluation and comparison of offset, well bit records, bit run cost equation, drill-off test and specific energy equation. The bit run cost equation is generally what provides the drilling engineer with a quick estimate of the offset bit run cost and, thus, the ability to compare bits. Bit run cost may be expressed as follows (Azar and Samuel 2007):

$$C_{di} = \frac{C_{bi} + C_r(T_{di} + T_{ti} + T_{ci})}{\Delta D_i} \quad (1)$$

where C_{di} is the drilling cost in \$/ft for bit run i , C_{bi} is the cost of bit run i in \$, C_r is rig cost in \$/h, T_{di} is the drilling time in h for bit i , T_{ti} is the trip time in h for bit i , T_{ci} is the connection time in h for bit i , and ΔD_i is the formation interval drilled in ft by bit i .

Drilling optimization

The study of cost per foot is useful in defining optimum, minimum cost drilling condition. A cost comparison of

O. S. Ardekani · S. R. Shadizadeh (✉)
Abadan Faculty of Petroleum Engineering, Petroleum University
of Technology, Northern Bowarde, 6318714331 Abadan, Iran
e-mail: shadizadeh@put.ac.ir

each bit run on all available wells in the area will identify the bits and operation conditions that yield minimum drilling costs. The drilling engineer provides his expected rig costs, bit costs, and assumed average trip time. Then, he can use the bit run cost equation (Adams and Charrier 1985).

Well planning and cost estimate

Preparing cost estimates for a well is the final step in well planning. The time required to drill the well has a significant impact on many items in the well cost. The cost of the footage drilled during a single bit run is the sum of three costs: bit costs, trip costs, and rig operation cost. The cost of the bit and the cost to trip are fixed for a particular bit run (Adams and Charrier 1985).

Drilling trip time depends on factors such as: well depth, hole size, surge and swab pressure, bottom hole assembly configuration, hoisting capacity, use of automatic pipe handling system, type of rig, hole problems, crew efficiency, and drilling regulations.

Trip time prediction models

Available rule of thumb for trip time estimation is 1 h/1,000 ft of well depth. Over the total drilling life of a well, this rule of thumb will be reasonably accurate (Adams and Charrier 1985). Short (1982) has used the following estimation for trip time; Trip time is taken as 0.8 h/1,000 ft to 10,000 ft; and 1.0 h/1,000ft from 10,000 to 15,000 ft; 1.2 h/1,000ft, from 15,000 to 20,000 ft. Adams and Charrier (1985) used Table 1 for estimation of trip time in well planning. This table was developed by several operators who have conducted field studies. Schofield et al. (1992) used the following relationships for trip time.

Table 1 Average trip times (Adams and Charrier 1985)

Depth, ft	Hole (Bit) size, in.		
	Small (<8.75)	Medium (8.75–9.875)	Large (>9.875)
2,000	1.5	3.0	4.5
4,000	2.5	4.2	5.75
6,000	3.5	5.4	7.0
8,000	4.7	6.5	8.0
10,000	5.8	7.25	9.0
12,000	7.0	8.25	10.25
14,000	8.25	9.25	11.50
16,000	9.75	10.25	12.50
18,000	11.00	11.25	13.75
20,000	11.8	12.25	15.0

The average circulating time prior to tripping out of hole was 1 ¼ h. The following relationships have been obtained by fitting the best line in a graph of trip time versus depth.

$$\text{Roundtrip (h)} = \left(2 + \frac{D(m)}{1,250}\right) * 2 \quad (2)$$

D is the depth in meters. In the case that there was a downhole motor:

$$\text{Roundtrip (h)} = \left(2 + 5 + \left(2 * \frac{D(m)}{1,250}\right)\right) \quad (3)$$

Falcao et al. (1993) used Eq. (4) for the round trip time which is obtained from field experience:

$$T(h) = \frac{3 \times D}{1,000} + 1 \quad (4)$$

where D is the bit depth, meter.

Methods

The main objective of many engineering investigations is to make predictions. Usually, such predictions require a formula to be found which relates the dependent variable to one or more independent variables. This technique in data analysis refers to multivariate statistical analysis. One of the main types of the multivariate statistical analysis is regression analysis. Another approach which can be used in this region is artificial neural networks. In this study, both a regression model and an ANN model are developed to predict drilling trip time from the predictor variables that are depth, open hole length, drill collar number, bit diameter size, mud weight, using downhole motor, and using top drive rig.

Regression analysis

Regression analysis is a statistical methodology that utilizes the relation between two or more quantitative variables so that a response or outcome variable can be predicted from the others. This methodology is widely used in business, the social and behavioral sciences, the biological sciences, and many other disciplines (Kutner et al. 2005).

Artificial neural networks

A neural network is basically a model structure and an algorithm for fitting the model to some given data. The network approach to modeling a plant uses a generic nonlinearity and allows all the parameters to be adjusted. In this way, it can deal with a wide range of nonlinearities. Learning is the procedure of training a neural

network to represent the dynamics of a plant. The neural network is placed in parallel with the plant, and the error between the output of the system and the network outputs, the prediction error, is used as the training signal. Neural networks have a potential for intelligent control systems because they can learn and adapt, they can approximate nonlinear functions, they are suited for parallel and distributed processing, and they naturally model multivariable systems. If a physical model is unavailable or too expensive to develop, a neural network model might be an alternative (Sumathi and Pa-neerselvam 2010).

Data gathering

The Dezful embayment is one of the most prolific areas in south of Iran and includes 45 oil fields, often associated with gas caps. Several of them are categorized as super giants as they contain 10–50 billions barrels of oil in place, i.e., Aghajari, Ahwaz, Bibihakimeh, Gachsaran, Mansuri, Marun, and Rag-e-Safid. This zone is characterized by intense structural depression and was formed as a result of the Late Cretaceous continental collision

between the Eurasian (central Iran) and Persian plates (Bordenave and Hegre 2005).

After reviewing drilling daily reports of the three oil fields (Ahwaz, Marun, and Gachsaran), 1,072 round trip records related to bit changes have been gathered. Each round trip record contains eight parameters, which are necessary for comparing round trips. In the following, these parameters are described in detail.

Depth

The measured depth of the well at the bit change which is in meters and labeled as *D*. The range of depth in the gathered data is between 104 and 5,268 m.

Using downhole motor

Directional drilling with a downhole motor has a significant effect on trip time. In directional drilling with a downhole motor, there is an especial bottom hole assembly, surface test, and shallow test for motor. This parameter, which is dimensionless and labeled as DM, takes two values: one or zero which indicates using or not using downhole motor in drill string.

Table 2 Correlation matrix between variables

		Trip time	<i>D</i>	OHL	DC	BS	MW	DM	TD
Trip time	Pearson correlation	1	0.775	0.136	0.470	−0.630	0.025	0.457	0.237
	Sig. (2-tailed)		0.000	0.000	0.000	0.000	0.441	0.000	0.000
	<i>N</i>	973	973	973	973	973	973	973	973
<i>D</i>	Pearson correlation	0.775	1	−0.042	0.587	−0.803	−0.048	0.119	0.158
	Sig. (2-tailed)	0.000		0.190	0.000	0.000	0.134	0.000	0.000
	<i>N</i>	973	973	973	973	973	973	973	973
OHL	Pearson correlation	0.136	−0.042	1	0.038	0.364	−0.122	0.098	0.018
	Sig. (2-tailed)	0.000	0.190		0.239	0.000	0.000	0.002	0.569
	<i>N</i>	973	973	973	973	973	973	973	973
DC	Pearson correlation	0.470	0.587	0.038	1	−0.563	−0.079	−0.116	0.124
	Sig. (2-tailed)	0.000	0.000	0.239		0.000	0.014	0.000	0.000
	<i>N</i>	973	973	973	973	973	973	973	973
BS	Pearson correlation	−0.630	−0.803	0.364	−0.563	1	0.035	−0.228	−0.102
	Sig. (2-tailed)	0.000	0.000	0.000	0.000		0.276	0.000	0.002
	<i>N</i>	973	973	973	973	973	973	973	973
MW	Pearson correlation	0.025	−0.048	−0.122	−0.079	0.035	1	−0.040	−0.005
	Sig. (2-tailed)	0.441	0.134	0.000	0.014	0.276		0.209	0.880
	<i>N</i>	973	973	973	973	973	973	973	973
DM	Pearson correlation	0.457	0.119	0.098	−0.116	−0.228	−0.040	1	0.048
	Sig. (2-tailed)	0.000	0.000	0.002	0.000	0.000	0.209		0.138
	<i>N</i>	973	973	973	973	973	973	973	973
TD	Pearson correlation	0.237	0.158	0.018	0.124	−0.102	−0.005	0.048	1
	Sig. (2-tailed)	0.000	0.000	0.569	0.000	0.002	0.880	0.138	
	<i>N</i>	973	973	973	973	973	973	973	973

Table 3 Model summary

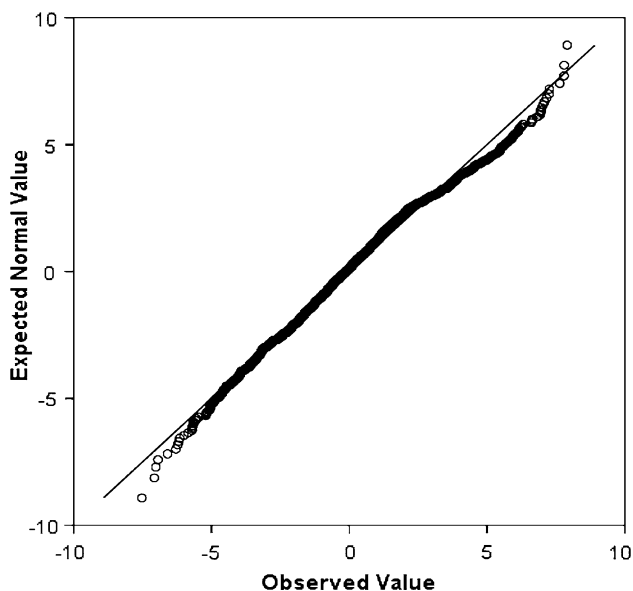
Model	<i>R</i>	<i>R</i> ²	Adjusted <i>R</i> ²	Std. error of the estimate
Equation 5	0.881	0.776	0.774	2.7629

Table 4 ANOVA table

Model		Sum of squares	df	Mean square	<i>F</i>	Sig.
Equation 5	Regression	25506.452	6	4251.075	556.872	0.000
	Residual	7374.302	966	7.634		
	Total	32880.754	972			

Table 5 Multiple regression coefficients

Model		Unstandardized coefficients		Standardized coefficients	<i>t</i>	Sig.	95 % confidence interval for <i>B</i>	
		<i>B</i>	Std. error	Beta			Lower bound	Upper bound
Equation 5	Constant	1.534	0.428		3.580	0.000	0.693	2.375
	D	0.004	0.000	0.653	33.393	0.000	0.004	0.004
	OHL	0.001	0.000	0.126	8.148	0.000	0.001	0.001
	DC	0.104	0.017	0.122	6.233	0.000	0.071	0.137
	MW	0.022	0.004	0.098	6.336	0.000	0.015	0.029
	DM	6.186	0.261	0.377	23.664	0.000	5.673	6.699
	TD	2.257	0.355	0.098	6.360	0.000	1.561	2.954

**Fig. 1** Normal probability plot

Using top drive rig

Two major kinds of rigs are top drive and rotary table rigs. This parameter is labeled as TD and takes two values: zero or one. Zero means that the rig is a rotary table, and one means that the rig is a top drive.

Drill collar number

The number of drill collars used in drill string. Running or pulling a drill collar into or out of a hole takes more time than a drill pipe; so it must be taken into account. It is dimensionless and labeled as DC.

Mud weight

When high weighted mud is used, the speed of tripping is less than usual, so the round trip time increases. Mud weight labeled as MW, is measured in pound per cubic feet (pcf) and its range in the collected data is 53–156 pcf.

Open hole length

The speed of tripping in open hole section is less than in cased hole section, so that the more open hole length the more round trip time. Open hole length is measured in meters and its range is between 0 and 3,988 m in the gathered records. This parameter is labeled as OHL.

Bit diameter size

Bit diameter size which is labeled as BS and measured in inches.

Trip time

The time interval of a round trip measured in hours. In the gathered records, trip time is between 2 and 39 h.

Models development and results

Ten percent of records have been left out as test data for validation of the models. So the models have been developed based on the 90 % remaining records. Seven predictor parameters have been used in regression analysis and ANN to predict trip time. These parameters are *D* (m), OHL (m), DC, BS (inch), MW (pcf), TD, and DM.

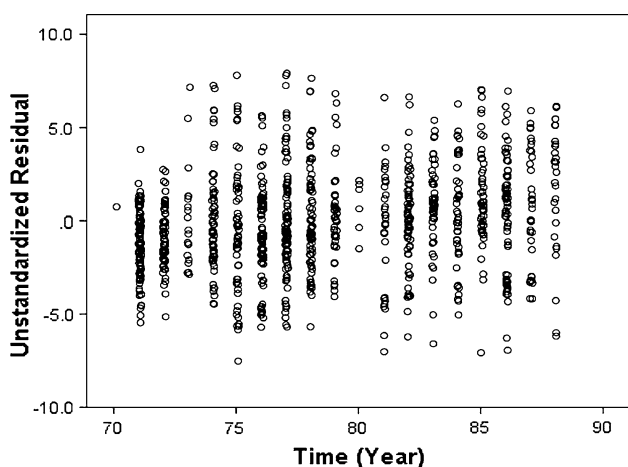
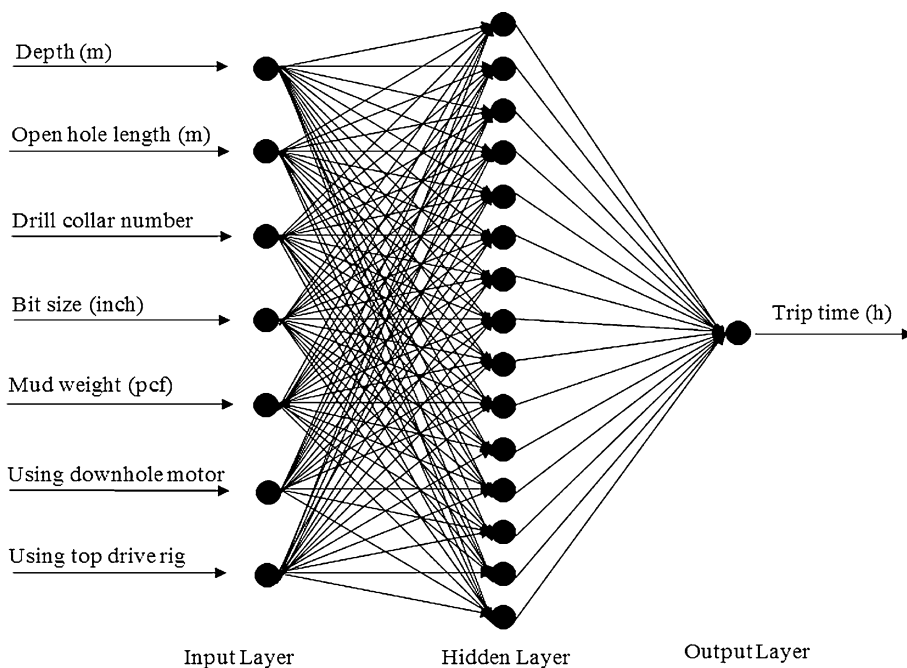


Fig. 2 Time sequence plot of the residuals

Fig. 3 Selected network architecture for trip time prediction



Statistical model

The statistical section of this study was done by SPSS software version 18. SPSS (Statistical Package for the Social Sciences) is a computer program used for statistical analysis. SPSS is among the most widely used programs for statistical analysis.

One of the first steps of calculating an equation with several independent variables is to prepare a correlation matrix for all the variables. This matrix (Table 2) shows the correlation between the dependent variable (trip time) and any other independent variable, and also the correlation among the independent variables. In each cell of Table 2, the first row shows the Pearson correlation coefficient. The second row (sig.) shows the accurateness of the correlation coefficients of the first row. In the third row, the number of cases that computed between variables has been shown. As shown in the correlation matrix Table 2, Depth has the highest correlation with trip time compared to other variables ($r = 0.77$). All of the predictor variables except BS have positive correlation with trip time. There is a negative high correlation between depth (*D*) and BS ($r = -0.80$), which is obvious and so the BS can be predicted by depth.

In regression analysis, at first all seven predictor variables have been used. But in the statistical inferences, the null hypothesis (i.e., the coefficient of each parameter equals zero, at the 5 % significance level) could not be rejected for one parameter, BS. Therefore, another multiple regression model must be calculated by removing the insignificant variable, BS.

After fitting the new linear model to a given data set, an assessment is made of the adequacy of fit. From Table 3, the value of R^2 is 0.77, showing that about 77 % of the total variations in the trip time can be accounted for the independent variables. To test whether the dependent variable (trip time) is related to predictor variables, the ANOVA table (Table 4) is used. Since P value (Sig.) related to F-statistic is less than the significance level (5 %), it is concluded that the dependent variable is related to predictor variables. Table 5

determines at the 5 % significance level, whether it appears that any of the predictor variables can be removed from the full model as unnecessary. As shown in Table 5, the entire coefficients for the new model are significant, i.e., P value of the t statistic for each coefficient is less than significance level (5 %), so all the predictor variables are useful as predictors of dependent variable (trip time). From the previous tables and discussions, it can be said that the appropriate obtained model is of the form below:

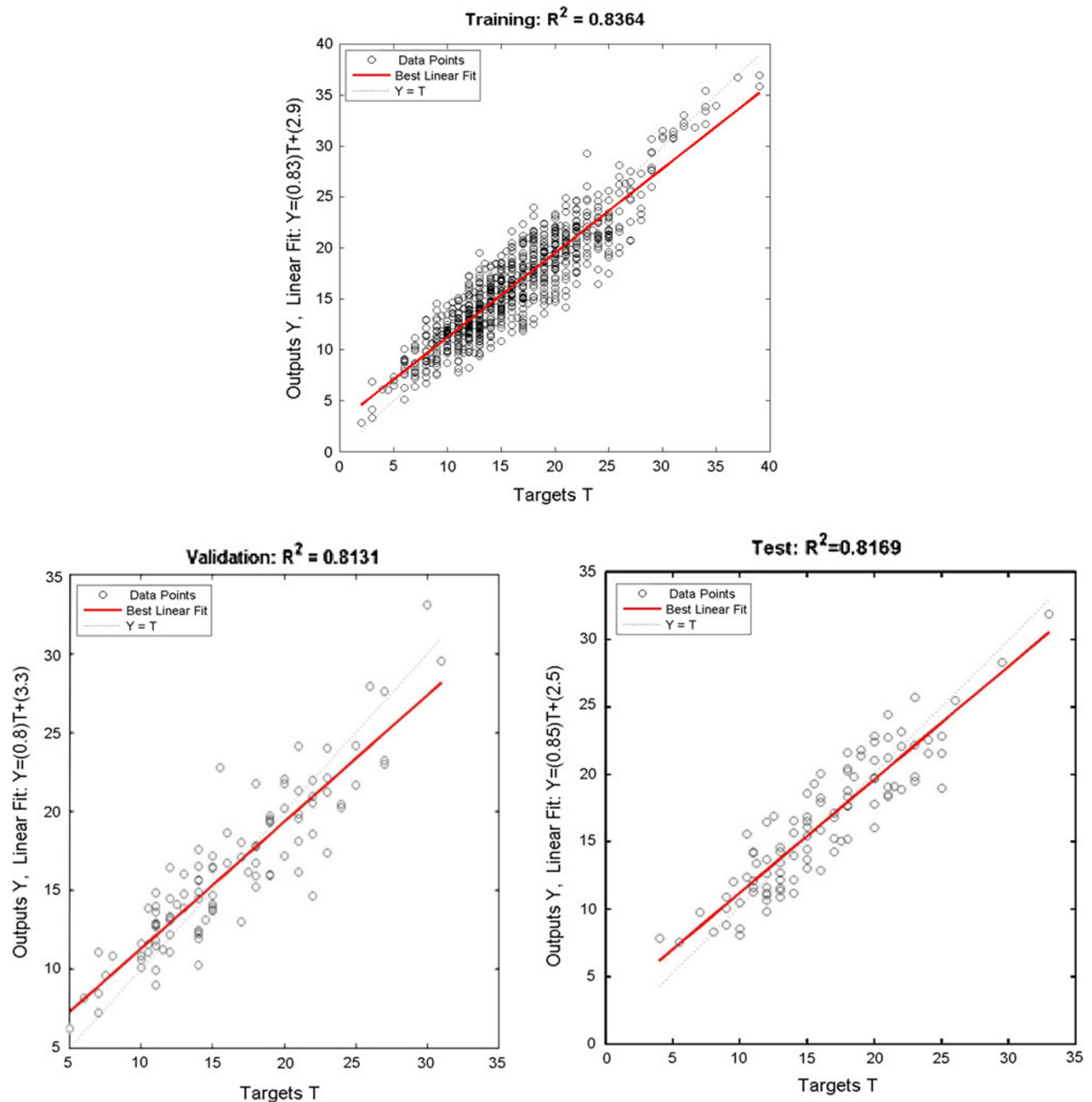


Fig. 4 Regression plots for training and validation and test data

$$\text{Trip time} = a_0 + a_1 * D + a_2 * \text{OHL} + a_3 * \text{DC} + a_4 * \text{MW} + a_5 * \text{DM} + a_6 * \text{TD} \tag{5}$$

The value of the coefficients of Eq. 5 and their units are as follows:

$$a_0 = 1.534 \text{ (hour)} a_1 = 0.004 \left(\frac{\text{hour}}{\text{meter}} \right) a_2 = 0.001 \left(\frac{\text{hour}}{\text{meter}} \right) a_3 = 0.104 \text{ (hour)}$$

$$a_4 = 0.022 \left(\frac{\text{hour}}{\text{pcf}} \right) a_5 = 6.186 \text{ (hour)} a_6 = 2.257 \text{ (hour)}$$

Aptness of model

After obtaining the residuals of regression model, residual plots are created, and it is decided whether or not it is reasonable to accept the assumptions of multiple regression analysis. Figure 1 shows the normal probability plot for the multiple regression model. The normal probability plot is used for evaluating the assumption that the distribution of the errors (residuals) is normal. The points in Fig. 1 fall reasonably close to a straight line, suggesting that the distribution of the error terms does not depart substantially from a normal distribution. Figure 2 shows time sequence plot of the residuals. The residuals in the sequence plot of Fig. 2 fluctuate in a more or less random pattern around the base line zero, which indicates validity of the assumption that errors are independent, and they have constant variance.

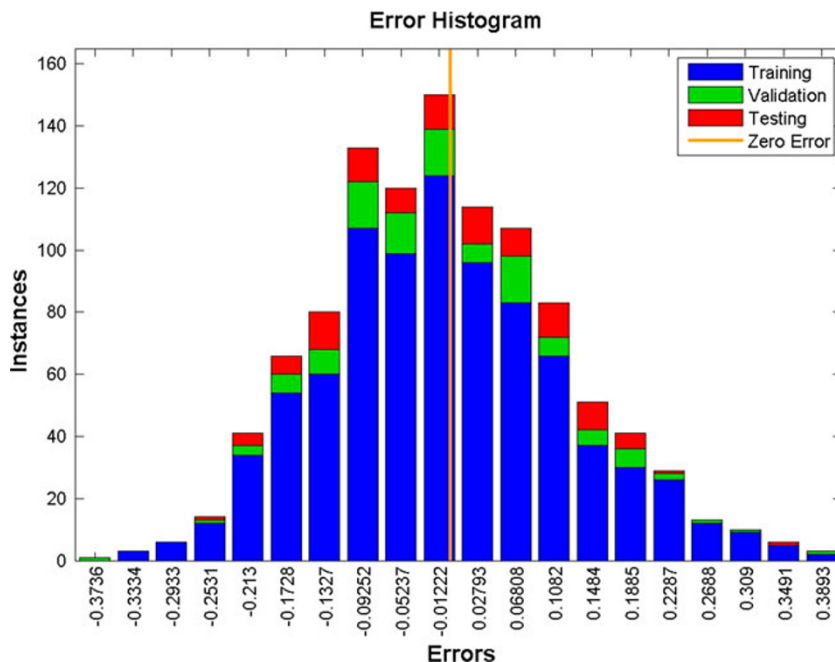
Artificial neural networks

Artificial neural network is highly dependent on the input and output data. Reliable data must be fed into the network to get the reliable output. So, data handling procedures before training the network is of a great importance. “Cross validation” approach was considered to split the available data in this study. This approach requires splitting the data into three representative subsets: training set to calibrate the model, a validation set to evaluate the calibration process at different stages, and a testing set to finally assess the performance of the calibrated model. Another important point to consider is that artificial neural network, like other statistically based models, generally only performs well when interpolating within the data range, they are provided with during the calibration or training phase. For that reason, the maximum and minimum values for each input parameter, as well as each output parameter, have to be contained in the training set (Goda et al. 2005).

Here, 80 % of data sets are randomly devoted to training and 10 % for validation and 10 % to testing. Besides, the data sets considered for training, cover all data range. Before training, it is often useful to scale the inputs and targets so that they are always within a specified range. In this research, the available data have been normalized into the range of -1 to 1.

For prediction using ANN, the MATLAB neural network toolbox has been used. A multilayer feed-forward network has been chosen as network architecture. Using command line operation and writing a code, many runs

Fig. 5 Error histogram plot for ANN model



have done for different networks with one and two hidden layers and different hidden neurons. The final network is a three layer feed-forward back propagation network, whose features are as follows: Levenberg–Marquardt as the training algorithm, MSE as performance function, two layers, seven neurons for input layer, 15 neurons for hidden layer, one neuron for output layer, “tansig” as activation function for hidden layer, and “purelin” for output layer. Figure 3 shows this network graphically.

After training, network performance must be checked. For validating the network, regression plots can be used,

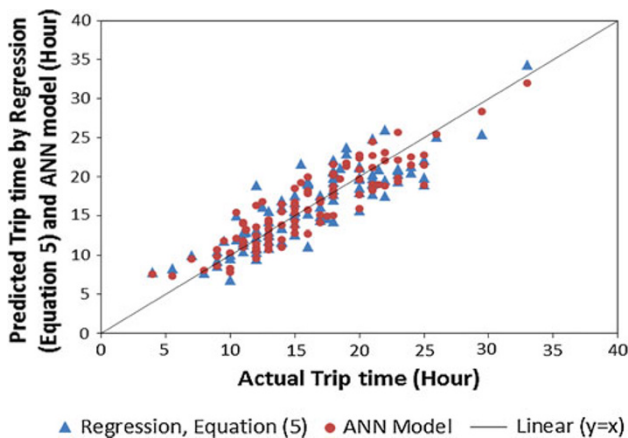
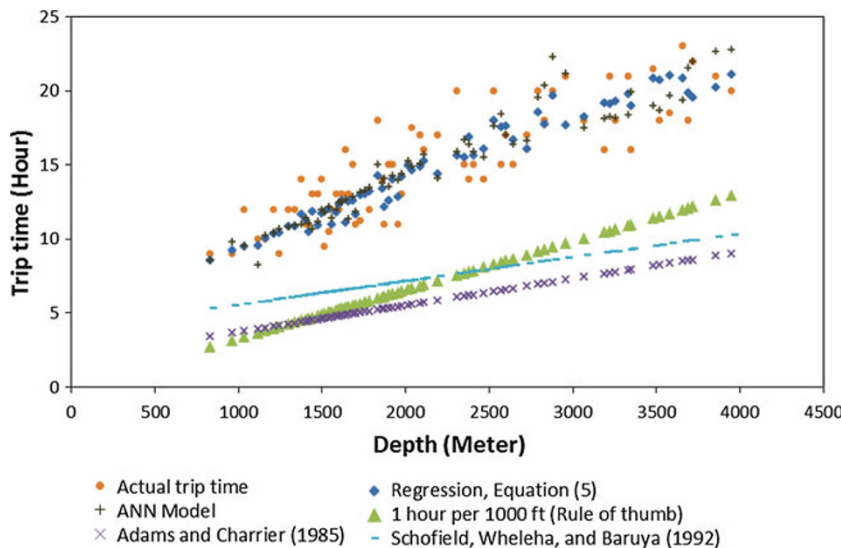


Fig. 6 Comparison of regression model with ANN model

Table 6 Performance table for multiple regression and ANN

Model	R	R ²	Adjusted R ²	Std. error of the estimate
Regression Equation 5	0.872	0.761	0.758	2.41517
ANN	0.906	0.821	0.819	2.05957

Fig. 7 Comparison between trip time estimation models



which show the relationship between outputs of the network and the targets. As shown in Fig. 4, three axes represent training, validation and testing data. Here, training data indicates a good fit. The validation and test results also show R^2 values >0.8 . The error histogram, Fig. 5, can be viewed to obtain additional verification of network performance. This plot shows the distribution of the network errors. It must be reminded that since the data have been normalized so the resulting errors are normalized too.

The histogram can give an indication of outliers, which are data points where the fit is significantly worse than the majority of data.

Discussion

Ten percent of collected data has not input into model developing for evaluating the performance of each developed model and comparison of the ANN model with multiple regression analysis. This is done by making a plot of the predicted trip times versus the actual trip times (Fig. 6) for test data. Table 6 shows the results for the multiple linear regression and artificial neural network model. It can be concluded that the ANN model predicts the trip time from the predictor variables better than the multiple linear regression model.

Despite the superior performance of the ANN models, they are generally considered to have the disadvantage of being less transparent than more conventional models (Goda et al. 2005).

Most of the previous trip time estimation models take into account just one parameter, depth; while in the models developed in this study the effect of other parameters is included like: mud weight, open hole length, drill collar number. For comparing the models developed in this thesis

with the previous models, the regression model Eq. 5 and the ANN model are used. This comparison has been shown in Fig. 7. The data used for comparison is the previous test data. By comparing the trip time predicted by developed models with previous models, it is observed that the developed models' outputs are about 75–100 % greater than outputs predicted by the previous models. This strange difference could be because of the tripping operation regulation in the southern Iranian oil company. The time of a round trip in drilling daily report of NIOC is not just pulling and running drill string into the hole. In addition to pulling and running drill string, trip time also includes mud circulation before trip out for cleaning the hole and well observation at the bottom hole, casing shoe, and in above of drill collars. Consequently, the trip time resulted by the developed models may have a considerable difference with similar case studies.

Conclusions

1. It is obvious that the developed models in this thesis are reliable only for southern Iranian oil fields.
2. Using downhole motor has a significant increasing effect on trip time. This is by the reason of special bottom hole assembly, surface test, and shallow test of downhole motor. Besides, it can be seen that the trip time in a top drive rig is more than a rotary table rig, which may arise from the lack of skills in working with top drive rigs.
3. It has been observed that the values of predicted trip time by the models developed in this thesis in southern Iranian oil fields are about 75–100 % greater than the trip times predicted by available trip time prediction models. The models developed in this thesis predict trip time more accurate than the available trip time production.
4. Although artificial neural networks provide more precise models than regression analysis, it is more

complicated. The power of neural networks appear when there is no idea of the functional relationship between dependent and independent variables. If an idea exists, that is independent parameters are known and it is clear how it effects on the dependent parameter, it would be better to use a regression model.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Adams NJ, Charrier T (1985) Drilling engineering: a complete well planning approach. PennWell Publishing Company, Tulsa
- Azar JJ, Samuel GR (2007) Drilling engineering. PennWell Corporation, Tulsa
- Bordenave ML, Hegre JA (2005) The influence of tectonics on the entrapment of oil in the Dezful embayment, zagros fold belt iran. *J Petroleum Geol* 28(4):339–368
- Bourgoyne AT, Millheim KK, Chenevert ME, Young Jr FS (1991) Applied drilling engineering. Soc Petroleum Eng, USA
- Falcao JL, Maida EE, Dumans CF, Dezen JF (1993). PDC bit selection through cost prediction estimates using crossplots and sonic log data. SPE/IADC paper 25733, SPE/IADC drilling conference, Amsterdam, 23–25 February
- Goda HM, Maier HR, Behrenbruch P (2005). The development of an optimal artificial neural network model for estimating initial water saturation. SPE paper 93307, Asia Pacific oil and gas conference and exhibition, Jakarta, 5–7 April
- Kutner MH, Nachtsheim CJ, Neter J, Li W (2005) Applied linear statistical models, 5th edn. McGraw-Hill/Irwin, New York
- Schofield TR, Wheleha OP, Baruya A (1992). A new fishing equation. SPE paper 22380, SPE international meeting on petroleum engineering, Beijing, 24–27 March
- Short JA (1982) Drilling and casing operations. PennWell Publishing Company, Tulsa
- Sumathi S, Paneerselvam S (2010) Computational intelligence paradigms: theory and applications using MATLAB. CRC Press, Taylor and Francis Group, USA