

***Modeling and Prediction of Drilling Rate of  
Penetration and Optimization of Parameters using  
Group Method of Data Handling Techniques (GMDH)***

By

**Ibrahim Shaikheldin Ibrahim Mohamed**

**ID: 17749**

Dissertation submitted in partial fulfilment

of the requirements for the

Bachelor of Engineering (honor)

(Petroleum Engineering)

January 2015

Universiti Teknologi PETRONAS

Bandar Seir Iskandar

31750 Tronoh

Perak Darual Ridzuan

CERTIFICATION OF APPROVAL

*Modeling and Prediction of Drilling Rate of Penetration and  
Optimization of Parameters using Group Method of Data Handling  
Techniques (GMDH)*

By

Ibrahim Shaikheldin Ibrahim Mohamed

17749

A project dissertation submitted to the

Petroleum Engineering Programme

Universiti Teknologi PETRONAS

In partial fulfilment of the requirement of the

BACHELOR OF ENGINEERING (Hons)

(PETROLEUM)

Approved by

---

(Dr. Mohammed Abdallah Ayoub)

UNIVERSITI TEKNOLOGI PETRONAS

TRONOH, PERAK

JAN 2015

## CERTIFICATION OF ORIGINALTY

This is to certify that I am responsible for the work submitted in this project, that the original work is my own except as specified in the references and acknowledgements, and that the original work contained herein have not been undertaken or done by unspecified sources or persons.

---

IBRAHIM SHAIKHELDIN IBRAHIM MOHAMED

## ABSTRACT

Rate of penetration (ROP) is a master parameter which has direct effect on drilling cost. So, this project is aimed to optimize drilling parameters and predict optimum rate of penetration (ROP).

Utilizing Group Method of Data Handling techniques (GMDH) for modeling and prediction, choosing GMDH particularly, because of its capability to find the nonlinear relationship between drilling parameters and drilling rate of penetration. The previous models which provided some of it in the literature, has limitation to meet the required accuracy.

At the first stage, Drilling Simulator had been used in order to optimize the parameters. After entering the all parameters needed in drilling simulator, start optimizing penetration rate by optimizing weight on bit (WOB) and rotary speed (N) which they have good effect on ROP. The result of the optimization of ROP had been proved by take one well as example and calculate the saved percentage of total drilling operation cost which was 13.69% this reduction result for cost of drilling operation only.

Then the second stage GMDH had been utilized to predict the optimum ROP using the optimized parameter after first stage. The model was developed by a total of 504 data sets. By the ratio of 2:1:1 the data was divided to training, cross validation and testing sets. Trend analysis as well as graphical and statistical tools had been utilized to make sure the correction of the model.

## Acknowledgements

Gracefully, I thank almighty Allah for surrounding me with his blessings, for giving me the courage and the determination to achieve such success.

My thanks and appreciation to my father's soul, Shaikheldin Ibrahim Mohamed, and my entire family for all their permanent support.

I would like to extend my utmost gratitude to my supervisor, Dr. Mohammed Abdalla Ayoub, for his help, care, patience as well as guidance in conducting this project study, despite all difficulties.

I must acknowledge as well, Mr. Elshiekh Mousa Batty, Drilling Engineer - at GREATER NILE PETROLEUM OPERATING COMPANY (GNPOC), and all other engineers in operation department, for their help during my internship at the company and continuous support for my final year project.

Finally, my warm thanks to all my friends, colleagues and lecturers at UTP for being there whenever I needed.

## Table of Contents

CERTIFICATION OF ORIGINALTY .....	II
<b>ABSTRACT</b> .....	III
List of Figures .....	VII
List of Table .....	VIII
CHAPTER (1).....	1
INTRODUCTION .....	1
<b>1.1 Background:</b> .....	1
<b>1.2 Problem Statement:</b> .....	2
<b>1.3 Objective:</b> .....	2
<b>1.4 Scope of Study:</b> .....	2
<b>1.5 Relevance of the Project:</b> .....	3
<b>1.6 Project Feasibility and Significance:</b> .....	3
CHAPTER (2).....	4
LITERATURE REVIEW .....	4
<b>2.1 ROP Models:</b> .....	5
<b>2.1.1 Overview of ROP Correlation Models:</b> .....	5
<b>2.1.2 Bourgoyne and Young's Model:</b> .....	5
<b>2.1.3 Warren model:</b> .....	6
<b>2.1.4 Bingham model:</b> .....	6
<b>2.1.5 Artificial Neural Network (ANN):</b> .....	7
<b>2.1.5 Group Method of Data Handling (GMDH):</b> .....	7
CHAPTER (3).....	8
METHODOLOGY.....	8
<b>3.1 Research Methodology:</b> .....	8
<b>3.2 Project Activities:</b> .....	9
<b>3.3 Project workflow:</b> .....	9
<b>3.4 Project Key Milestone:</b> .....	10
<b>3.5 Gantt chart(1):</b> .....	11
<b>3.6 Project Stages:</b> .....	12
<b>3.6.1 Optimization Stage:</b> .....	12

3.6.2 Modeling and Prediction Stage: .....	14
3.7 Trend analysis:.....	15
3.8 Graphical Error Analysis: .....	15
3.8.1 Cross Plot:.....	15
3.8.2 Error Distribution: .....	15
3.9 Tools:.....	16
<b>CHAPTER (4) .....</b>	<b>17</b>
<b>RESULT AND DISCUSSION .....</b>	<b>17</b>
4.1 Data gathering and processing:.....	17
4.2 Optimization Parameters: .....	18
4.3 GMDH Model:.....	21
4.3.1 Summary of the GMDH model's Equation: .....	22
4.3.2 Trend Analysis:.....	22
4.3.4 Statistical Error Analysis:.....	25
4.3.5: Cross Plot:.....	25
4.4 Error distribution for GMDH model: .....	27
4.5 Comparison between proposed model and others models .....	29
<b>CHAPTER (5).....</b>	<b>32</b>
<b>CONCLUSION AND RECOMMENDATION.....</b>	<b>32</b>
5.1 Conclusion: .....	32
5.2 Recommendations: .....	33
<b>REFERENCES: .....</b>	<b>34</b>
<b>APPENDICES .....</b>	<b>36</b>

## List of Figures

Figure 3.1 Research Methodology .....	8
Figure 3.2 Project Workflow .....	10
Figure 3.3 Gantt chart (1).....	11
Figure 3.4 Gantt chart (2).....	12
Figure (3.5) Formation requirement .....	12
Figure (3.6) Geology Summary.....	13
Figure (3.7) Down Hole Graphics.....	14
Figure 4.1 Sample of data preparation. ....	17
Figure 4.2 Measured ROP using DrillSIM500.....	18
Figure 4.3 Daily Drilling Report (DDR).....	19
Figure 4.4 plan day's vs depth before and after the optimization.....	20
Figure 4.5 schematic diagram of proposed GMDH.....	21
Figure 4.6 Rotary Speed of proposed model .....	23
Figure 4.7 Rotary Speed proposed by numerous authors .....	23
Figure 4.8 Weight on Bit of proposed model .....	24
Figure 4.9 Weight on Bit by numerous authors .....	24
Figure 4.10 Cross Plot Rate of Penetration of Training Sets for GMDH .....	26
Figure 4.11 Cross Plot of Rate of Penetration for Validation Sets for GMDH .....	26
Figure 4.12 Cross Plot of Rate of Penetration for Testing Sets GMDH.....	27
Figure 4.13 Error distribution for the training sets .....	28
Figure 4.14 Error distribution for the validation sets .....	28
Figure 4.15 Error distribution for the testing sets .....	29
Figure 4.16 Cross Plot of Bingham Model vs Actual ROP .....	30
Figure 4.17 Cross Plot of BYD Model vs Actual ROP.....	30
Figure 4.18 Cross Plot of Proposed Model by GMDH vs Actual ROP.....	31



## List of Table

Table 3.1: summary of computer programs used.....	16
Table 4.1 sample of the Optimization .....	19
Table 4.2 Calculation of the cost.....	21
Table 4.3 statistical Parameters for GMDH.....	25
Table 4.4 Statistical Parameters Comparison .....	31

## Abbreviations and Nomenclatures

GMDH: Group Method of Data Handling.

ROP: Rate of Penetration (m/hr).

WOB: Weight on Bit (ton).

N: Rotary Speed (rotation per mint).

DDR: Daily Drilling Report.

BYD: Bourgne and Young's Model.

ANN: Artificial Neural Network.

# CHAPTER (1)

## INTRODUCTION

### **1.1 Background:**

In Oil and Gas industry all considerations are focusing on the profit, which indicates reduction in expenditure. Drilling is the highest cost amongst all the phases in oil and gas industry, so all researchers in the drilling filed are working to optimize various parameters which can affect drilling cost. Rate of penetration (ROP) play as the master parameter, which has direct effect in drilling cost, therefore all models are utilized to predict the optimum rate of penetration (ROP). There are too many parameters affecting penetration rate such as weight on bit (WOB) and floundering phenomena, drilling rotation speed (N), mud rheology, hole cleaning, formation hardness and differential pressure. So, it is complicated to get logical connection among all these parameters to assist an appropriate rate of penetration (ROP). To recognize the complex connection between these variables Group Method of Data Handling (GMDH) is the best choice as all researcher agreed with it.

Osman & Abdel-Aal (2002) say that Group Method of Data Handling techniques (GMDH) is an inductive modeling method built on the principles of self-organization. This modeling approach has been used widely in many areas such as medical diagnostics, weather modeling, marketing and environment systems.

In last 35 years, GMDH is developing as a method of inductive modeling and forecasting of complex systems (Godefroy et al, 2012). Therefore, GMDH modeling approach has been proposed as an alternative modeling tool to predict and propose new correlations, which can minimize the limitations of the existing correlations.

Partial models are the various component subset of base function in GMDH technique. Using least squares method to calculate the known coefficients. Self-organization is a gradually increase the number of partial model component to come out with model structure with optimal complexity indicated by minimum value of an external criterion.

. In literature, there are several models to appreciate rate of penetration (ROP) such as Bingham Model, Modified Warren Model, Young Model and Bourgoyne. All these models used to estimate and predict appreciated (ROP). The purpose of this project to model and predict the drilling rate of penetration and optimize drilling parameters using (GMDH).

### **1.2 Problem Statement:**

- Drilling operation is one of the costly operation in Oil and Gas industry, rate of penetration (ROP) has a direct effect on cost reduction of drilling operation. However, because of the complicated relationship between the ROP and other parameters (Nonlinear relationship), the previous models available in literature have high limitation to provide the required accuracy.
- GMDH recognized by it is capability to discover the nonlinear relationship between the input parameters.

### **1.3 Objective:**

- Developing a model for using real field data from Sudanese oil field.
- Validating the model using synthetic data (trend analysis).
- Optimizing the ROP parameters using Drilling Simulator.
- Testing the newly proposed model using real data and against the best available models in literature.

### **1.4 Scope of Study:**

The scope of study is mainly to model the rate of penetration to optimize of the parameters that will affect reduction of the cost by using (GMDH). The study is divided into two stages; the first stage is to predict the proper ROP. After predicting stage is successfully done, the second stage focuses on optimizing the parameters which lead to cost reduction. This project involves the understanding and ability to deal with the (GMDH) techniques and also involves the understanding of modeling and predicting the ROP and optimize the important parameters. Proper understanding of all these parameters are important in order to keep this project work on the right track.

### **1.5 Relevance of the Project:**

This project is relevance to the author's field since drilling engineer is playing main role in petroleum engineering. The rate of penetration (ROP) is the significant factor in drilling operation which can use to redact the cost on it.

In this project author has to deal with landmark software and GMDH using landmark to do the optimization for the parameters and using GMDH to predicting proper rate of penetration (ROP).

### **1.6 Project Feasibility and Significance:**

As shown in chapter three methodology chapter, the author amid to complete the project and come out with the result within the time frame. The details in the research methodology and project workflow which were linked with the given time in Gantt chart.

The significant of this project is to find the complex relationship between ROP and other drilling parameters which the previous models have limitation to achieve the required accuracy, by using landmark and GMDH predicting optimum ROP which will lead to reduce the cost and the end of the day.

## CHAPTER (2)

### LITERATURE REVIEW

There are many techniques that utilized for reduction of drilling operation cost. This can be achieved by optimize time of operation since time is always money in drilling operation. Time taken to drill any well in drilling operation can be represented by Penetration rate (ROP). Therefore Drilling Rate of Penetration plays main role in drilling optimization. Drilling Model must be developed to come out with rate of penetration.

Drilling models are always find the best mathematical relationship between ROP and other drilling parameters that have important effect on it. Because of the uncertain drilling variables there is no direct or exact mathematical relation for rate of penetration and other drilling parameters, and also their relationship are complex and nonlinear.

Penetration rate can be affected by many parameters such as:

Weight on bit (WOB), bit hydraulic, bit type, rotary speed (N), formation characteristic and mud properties etc, are the parameters affecting rate of penetration. Here, are lots of models that have been proposed for rate of penetration such as Bourgoyne and Young model, Artificial Neural Network (ANN), Bingham model and Warren model, etc.

First step is to review the background in several rate of penetration models but before that there is one method lowering drilling cost, which is cost per foot analysis. That aimed to optimize the rate of penetration. It is based on the optimum drilling operation condition of bit run and the criteria of bit selection or respected bit selection. It can estimate the cost per foot as follow:

$$Cf = \frac{(tr + tc + tt)Cr + trCm + Cb}{\Delta D}$$

For this equation  $C_f$  is the cost per foot drilled cost per unit depth  $\left(\frac{\$}{ft}\right)$ ,  $C_r$  is the fixed operating cost per time  $\left(\frac{\$}{hr}\right)$ ,  $C_b$  is bit cost (\$),  $C_m$  is the down hole motor cost  $\left(\frac{\$}{hr}\right)$ ,  $\Delta D$  is the drilled depth(ft),  $tr$  is bit rotating time (hr),  $tt$  is the total trip time (hr),  $tc$  time pipe connection (hr).

If drilling rate is high the drilling cost will be reduce from the drilling cost equation so ROP can play main role to reduce the cost . So, we can choose one of the models to optimize ROP. There are common models used to optimize ROP.

## **2.1 ROP Models:**

### **2.1.1 Overview of ROP Correlation Models:**

Rotary speed, bit type and many parameters can be selected from the offset wells and this parameters have significant effect on rate of penetration. It will be done by analyzing the performance of single section after separate the field into sections depending on their geological likeness.

The relation between the rate of penetration and the drilling variables is complicated, so there are lots of mathematical models that have been suggested to describe it. Majority of these models are rely on the controllable variables and properties of the formation. Controllable variables such as rotary speed and pump pressure, etc. Here are some common models are summarized below.

### **2.1.2 Bourgoyne and Young's Model:**

Bourgoyne and Young's model (Bourgoyne et al., 1991) used eight controllable variables to describe the relationship between rate of penetration and some of the drilling variable. These eight variables rely on ground formation type and should be determined from offset wells or gathering data in advance.

Bourgoyne and Young's model is one of the widely used models in common practices. In any models, there are many unknown parameters and coefficients. Here, this unknown should be calculated from offset wells based on the drilling experiences in the field because these coefficients have a significant impact on the accuracy of the model.

BYD creators proposed multiple regression method to find the unknown coefficients, but applying multiple regression method is not reliable that it can procedure to meaningful results physically, and also number data point limit is affecting this method. So, recently there are many new mathematical techniques applied to calculate these unknown coefficients, to reach the meaningful result. Example of these methods is Nonlinear least square data fitting with trust –region method is a technique apply to the problem. In some researches to determine the optimum unknown parameters in BYD, they use Genetic Algorithm (GA), to be sure that it can reach the meaningful result. Comparing between GA and trust region method GA is more accurate. For Roller cone bits, BYD proposed the below equation.

$$ROP = f1 \times f2 \times f3 \times f4 \times f5 \times f6 \times f7 \times f8$$

Where, ROP is rate of penetration  $\left(\frac{ft}{hr}\right)$ ,  $f1$  is the function of the formation drill ability (mud type, bit type, formation strength), symbolize the impact of compaction on the penetration rate represent by  $f2$ ,  $f3$  &  $f4$ , signifies the overbalance on ROP,  $f5$  &  $f6$  respectively model the effect of bit weight and rotary speed on ROP, effect of tooth wear and bit hydraulic represent by  $f7$  &  $f8$  respectively.

### **2.1.3 Warren model:**

Perfect-Cleaning Model or the process of the drilling using tri cone bit model was founded by Warren in 1987 and modified by Hareland and Hoberock 1993. The idea came from the rate of cutting removal from the bit is equal to the rate at which the chips are formed under steady state drilling condition. So, proposed that the ROP can be controlled by cutting removal process, cutting generation process or combination of both. The perfect cleaning model is developed to imperfect cleaning model because in field or real practices the rate of penetration is remarkably inhibited by the rate cuttings removal under the bit.

### **2.1.4 Bingham model:**

Maurer model has been modified to present an experimental model which called Bingham model, simple model valid for low value of rotary speed and weight on bit. The answer usually less reliability because neglecting important factor which is the depth.



### **2.1.5 Artificial Neural Network (ANN):**

ANN is a well-liked techniques among most of the researchers due to it is ability to recognize the nonlinear relationship with available data. Hajizarid (2007) established neural network for reservoir engineer including initial data which are distributed into three section which are training, validation and testing. The best performance come when we training our data many times in the intelligent system. Input, hidden and output are the three layers which neural network consist of .To help the model to be more reliable and more accurate the prediction model is built again, do training the model with extra data point. So the generalization stop improving the training stop automatically. Measure the network generalization with validation data come to stop when the generalization stop improving. Independent measurement of performance come out with test data before and after training. so can indicate that you have to do more training to get better output or you are satisfied .

### **2.1.5 Group Method of Data Handling (GMDH):**

There are always some limitation in modeling to overcome with this limitation propose Group Method of Data Handling to help on that. GMDH is combination of the advantage of neural network with those advanced statistic method to provide faster, easier to use more accurate tool (Abdel-Aal et al,1997).

GMDH or also known as polynomial neural networks, abductive and statistical learning networks is an algorithm modeling tool for identifying nonlinear relations between input and output variables (Oh & Pedrycz, 2002).

GMDH meets expectations by building successive layers with connection. The layers are basic polynomial terms which are made by utilizing straight and nonlinear relapses. The first layer is constructed by registering relapses of the input variables and after that picking the best ones. The second layer is made by processing relapses of the qualities in the first layer alongside the input variables. This process continues until the net stops getting better (Ward Systems Group Inc., 2008).

## CHAPTER (3)

### METHODOLOGY

This chapter is explain the methodology used to optimize the parameters and build the model and prediction of drilling rate of penetration (ROP). Utilizing Drilling simulator 500 for optimization and Group of Method Data Handling techniques (GMDH) for building a model and prediction.

#### 3.1 Research Methodology:

This project will apply GMDH Techniques to predict the proper rate of penetration (ROP) and to optimize the parameters which can help to reduce the cost of drilling operation in future wells. The research methodology is illustrated in the following chart:

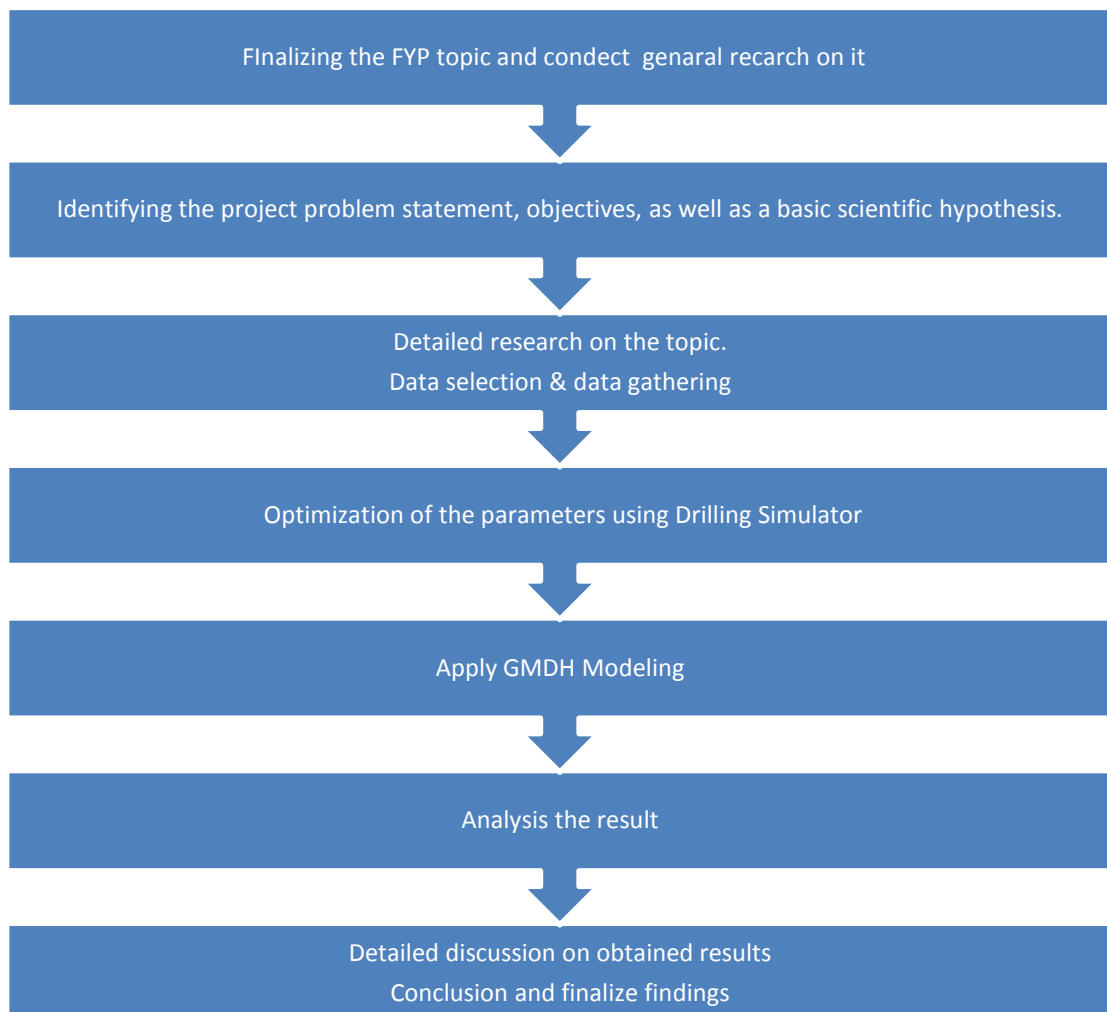


Figure 3.1 Research Methodology

### **3.2 Project Activities:**

The project activities can be divided to two categories:

The first category:

- Start with researching and reading to make strong background about the rate of penetration and other drilling parameters such as weight on bit, rotary speed, etc, and its effect on the drilling operation.
- Literature review had been done through reading SPE papers and referring to some books such as Apply Drilling Engineer and Drilling Engineering a Complete well.
- Start learn about how to use Drilling Simulator which will be used to do the optimization.
- Read more details about GMDH techniques.

Second category university requirement which are:

- Extended Proposal
- Proposal defense (oral presentation).
- First draft Interim report.
- Final draft Interim report
- Progress report.
- Pre SEDEX.
- Dissertation.
- Technical paper.
- Viva.

### **3.3 Project workflow:**

The workflow is schemed below to clear the picture about the project activities.



Figure 3.2 Project Workflow

### 3.4 Project Key Milestone:

- Collect real data that involve desirable parameters which will help in the project form the industry.
- Preparation of the data.
- Study more about GMDH techniques and the required program to apply it.

- Use Drilling Simulator to optimize the parameters.
- Apply GMDH to model and predict ROP.
- Compare the result with the real data and see the accuracy of the techniques.
- Improving and adding information to this research paper.

### 3.5 Gantt chart (1):

Discretion	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Topic selection														
Background study about the project														
Submission of Extended proposal														
Literature Research about (ROP&GMDH) Proposal defense														
Gathering data & parameters involve														
Submission of interim draft report														
Submission of interim report														

Figure 3.3 Gantt chart (1)

### Gantt chart (2):

Discretion	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Work on the optimization using Drilling simulator														
Apply GMDH and submission of progress report														



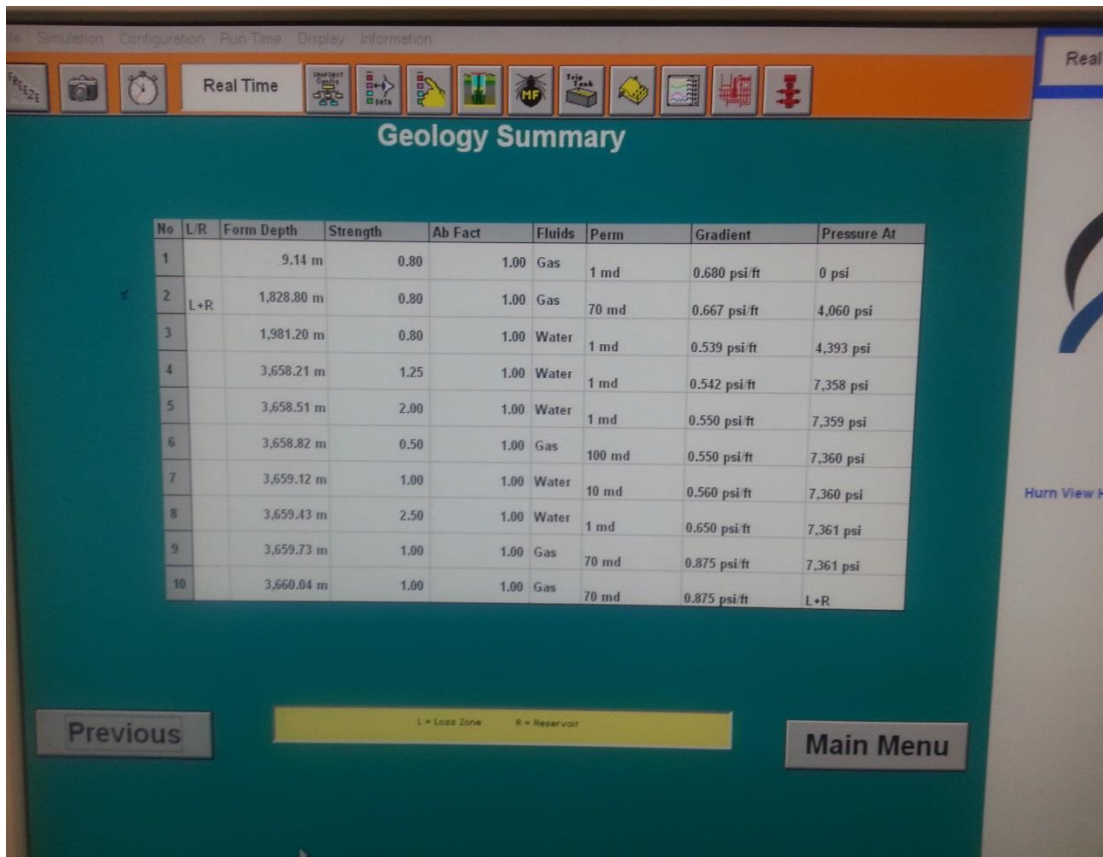


Figure (3.6) Geology Summary

Down Hole Graphics used to monitor the drilling as shown in figure (3.7), which can be used to monitor the hole depth, shoe depth and bit depth which indicate are you drilling now or not the bit is off bottom or not.

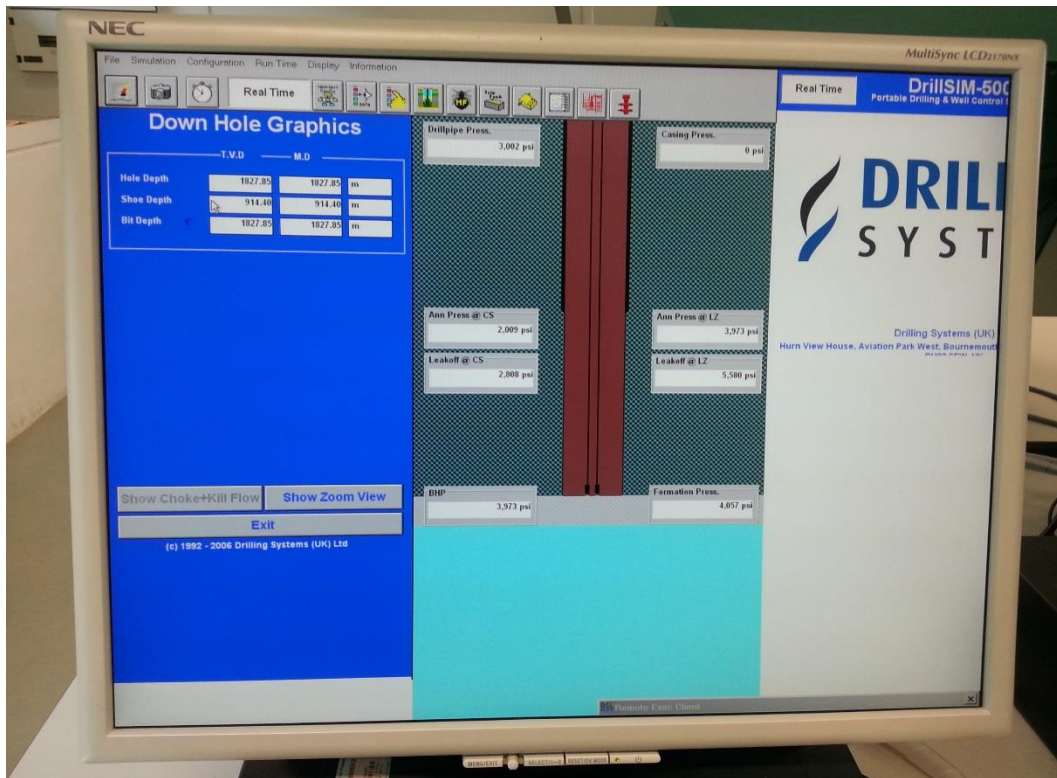


Figure (3.7) Down Hole Graphics

After the entering the formation and geological summary and other requirement of the Drilling Simulator the author stated optimization of the data, and optimized the Drilling Rate of Penetration ROP using Weight on Bit (WOB) and Rotary Speed (N) (RPM) rotation per mint. The result of the optimization will be discussed with father explanation and show the effect of the optimization on the drilling operation cost.

### 3.6.2 Modeling and Prediction Stage:

In this stage Group Method of Data Handling Techniques had been used to build the model. And predict the optimum rate of penetration (ROP). In this Project, MATLAB software, had been used because of it is ability to give a flexible programming and graphic visualization. MATLAB provides an excellent way to keep an eye on the performance of the validation, training and testing data sets. all together which facilitate the optimization process and the sensitivity analysis.

To ensure the parameters were well optimized, a MATLAB code was developed and trained the parameters to get optimum result. For artificial neural network model, the



used of cross validation data set was to avoid the problem of under-fitting and over-fitting.

To check the generality and stability of the model Cross validation data set is presented to the network after each epoch of training.

Development of Group Method of Data Handling techniques model after choosing all the optimized parameters after first stage. The GMDH is inductive approach based on sorting out the data and selection of the best solution by minimum external criterial characteristic. Regularity criterion is a polynomial of GMDH techniques represent by the input regime to output through the application. And often represented by Average Absolut Percentage Error (APPE).

### 3.7 Trend analysis:

To check this model is physically correct or not trend analysis should be present. In this project the trend analysis made for rotary speed (N)(RPM) and weight on bit (WOB) which is confirm the physically with the other models and the result of the Drilling Simulator. The synthetic set first trend the rotary speed is change while the other parameters was changed. And the second trend the WOB is change while the other parameters is constant.

### 3.8 Graphical Error Analysis:

Graphical analysis are tools used to aid in accuracy and visualization of the new model, the graphical error used are:

#### 3.8.1 Cross Plot:

The cross plot are used to compare between the all models new one and old one in literature. Using a 45° straight line between predicted and measured rate of penetration. When the value is closer to the line that means that better result between the measured and predicted.

#### 3.8.2 Error Distribution:

For the proposed GMDH model (training, validation and testing) the error distribution shows the error showing histograms. The normal distribution curve had been fitted to each one of them. The stander deviation was equal to 1.0 and the 0% mean was said

to the normal distribution error.in this case the normal distribution was used to describe the error tendency around the mean.

### 3.9 Tools:

In this project the tools used for two stage the first stage optimization utilizing drilling simulator (DrillSIM500). Second stage using MATLAB software and for modeling purpose GMDH Algorithm at MATLAB software was used.

**Table 3.1: summary of computer programs used**

Tool	Function
Drilling Simulator (DrillSIM500)	Optimization the parameters.
GMDH at MATLAB software	Modeling and Prediction.
Microsoft Office Word	To write a report, data, etc.
Microsoft Office Excel	To prepare data sheet and calculation.

## CHAPTER (4)

### RESULT AND DISCUSSION

This chapter will start with the data gathering and selection as it is effect on drilling operation. Then will move to the optimization stage result and discussion and then followed by GMDH developing model and trend analysis. Finally, the statistical and graphical comparison with of the developed model.

#### 4.1 Data gathering and processing:

The data was collected from real oil field and it represent six different wells. The most difficult job was how to get the data. The author extract this data from Daily Drilling Report (DDR), which is the record of everyday operation, the sample of the DDR at the appendix.

The main data extracted from Daily Drilling Report are bit size, rotary speed (N), weight on bit (WOB), mud weight (MW), depth, torque, flow (GPM), stand pipe pressure (SPP), stroke per min for pump and the most important one is actual ROP (m/hr) all these data put it in one excel sheet. Figure 4.1 shows sample of the data preparation.

Depth	Bit Size	N(RPM)Min	N(RPM) Max	N optimized	(Min)WOB	(Max)WOB	optimized W	(Min)MW	(Max)MW	Torque	(PSI)SPP	(GPM)FLOW	ROP m/hr
0	17.5	90	100	100	4	6	6	8.4	8.6	4	1060	680	16
1660	8.5	90	100	150	6	8	9.5	11.2	11.3	2.83	2030	550	10.8
1660	8.5	130	135	150	4	8	9.5	11.2	11.3	1.16	2030	550	9
1660	9.875	90	100	150	2	4	9.5	11.2	11.3	4	2030	550	11
1660	9.875	110	115	150	4	6	9.5	11.2	11.3	3.25	2030	650	11.3
1660	9.875	100	110	150	4	6	9.5	11.2	11.3	3	2030	650	11
1660	9.875	100	110	150	5	6	9.5	11.2	11.3	3	2030	650	11

Figure 4.1 Sample of data preparation.

504 data set had been collected and used in the optimization and modelling.



Country	Province	Latitude	Longitude	Block Name	Field Name	Contractor	Rig Name	Type of Rig	Daily Cost	\$33,038	AFE No.	AFE Cost
Sudan	S Kordofan	10°17'15" N	29°21'47" E	2B	Heglig	ZPEP	Rig 797	Drilling	Cum. Cost	\$928,506	142236	\$2,278,190
Ann. BOP Size	W.P.(psi)	Pipe Ram Size	W.P.(psi)	Blind Ram Size	W.P.(psi)	Section A Size	W.P.(psi)	RKB to GL	GL Elevation	Move Date	06-May-2014 00:01	
13 5/8	5000	5	5000	5	5000	13 5/8"	3000	7m	400m	Spud Date	18-May-2014 20:00	
Last Csg Size(m)	Csg Shoe	Float Collar	Cement Top	Last BOP test	Csg Test	Next Csg Size	Planned Depth	Kill Speed(SPM)	Kill Pres(psi)	TD Date		
9 5/8	558	545		05/22/2014	05/22/2014	7	1650			Release Date		
M. Depth mkb	TVD mkb	Last Depth	Progress m	Daily Hours	Daily ROP	Cum. Drig Hrs	Well Avr. ROF	Authorized Depth	Survey Tool	Survey Depth	Inc.(deg.)	Azi.(deg.)
1516	1516	1312 m	204	16.1	12.7	51.50	17.50 m/h	1650 m	wire line	549 m	1	
Mud Type	Hole Loss	H. Progr. Loss	Surface Loss	S. Progr. Loss	Lag Check@	Lag Time (min)	Total Gas	Trip Gas	Bg Gas	Total Fuel	Fuel Used	Fuel Stock
KCL												93,000
Mud Properties	Agent Used	Daily Acum.	Mud Pumps	Pump 1	Pump 2	Bit Descriptio	Current Bit	Previous Bit	BHA No.	2		
Mud WT(ppg)	10.9	Bentonite	Make	3NB 1300C	3NB 1300 C	Run n.	2	1	BHA Name	Rotary BHA		
Viscosity(s)	51	Soda Ash	Type	TRIPLEX	TRIPLEX	Bit n.	2RR	1	BHA Elements	OD(m)	Length(m)	
PV(mPa.S)	20	Caustic Soda	Strk length mm	305.0	305.0	Size(m)	8 1/2	12.25	BIT	8.5	0.23	
YP(Pa)	26	XCD	Liners (mm)	160	160	Type	MRS619KLAPX	GT 117	BS	6.5	0.9	
Gels 10/10(Pa)	5/6	KCl	Stk per min	60	60	Nozzles	6*14	3*16,1*14	2 DC	6.5	18.07	
API FL(ml)	4.8	Flowzan	Hydraulic	Current Bit	Previous Bit	Serial Numr	SC2250	70	STAB	8.375	1.83	
HTHP FL(ml)		Barite	Bit Size(in)	8 1/2	12 1/4	Depth in (m)	558	90	8 DC	6.5	72.39	
MBT(g/l)	2.5	Silicate	SPP	1710	790	Depth out (m)	SIH	558	JAR	6.5	5.66	
PH	12	PACLV	Flow (gal/min)	550	500	Mtrs.Drilled	958	468	DC	6.5	9.02	
Sand(%)	0.3	PAC-RL	A.V.(m/min)DP	105.00	42.00	Rot.Hours	61.2	32.6	HWDP	5	110.33	
Solid(%)	10	QS-2 (M+C)	A.V.(m/min)DC	140.00	61.00	AVR ROP	15.6	14				
	4700	PalBio 35%	TFA(sqin)	0.90	0.74	WOB (KDaN)	8 tone	4-6				
	5	Sodium Bicar	Hydraulic HP	116	296.00	RPM	100	100				
	5	PAL CAL 130	Jet Vel (m/s)	60	92.00	LADC	PDC	117				
	15	AVR Cost	Bit P. Drop (psi)	363.00	726.00	Make	Smith	PDC				

Figure 4.3 Daily Drilling Report (DDR)

After been sure that the data enter it was match the actual start optimizing the ROP by increasing the rotary speed (N) and weight on bit (WOB) until the author got the optimum of WOB and N which lead to the optimum ROP. Record all the output data for the optimization.

Table 4.1 sample of the Optimization

Depth (m)	Min (N)	Max (N)	Optimized (N)	Min (WOB)	Max (WOB)	Optimized (WOB)	Actual (ROP) (m/hr)	Optimized (ROP)(m/hr)
120 (SH)	110	160	135	4	15	9.5	12.2	24
440(Sst)	90	160	140	4	15	11	14	31
1320(Sst/SH)	110	160	150	6	15	10	11	24

Taking one well as example to see effect of the optimization on the drilling operation cost.

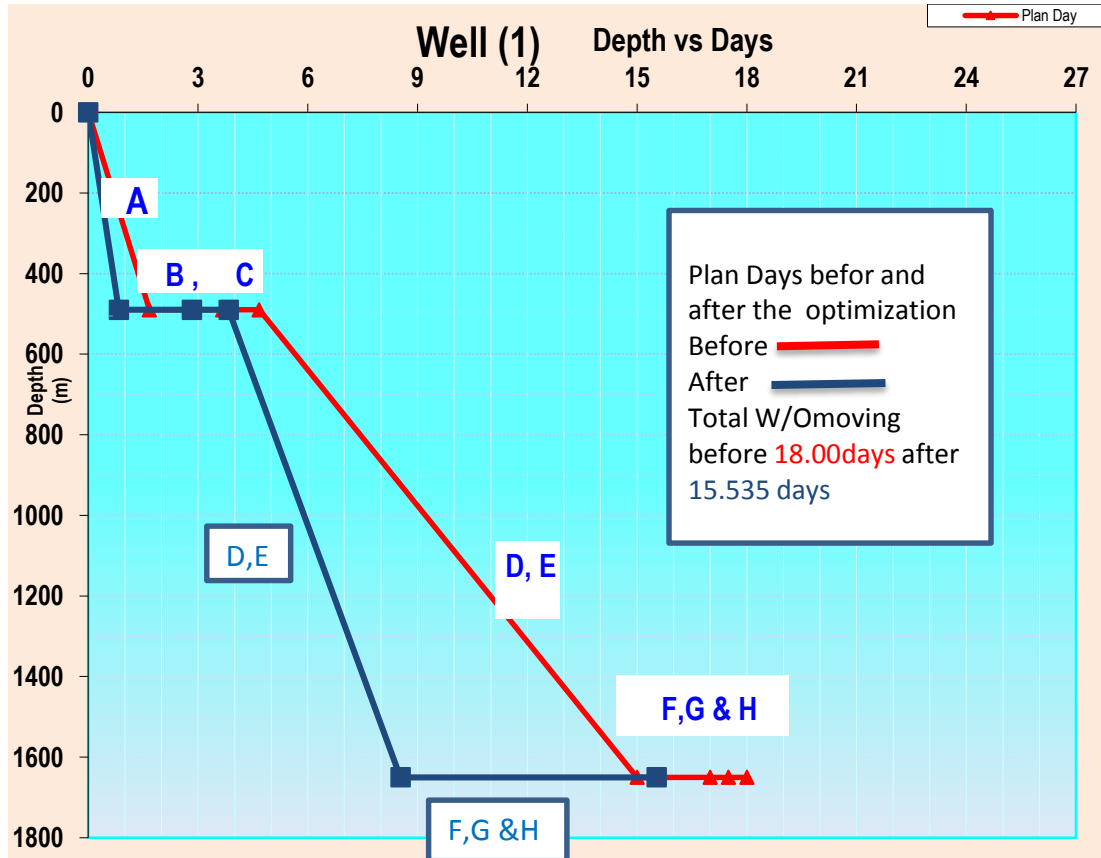


Figure 4.4 plan day's vs depth before and after the optimization

The represented of (A-H), A - Drill 12 1/4" surface hole to 490m, B - Run 9 5/8" csg, C - Cementing & Install Wellhead, D- Drill 8 1/2" hole Section, E- Log Production hole, F - Run 7" csg, G- Cementing and H - Rig Down.

After the proposed of the days be shorter because of the drilling operation fast specially when drilling the surface hole and the main hole. So, the author will take one of the wells to calculate the saved cost. The next table will show the saved cost and the percentage from the total cost. All the cost by USD dollar, the cost of the cost/day =100,854.07 USD/day, Operation cost while Moving =75,000 USD/day

Operation cost while Drilling =82,210 USD/day.

Optimization happen only during the drilling without moving.

**Table 4.2 Calculation of the cost**

Total cost before the optimization	1,479,780 USD
Total cost after the optimization	1,277,835 USD
Saved	202,647.65 USD
Percentage of saving amount	13.69%

### 4.3 GMDH Model:

The final model is consist of two layers, the parameter input was depth, bit size, rotary speed, weight on bit and mud weight. Among these five inputs only three input parameters had shown effect on the prediction on the rate of the penetration (ROP).

This topology was achieved after serious of the optimization process which done by keep an eye on the performance of the GMDH networks until the best output give excellent result.

Next figure will show the schematic diagram of proposed GMDH topology. To predict the rate of penetration (ROP) the parameter effect at the last model is bit size, rotary speed (N) and weight on bit (WOB).

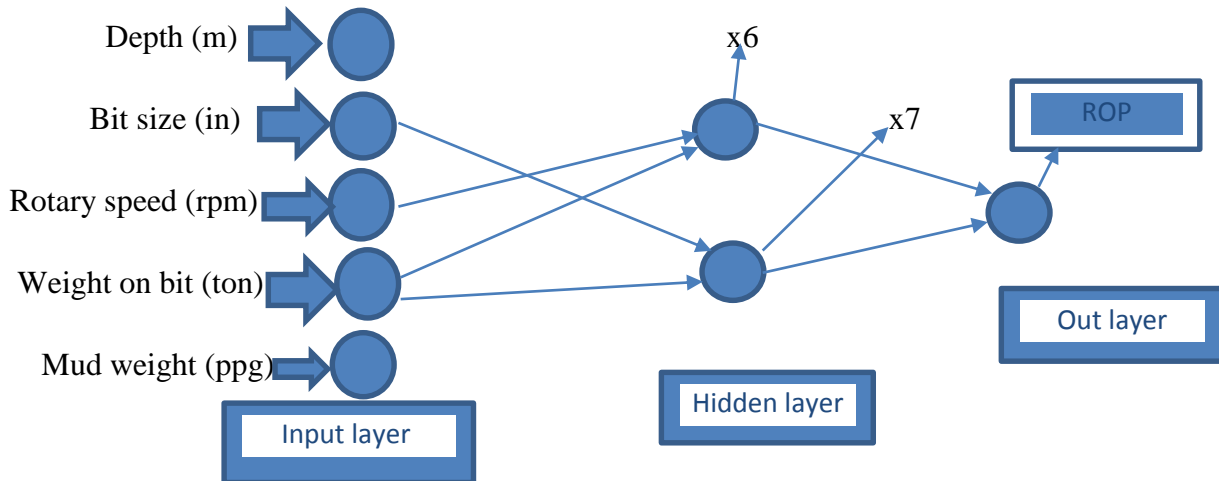


Figure 4.5 schematic diagram of proposed GMDH model

### 4.3.1 Summary of the GMDH model's Equation:

As described before the model consist of two layers as follows:

Number of layers =2

Number of used parameters to predict ROP =3

Layer #1

Number of neurons: 2

$$x_6 = 31.483938 - 4.4347767 * WOB - 0.10579705 * N + 0.08543123 * N * WOB - 0.20368996 * WOB^2 - 0.002369034 * N^2$$

$$x_7 = -77.949264 + 9.5920327 * WOB + 6.5993326 * Bd - 0.34635192 * Bd * WOB - 0.10765329 * WOB^2 - 0.12919722 * Bd^2$$

Layer #2

Number of neurons: 1

$$y = 0.7959967 + 2.534395 * x_7 - 1.5833712 * x_6 + 0.1221932 * x_6 * x_7 - 0.10546542 * x_7 * x_7 - 0.01597001 * x_6 * x_6$$

Where: WOB= weight on bit

N= rotary speed

Bd= bit size

### 4.3.2 Trend Analysis:

On this part of the trend analysis the author was focus on weight on bit (WOB) as well as rotary speed because they have a very important effect on the physical correction of the model. On the other hand, compare these trends with other models found on Apply Drilling Engineering book. So, well start with rotary speed (N) in this proposed model and others models.



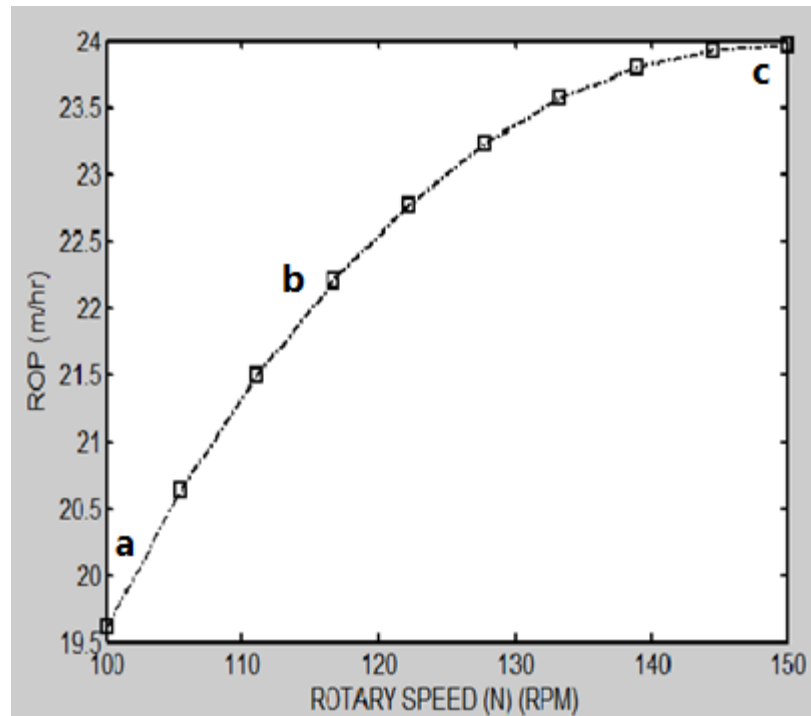


Figure 4.6 Rotary Speed of proposed model

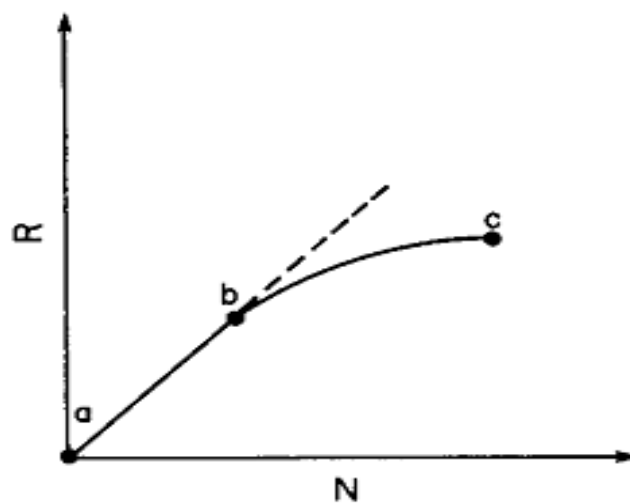


Figure 4.7 Rotary Speed proposed by numerous authors

The typical penetration rate (ROP) vs the rotary speed (N) shown at the figure 4.6 and 4.7, represent the proposed model and other numerous authors respectively. Typically this plot obtain by changing in rotary speed (N) and make the other variable constant. Clearly shown that the ROP increase linearly with rotary speed at low value. At the higher value of rotary speed observe that the respond of penetration rate to increase is

diminishes. The justification of poor respond of penetration rate at higher rotary speed (N) usually attributed to less efficient of bottomhole cleaning.

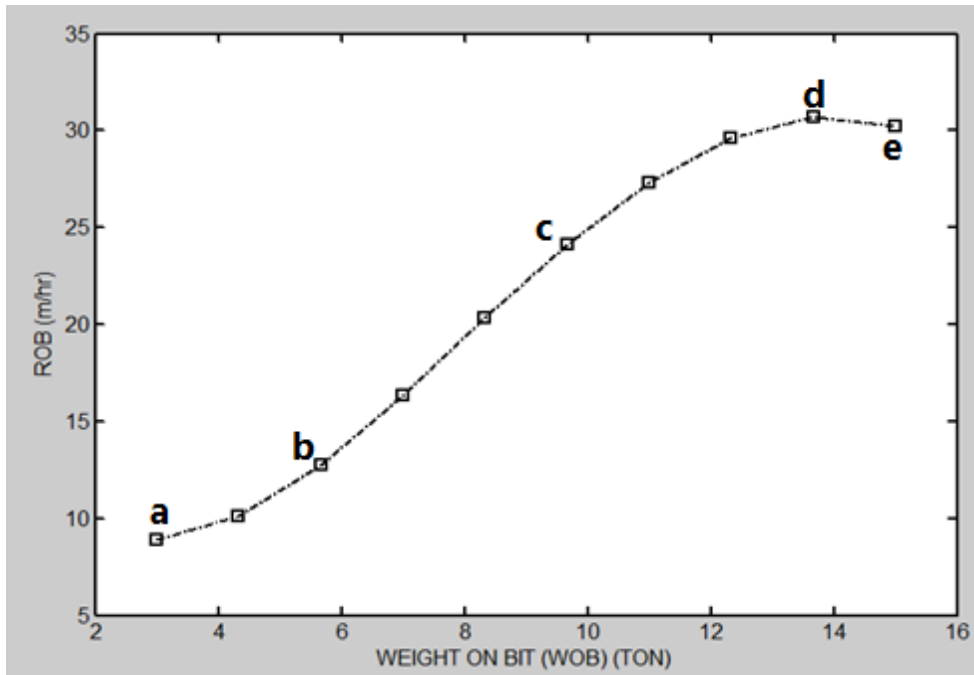


Figure 4.8 Weight on Bit of proposed model

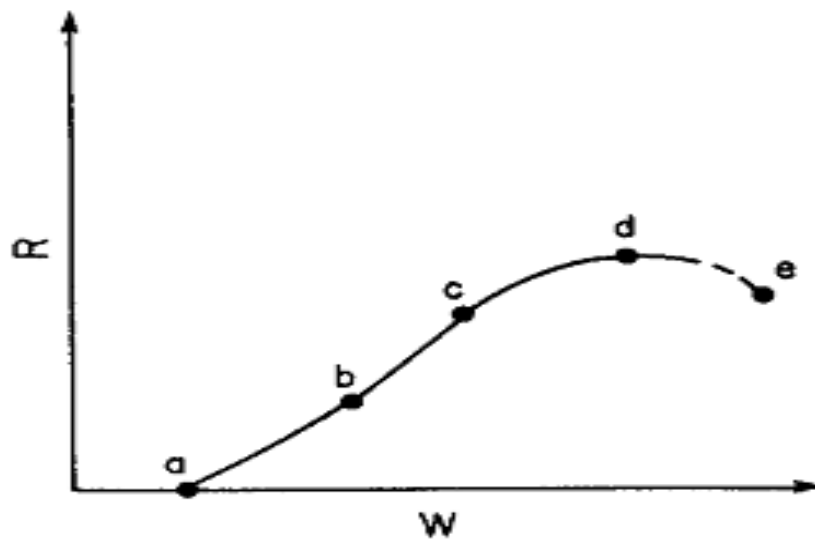


Figure 4.9 Weight on Bit by numerous authors

Figure 4.8 and 4.9 represent the typical plot of weight on bit Vs penetration rate done by proposed model and other numerous authors. Before the threshold bit weight is applied at point (a) there was no important penetration rate was obtained. Then, observed rapidly increasing of the penetration rate with increasing in weight on bit in segment (ab), a linear curve is observed at moderate bit weight (bc). Moreover, at higher values of weight on bit, subsequent increase in bit weight cause slight improvement in penetration rate (cd). In some cases, extremely high value of weight on bit lead to a decrease in penetration rate (de).

#### 4.3.4 Statistical Error Analysis:

Absolute Average Percentage Error (AAPE) and the coefficient of determination ( $R^2$ ), was the statistical error used as shown in table 4.3.

**Table 4.3 statistical Parameters for GMDH**

Statistical Parameters	Training	Validation	Testing
AAPE	5.5484	6.8428	4.9129
$R^2$	0.9211	0.8716	0.9152

#### 4.3.5: Cross Plot:

Cross plot of predicted rate of penetration (ROP) verses measure rate of penetration for the proposed GMDH model represent by figures from 4.10 to 4.12.

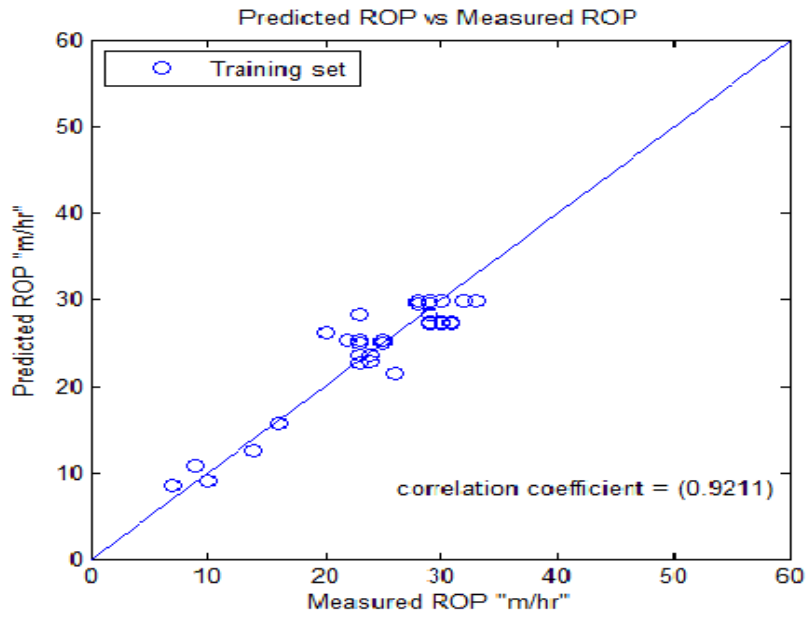


Figure 4.10 Cross Plot Rate of Penetration of Training Sets for GMDH

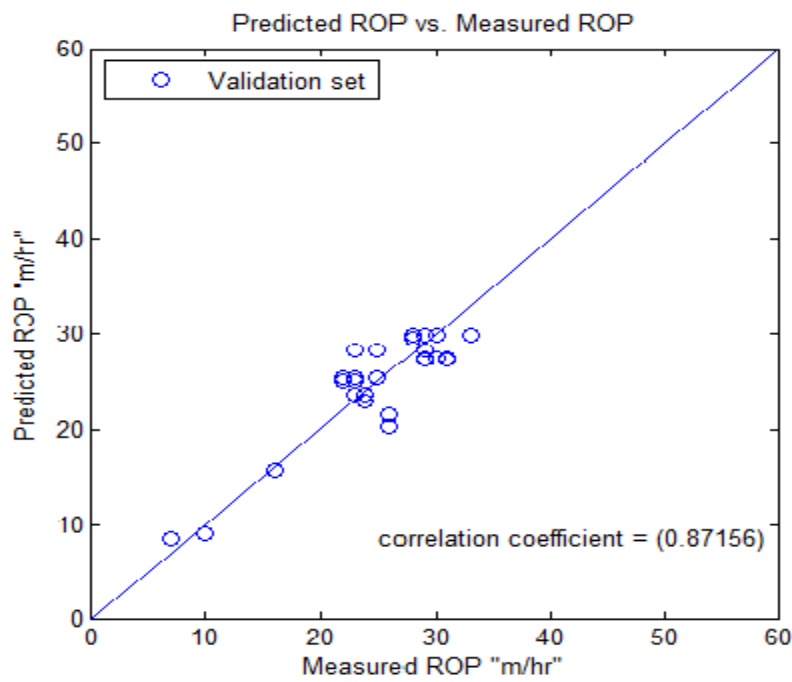


Figure 4.11 Cross Plot of Rate of Penetration for Validation Sets for GMDH

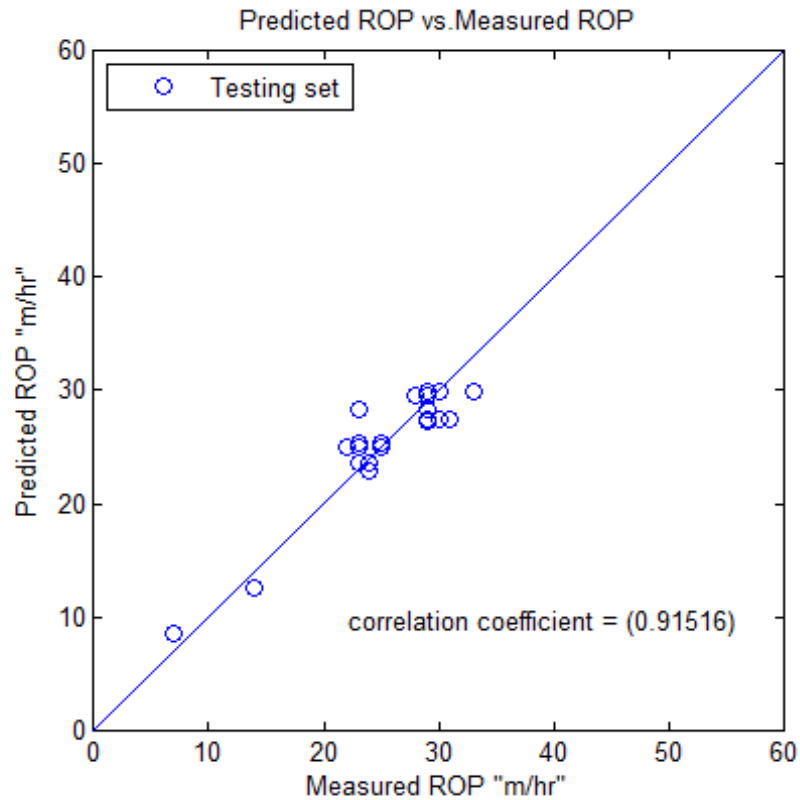


Figure 4.12 Cross Plot of Rate of Penetration for Testing Sets GMDH

#### 4.4 Error distribution for GMDH model:

The upcoming figures shown the error distribution for Training, Validation and Testing. The significant of the error distribution appear because of the clear idea about the performance of the model for all data sets.

In this proposed model result, all sets have normal distribution without any observation of shifting towards negative or positive which indicates a good estimation.

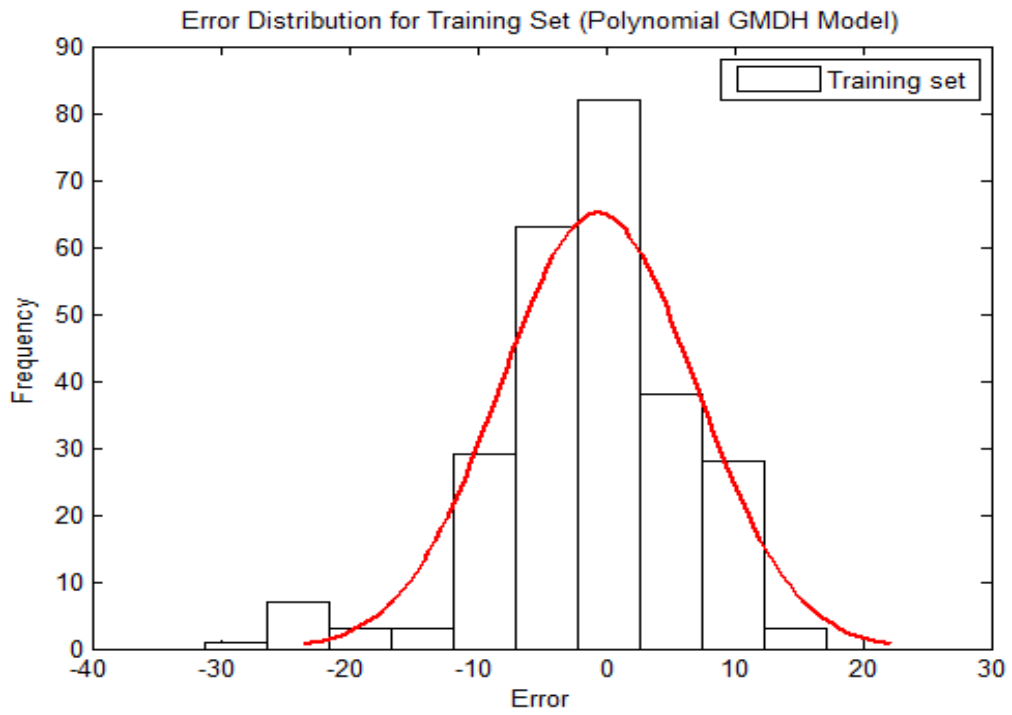


Figure 4.13 Error distribution for the training sets

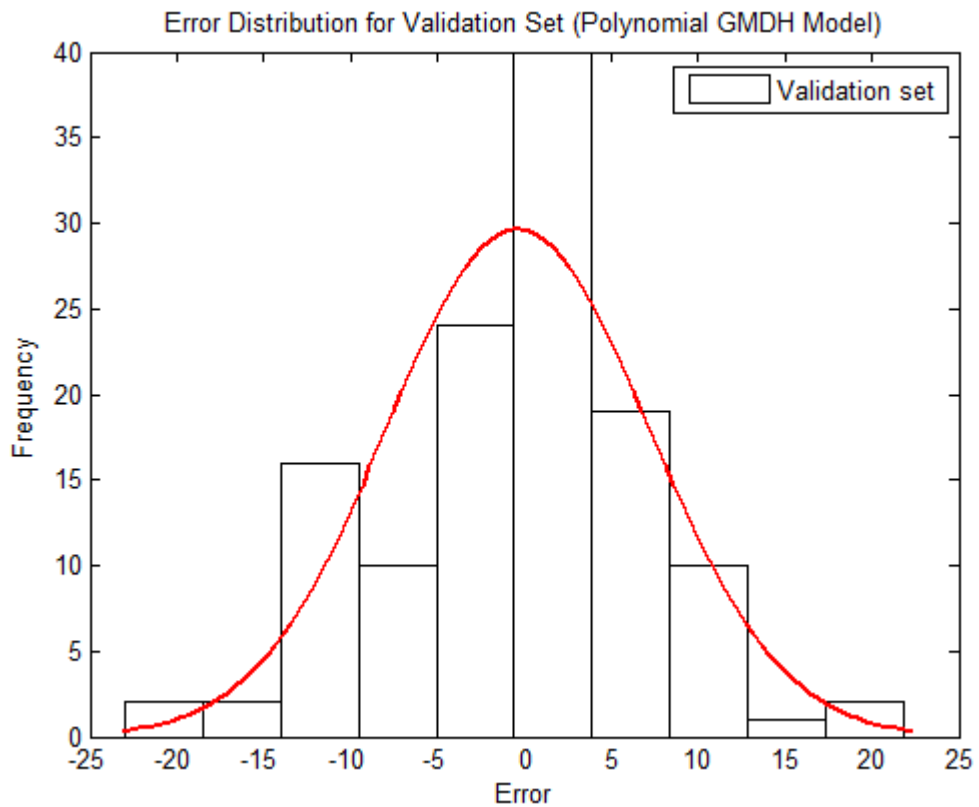


Figure 4.14 Error distribution for the validation sets

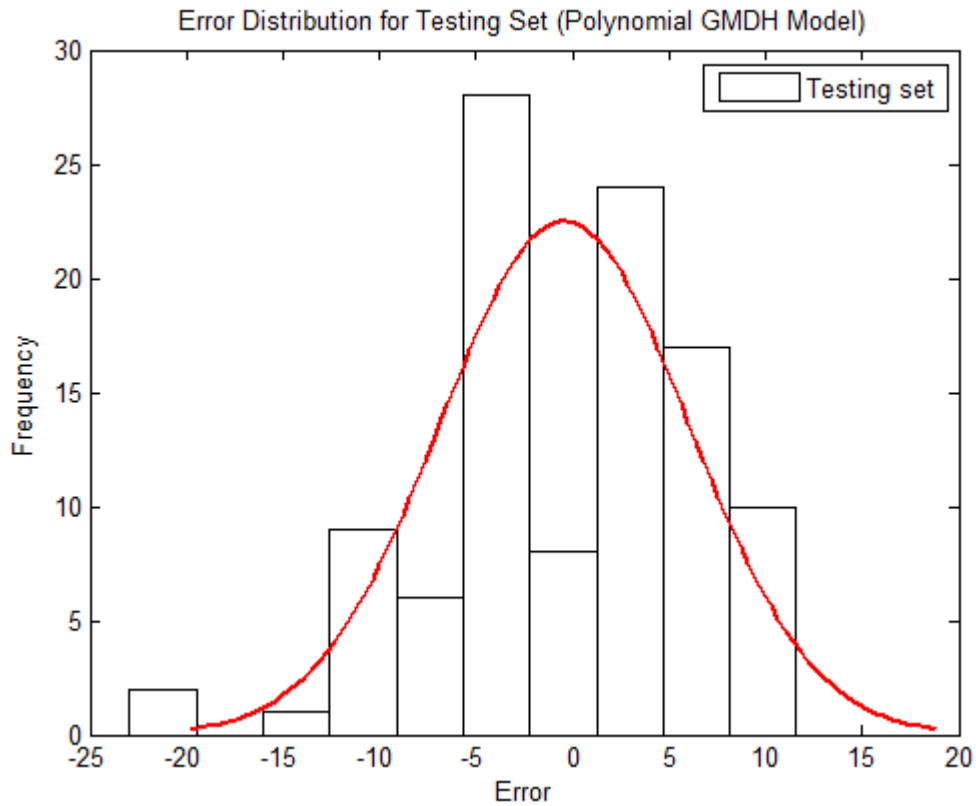


Figure 4.15 Error distribution for the testing sets

#### 4.5 Comparison between proposed model and others models

The model proposed had been determined by using Group Method of Data Handling Techniques (GMDH), specifically choose GMDH technique because of it is ability to find nonlinear relationship between the input data. Therefore, the author did some comparison between the proposed model and two method Bourgne and Young's model (BYD) and Bingham model using the 504 data sets, which was used by GMDH to propose the model. The result shown by cross plot for the three Bourgne and Young, Bingham and proposed model. The Absolute Average Percentage Error (AAPE) and the Coefficient determination ( $R^2$ ) was done for the statistical error analysis.

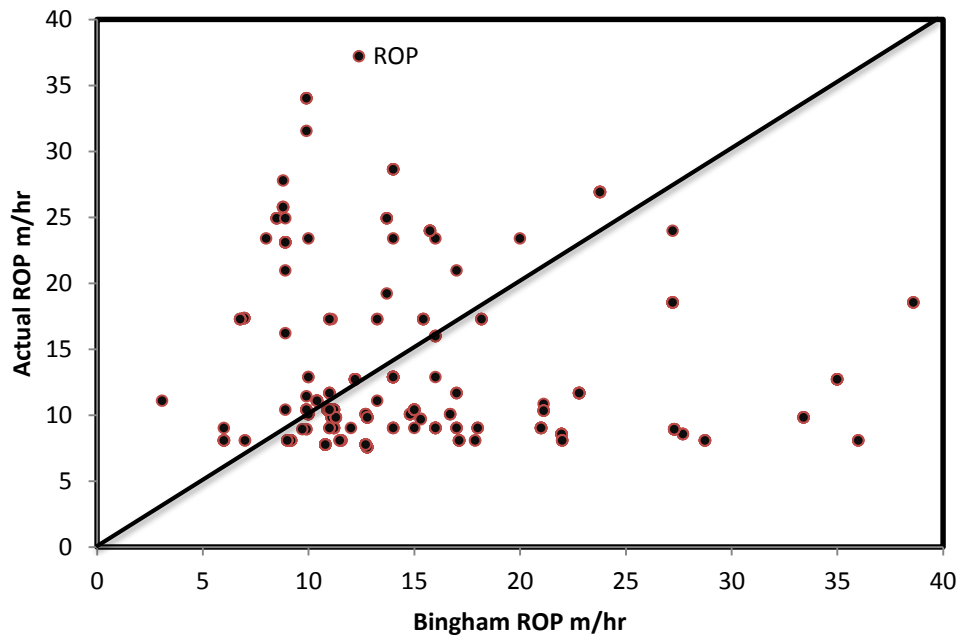


Figure 4.16 Cross Plot of Bingham Model vs Actual ROP

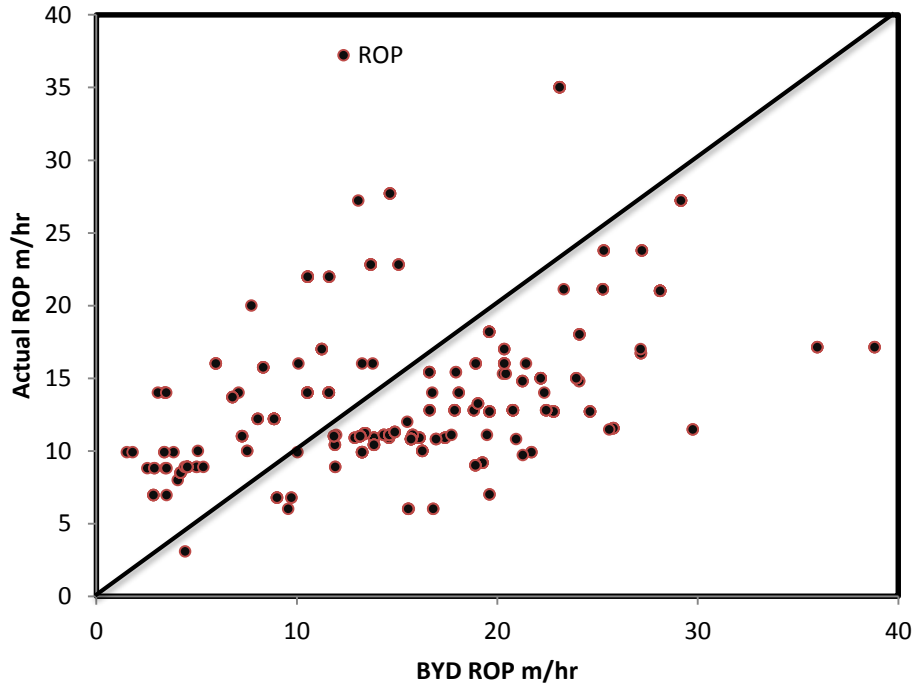


Figure 4.17 Cross Plot of BYD Model vs Actual ROP



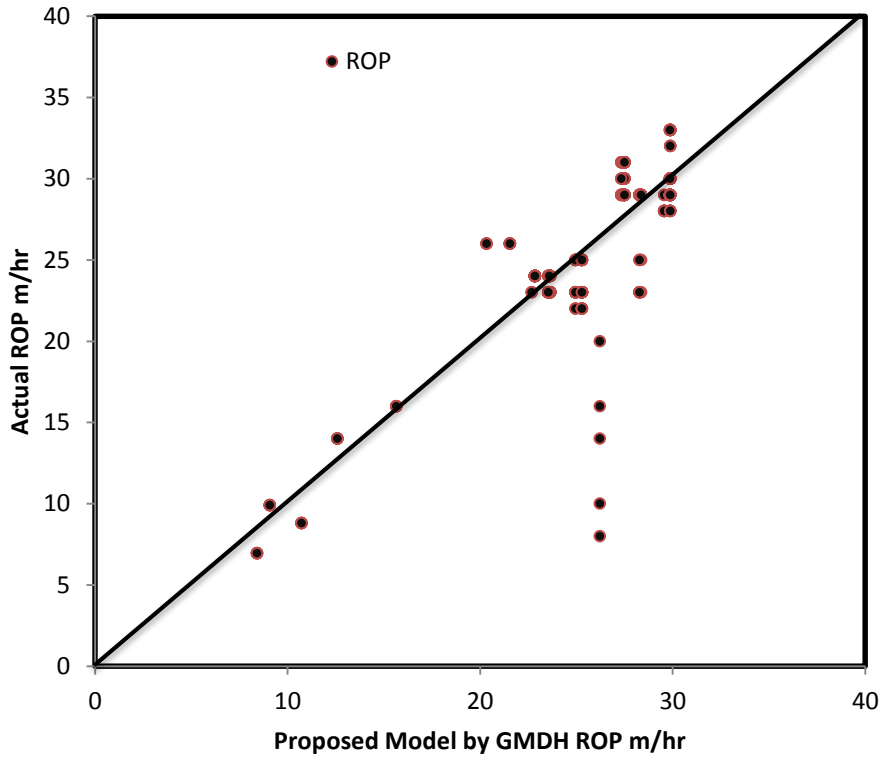


Figure 4.18 Cross Plot of Proposed Model by GMDH vs Actual ROP

The table 4.4 show the Statistical Parameters of all three models BYD, Bingham and Proposed model by GMDH.

**Table 4.4 Statistical Parameters Comparison**

Model	AAPE	( $R^2$ )
Bingham	43.3743	0.364
BYD	51.442	0.2600
GMDH	4.9129	0.9152

## CHAPTER (5)

### CONCLUSION AND RECOMMENDATION

#### 5.1 Conclusion:

Based on the result and discussion done in this study, the conclusion can be summaries on the on three main points, firstly, first stage optimization, secondly modeling and prediction, finally compare between the previous methods.

Firstly, the optimization was done using Drilling simulator (DrillSIM500) which is available in UTP, usually used for well control training. In this project the used for optimization and give good result. The result of the optimization is saving 13.69% of the Drilling Operation cost.

Secondly, modelling Rate of penetration (ROP) using GMDH techniques which shows high ability to find good relationship between drilling parameters and drilling rate of penetration with high accuracy. The GMDH achieved correlation coefficient of 91.52% and the average absolute percentage error (AAPE) of 4.9129%. The trend analysis confirm that the model is physically correct, because it match the previous model in Apply Drilling Engineering book which have typical respond of penetration rate to increase in weight on bit and rotary speed. In some cases, extremely weight on bit can decrease the penetration rate and this behavior called Floundering. Also, at high rotary speed sometimes respond poor penetration rate which can be justified attributed to less efficient bottomhole cleaning.

Finally, compare between Bourqne and Young's model, Bingham model and GMDH model. The correlation coefficient for BYD is 0.26, Bingham 0.364 and 0.9152. GMDH achieved good model can be used for prediction Drilling rate of penetration (ROP).

## 5.2 Recommendations:

Drilling simulator can be used to optimized the parameters which will lead to reduce the cost of the drilling, so, update the parameters range used in specific field by optimized parameters after doing simulation to the previous wells.

Group Method of Data Handling (GMDH) all researcher and development should be work to improve the code as well as the prediction process, because it will improve the accuracy of the proposed model. Since the GMDH is eliminate the parameters have no effect on the output. Use wide range of data sets.

## REFERENCES:

- Andre .L.S , Lianos .C, Coiho .L, "Comparing Artificial neural net work implementations for prediction and modleing of dynamical systems",vol.5,602-609,2012.
- Bahari M.A, Nejatimoharrami .F, Rajaei.R, " Forecasting the rate of penetration using Bourgoyne and young model Associated with Genetic Algorithm",1-5.
- 
- Bakhshi .J, Kesharazi .R,R , "Real-time prediction of rate of penetration during drilling operation in oil and gas wells", represented at the 46th ,US Rock Mechenics/ Geomechanics sympcsiam held in Chicago , USA, 1-7 , 24-27 June 2012.
- Bataee .M, Irawan.S , Kaunyab .M, " Artificial Neural network model for prediction of Drilling rate of penetration and optimization of parameters", vol.57 (2), 65-70, July 9,2013.
- Bataee .M, Kamyub .M, Ashena.R, "inverstigation of various (ROP) Models and optimization of drilling parameters for Ppc and roller - cone bits in shadegoan oil field", presented at the CPS/SPE international oil and gas conference and Exhibition in China-Beijin,1-10, 8-10 June 2010.
- Bielstein W.D, Cannon .G.E , "Factors affecting the rate of penetration of rock bits " ,61-78.
- Carlos M.C.Jacinto, P.J- Ereitas Filho,S.M. Nassar, M.Roisenberg, Diego.G. Rodrigues,Mariana D.C.lima,"Optimization models and prediction og drilling rate ROP for the brazilian pre-salt layer", Enrico Zio Piero Baraldi Ed.AIDIC servizi S.R.L, 2013 Pp .823-828.
- James.C, Howland III, S.Mark Voss,"Natural gas prediction using the group method of data Andling", college of Technical sciences ,Montana stat University -Northern , Havre , Montana , USA , 1-5.
- Lee B.y , Liu .H.S , Tarnng .Y.S, " Modeling and optimization of drilling process", vol.74, (149-157) , 22 October 1996 .
- Luzar.M, Witozak.M,"Agmoh toolbox for neural network-based modelling ", institute of control and computation engineering, university of Zielona Gora ul.Pogorna 50,65-246.
- Luzar .M , Witczak.M," Agmdh toolbox for neural network-based Modelling ", institute of control and computation Engineering , university of zielona Pogorna 50,65-246.

- Monazami .M, Hashemi . A,Shahbazian.M, " Drilling rate of penetration prediction using artificial Neural network = A case study of one of Iranian southern oil fields ", Ahwaz,Iran ,244-622,2012.
- Mostofi.M, Shabazi .K, Rahimzadeh .H, Rastegar.M, " Drilling optimization based on the ROP model in one of the Iranian oil fields ", presented at the CPS/SPE international oil and gas conference and Exhibition in china , Beijin,China ,8-18 June 2010.
- Popov Andrey ,"Genetic Algorithms for optimization -Application in controller design problems", programs for MATLAB,2003.
- Rahimzadeh .H, Mostofi .M, Hashemi .A, Salahshoor.K," Comparison of the penetration rate models using fird pata for one of the gas fields in persian galf area ", presented at the CPS/SPE international oil& gas conference and exhibition in China held , Beijing , China 8-10 June 2010.
- Rastegar .M , Hareland .G, Nygaard .R , Bashri .A,"Optimization of Multiple bit runs based on ROB models and cost equation= A new methodology applied for one of the persian gulf carbonate fields " , presented at the IAD/SPE Asia pacific drilling technonlogy conference and exhibition , Jakarta , Indonesia, 25-27 Aguast 2008.
- Samsudin.R, Saad Puteh,"A time series forecasting model using group method of data handling (GMDA)," Department of soft ware Engineering , University Technology of Malaysia , Neurocom puting ,70=2913-2923.

## APPENDICES

Country	Province	Latitude	Longitude	Block Name	Field Name	Contractor	Rig Name	Type of Rig	Daily Cost	\$53,038	AFE No.	AFE Cost
Sudan	S Kordofan	10°17'15" N	29°21'47" E	2B	Heglig	ZPEP	Rig 797	Drilling	Cum. Cost	\$928,506	142236	\$2,278,190
Ann. BOP Size	W.P.(psi)	Pipe Ram Size	W.P.(psi)	Blind Ram Size	W.P. (psi)	Section A Size	W.P. (psi)	RKB to GL	GL Elevation	Move Date	06-May-2014 00:01	
13 5/8	5000	5	5000	5	5000	13 5/8"	3000	7m	400m	Spud Date	18-May-2014 20:00	
ast Csg Size(in)	Csg Shoe	Float Collar	Cement Top	Last BOP test	Csg Test	Next Csg Size	Planned Depth	Kill Speed(SPM)	Kill Pres(psi)	TD Date		
9 5/8	558	545		05/22/2014	05/22/2014	7	1650			Release Date		
M. Depth mkb	TVD mkb	Last Depth	Progress m	Daily Hours	Daily ROP	Cum. Drlg Hrs	Well Avr. ROF	Authorized Depth	Survey Tool	Survey Depth	Inc.(deg.)	Azi.(deg.)
1650	1650	1516 m	134	12.4	10.8	51.50	17.50 m/h	1650 m	wire line	549 m	1	
Mud Type	Hole Loss	H. Progr. Loss	Surface Loss	S. Progr. Loss	Lag Check@	Lag Time (min)	Total Gas	Trip Gas	Bg Gas	Total Fuel	Fuel Used	Fuel Stock
KCL												93,000
Mud Properties	Agent Used	Daily	Acum.	Mud Pumps	Pump 1	Pump 2	Bit Description	Current Bit	Previous Bit	BHA No.	2	
Vud WT(ppg)	11.0	Bentonite		Make	3NB 1300C	3NB 1300 C	Run n.	2	1	BHA Name	Rotary BHA	
Viscosity(s)	49	Soda Ash		Type	TRIPLEX	TRIPLEX	Bit n.	2RR	1	BHA Elements	OD(in)	Length(m)
PV(mPa.S)	20	Caustic Soda		Strk.length mm	305.0	305.0	Size(in)	8 1/2	12.25	BIT	8.5	0.23
YP(Pa)	26	XCD		Liners (mm)	160	160	Type	MRS619KLAPX	GT 117	BS	6.5	0.9
Gels 10/10(Pa)	5/6	KCl		Stk per min	60	60	Nozzles	6*14	3*16,1*14	2 DC	6.5	18.07
API FL(ml)	4.8	Flowzan		Hydraulics	Current Bit	Previous Bit	Serial Numr	SC2250	70	STAB	8.375	1.83
HTHP FL(ml)		Barite		Bit Size(in)	8 1/2	12 1/4	Depth in (m)	558	90	8 DC	6.5	72.39
MBT(g/l)	2.5	Silicate		SPP	1720	790	Depth out (m)	558	558	JAR	6.5	5.66
PH	12	PAC LV		Flow (gal/min)	550	500	Mtrs.Drilled	1092	468	DC	6.5	9.02
Sand(%)	0.3	PAC -RL		A.V.(m/min)DP	105.00	42.00	Rot.Hours	73.6	32.6	HWDP	5	110.33
		QS-2 (M-C)		A.V.(m/min)DC	140.00	61.00	AVR ROP	14.8	14			
		PaBio 35%		TFA(sqin)	0.90	0.74	WOB (KDaN)	6-8	4-6			
		Sodium Bicar		Hydraulic HP	116	296.00	RPM	100	100			
		PAL CAL 130		Jet Vel (m/s)	60	92.00	IADC	PDC	117			
		AVR Cost	108.1	Bit P. Drop (psi)	363.00	726.00	Make	Smith	PDC			

### Sample of DDR

Serial No.	TDR 21057	Serial No.	Hours	Code	Sub	Phase Code	PT/NPT	Operations Descriptions	Import PM Report
0000	1000	10	RR	05	D08	NPT	Rig Zero rate:P/U Swivel & Kelly bar, install Kelly bushings, Connect Kelly		
1000	1045	3/4	RO	04	D08	PT	Rig back on day rate: Precautionary wash dn f/- 500m to FC AT 545.5m.		
1045	1430	3 3/4	DR	05	D08	PT	C/O F/C @ 545.5m and hard CMT f/- 456m to 558m Shoe.		
1430	2400	9 1/2	DR	01	D08	PT	<p>Drlg 8 1/2" main hole from 558m to 655M. Hole good. Pumping 30 bbl Hi-vis every 9 connections</p> <p>00.00 - 04.00: Drlg 8 1/2" main hole f/- 655m &gt; Rig has 10 hrs Zero rate.</p> <p>Last formation: Baraka shl: 980m .. Last sample = SS 60%.. CLY ST = 40% Last survey @ 549m = 1*</p> <p>Material Requirement New : Casing req. #3, GTS trailer with sides to back load old bits. Pending : 7" CSG &amp; equipment Received : Chemical Req #2. 29 x7" CSG</p> <p>NOTE: No TDS. Rig using Kelly, Day rate is changed as per contract.</p> <p>Petronas trainee: Ibrahim Shaikheldin Ibrahim Mohammed Arrive 05/20/2014 Last Inspection 5" DP : Feb 2014. Next Inspection: Aug 2014. POB= Nationals = 28. Expats = 22 Fresh Water Truck: GTS 7324 ----- Vacuum Tanker: GTS 7003</p>		

The author name at the DDR

## Appendices- GMDH code

```
clc;
% Final Year Project
% Ibrahim Shaikeldin
% Universiti Teknologi PETRONAS
%
% the aim is to clear all input and output from the Command Window
% display, giving you a "clean screen."
clf; % it deletes from the current figure all graphics objects
clear all;%Clears all variables and other classes of data too.
close all;% it force deletes all figures (hidden and non-hidden
strings)
tic;
%
% Step (1) Reading the input file
% =====
% Loads data and prepares it for a neural network.
ndata= xlsread('all_data.xls');
ndata= xlsread('ROP3.xlsx');
%50% of data will be used for training
%25% of data will be used for cross-validation
%25% of data will be used for testing
for i=1:257
    atr(i,:)=ndata(i,:);
end
for i=258:383
    aval(i-257,:)=ndata(i,:);
end
%
for i=384:length(ndata)
    atest(i-383,:)=ndata(i,:);
end
Ytr=atr(:,1);
Xtr=atr(:,2:6);
Xtst=atest(:,2:6);
Ytst=atest(:,1);
Yv=aval(:,1);
Xv=aval(:,2:6);
[model, time] = gmdhbuild(Xtr, Ytr, 2, 0, 6, 0, 2, 2, 0, Xv, Yv,1);
gmdheq(model, 8);
[Yqtst] = gmdhpredict(model, Xtst);
[Yqval] = gmdhpredict(model, Xv);
[Yqtr] = gmdhpredict(model, Xtr);
[MSE, RMSE, RRMSE, R2] = gmdhtest(model, Xtst, Ytst);

% Evaluating Relative Error for training set:
%=====
Et1=(Ytr-Yqtr)./Ytr*100;
[q,z] = size(Et1);
figure
plot(Ytr,Yqtr,'o')
grid off
set(gcf, 'color', 'white')
axis square

title('Predicted ROP vs Measured ROP');
xlabel('Measured ROP "m/hr"');
ylabel('Predicted ROP "m/hr"')
legend('Training set', 'location', 'Northwest')
```



```

% Adding Reference Line with 45 degree slope
line([0 ; 60],[0 ; 60])
%HINT: Select the y-value based on your data limits
hold
% Evaluating the correlation coefficient for training set:
% =====
Rt1=corrcoef(Yqtr,Ytr);
Rt11=min(Rt1(:,1));
gtext(['correlation coefficient = ( ' num2str(Rt11) ')']);
hold

% Adding Reference Line with 45 degree slope
%line([0 ; 60],[0 ; 60])
%HINT: Select the y-value based on your data limits

% Evaluating Relative Error for validation set:
% =====
Ev1=(Yv-Yqval)./Yv*100;
[m,n] = size(Ev1);
figure

plot(Yv,Yqval,'o')
grid off
set(gcf, 'color', 'white')
axis square
title('Predicted ROP vs. Measured ROP');
xlabel('Measured ROP "m/hr"');
ylabel('Predicted ROP "m/hr"');
legend('Validation set', 'location', 'Northwest')
% Adding Reference Line with 45 degree slope
line([0 ; 60],[0 ; 60])
%HINT: Select the y-value based on your data limits

% Evaluating the correlation coefficient for validation set:
% =====
% for the first target Pressure Drop
Rv1=corrcoef(Yqval,Yv);
Rv11=min(Rv1(:,1));
gtext(['correlation coefficient = ( ' num2str(Rv11) ')']);
hold

% Evaluating Relative Error for testing set:
% =====
% for the first target Pressure Drop
Ett1=(Ytst-Yqtst)./Ytst*100;
[m,n] = size(Ett1);
figure
%
plot(Ytst,Yqtst,'o')
grid off
set(gcf, 'color', 'white')
axis square

title('Predicted ROP vs.Measured ROP');
xlabel('Measured ROP "m/hr"');
ylabel('Predicted ROP "m/hr"');
legend('Testing set', 'location', 'Northwest')
% Adding Reference Line with 45 degree slope
line([0 ; 60],[0 ; 60])
%HINT: Select the y-value based on your data limits

```

```

% Evaluating the correlation coefficient for testing set:
% =====
Rtt1=corrcoef(Yqst,Ytst);
Rtt11=min(Rtt1(:,1));
gtext(['correlation coefficient = ( ' num2str(Rtt11) ' )']);
hold
% plotting the histogram of the errors for training set:
% =====
figure
histfit(Et1,10)
%hist(Et1,10)
h = findobj(gca,'Type','patch');
set(h,'FaceColor','w','EdgeColor','k')
title('Error Distribution for Training Set (Polynomial GMDH
Model)');
legend('Training set')
xlabel('Error');
ylabel('Frequency')
set(gcf, 'color', 'white')
hold

% plotting the histogram of the errors for validation set:
% =====
figure
histfit(Ev1,10)
%hist(Ev1,10)
h = findobj(gca, 'Type', 'patch');
set(h, 'FaceColor', 'w', 'EdgeColor', 'k')
title('Error Distribution for Validation Set (Polynomial GMDH
Model)');
legend('Validation set')
xlabel('Error');
ylabel('Frequency')
set(gcf, 'color', 'white')
hold

% plotting the histogram of the errors for testing set:
% =====
figure
histfit(Ett1,10)
%hist(Ett1,10)
h = findobj(gca,'Type','patch');
set(h,'FaceColor','w','EdgeColor','k')
title('Error Distribution for Testing Set (Polynomial GMDH Model)');
legend('Testing set')
xlabel('Error');
ylabel('Frequency')
set(gcf, 'color', 'white')
hold

% Estimating the residuals for training set:
% =====
figure
Errorrt1 = Yqtr-Ytr;
plot(Errorrt1,':ro');
grid off
set(gcf, 'color', 'white')
title('Residual Graph for Training Set (Polynomial GMDH Model)')
legend('Training Set')
xlabel('Data Point No')

```

```

ylabel('Errors')
hold
% Estimating the residuals for validation set:
% =====
figure
Errorv1 = Yqval-Yv;
plot(Errorv1, ':ro');
grid off
set(gcf, 'color', 'white')
title('Residual Graph for Validation Set (Polynomial GMDH Model)')
legend('Validation Set')
xlabel('Data Point No')
ylabel('Errors')
hold
% Estimating the residuals for testing set:
% =====
figure
Errortt1 = Yqtst-Ytst;
plot(Errortt1, ':ro');
grid off
set(gcf, 'color', 'white')
title('Residual Graph for Testing Set (Polynomial GMDH Model)')
legend('Testing Set')
xlabel('Data Point No')
ylabel('Errors')

% *****
% STATISTICAL ANALYSIS:
% *****
% Training set:
% =====
% Determining the Maximum Absolute Percent Relative Error
MaxErrt1 = max(abs(Et1));

% Evaluating the average error
Etavg1 = 1/q*sum(Et1);

% Evaluating the standard deviation
STDT1 = std(Errort1);

% Determining the Minimum Absolute Percent Relative Error
MinErrt1 = min(abs(Et1));

% Evaluating Average Absolute Percent Relative Error
% =====
AAPET1 = sum(abs(Et1))/q;

% Evaluating Average Percent Relative Error
% =====
APET1 = 1/q*sum(Et1);

% Evaluating Root Mean Square
% =====
RMSET1 = sqrt(sum(abs(Et1).^2)/q);

% Validation set:
% =====
% Determining the Maximum Absolute Percent Relative Error
MaxErrv1 = max(abs(Ev1));

```

```

% Determining the Minimum Absolute Percent Relative Error
MinErrv1 = min(abs(Ev1));

% Evaluating the average error
Evavg1 = 1/m*sum(Ev1);

% Evaluating the standard deviation
STDV1 = std(Errorv1);

%
% Evaluating Average Absolute Percent Relative Error
% =====
AAPEV1 = sum(abs(Ev1))/m;

% Evaluating Average Percent Relative Error
% =====
APEV1 = 1/m*sum(Ev1);

% Evaluating Root Mean Square
% =====
RMSEV1 = sqrt(sum(abs(Ev1).^2)/m);

% Testing set:
% =====
% Determining the Maximum Absolute Percent Relative Error
MaxErrtt1 = max(abs(Ett1));

% Determining the Minimum Absolute Percent Relative Error
MinErrtt1 = min(abs(Ett1));

% Evaluating the average error
Ettavg1 = 1/m*sum(Ett1);

% Evaluating the standard deviation
STDTT1 = std(Errorrtt1);

% Evaluating Average Absolute Percent Relative Error
% =====
AAPETT1 = sum(abs(Ett1))/m;

% Evaluating Average Percent Relative Error
% =====
APETT1 = 1/m*sum(Ett1);

% Evaluating Root Mean Square
% =====
RMSETT1 = sqrt(sum(abs(Ett1).^2)/m);

% =====
%-----
%-----
%Simulation: Variation ROTARY SPEED (N) while fixing the other
parameters
% % -----ROTARY SPEED (N) -----
%-----

```

```

ps1=[linspace(160,160,10); % DEPTH [min=0      max=1660  mean=830]
linspace(12.5,12.5,10);%BIT SIZE[min=8.5    max=24  mean=11.2061]
linspace(100,150,10);%ROTARY SPEED [min=100   max=150
mean=141.3294]
linspace(9.5,9.5,10);%WOB [min=3      max=15   mean=9.960317]
linspace(9.4,9.4,10)];%MUD WEIGHT[min=8.6   max=11.3
mean=10.31746]

% Now simulate
[Yq_ratio1]=gmdhpredict(model, ps1);
% Plot Figures for ROTARY SPEED (N) Variation
figure
px1=plot(ps1(:,3),Yq_ratio1(:,1),'-rs');
set(gca,'YGrid','off','XGrid','off')
set(gca,'FontSize',12,'LineWidth',2);
set(px1,'LineStyle','-','LineWidth',1.5,'Color','k','MarkerSize',6)
xlabel('ROTARY SPEED (N) (RPM)','FontSize',12)
ylabel('ROP (m/hr)', 'fontsize',12)
% =====

%Simulation: Variation WEIGHT ON BIT (WOB) while fixing the other
parameters
% % -----WEIGHT ON BIT (WOB) -----
-----

ps2=[linspace(160,160,10); % DEPTH [min=0      max=1660  mean=830]
linspace(12.5,12.5,10);%BIT SIZE[min=8.5    max=24  mean=11.2061]
linspace(135,135,10);%ROTARY SPEED [min=100   max=150
mean=141.3294]
linspace(3,15,10);%WOB [min=3      max=15   mean=9.960317]
linspace(9.4,9.4,10)];%MUD WEIGHT[min=8.6   max=11.3
mean=10.31746]

% Now simulate
[Yq_ratio4]=gmdhpredict(model, ps2);
% Plot Figures for WEIGHT ON BIT (WOB) Variation
figure
px2=plot(ps2(:,4),Yq_ratio4(:,1),'-rs');
set(gca,'YGrid','off','XGrid','off')
set(gca,'FontSize',12,'LineWidth',2);
set(px2,'LineStyle','-','LineWidth',1.5,'Color','k','MarkerSize',6)
xlabel('WEIGHT ON BIT (WOB) (TON)','FontSize',12)
ylabel('ROB (m/hr)', 'fontsize',12)

```

