A NEW CORRELATION FOR OIL FORMATION VOLUME FACTOR OF OIL AND GAS MIXTURE USING GROUP METHOD OF DATA HANDLING; AN EMPIRICAL APPROACH

AHMAD AZHARI ELHADI ELJEZOOLY

PETROLEUM ENGINEERING

UINVERSITY TEKNOLOGI PETRONAS

MAY 2014

CERTIFICATION OF APPROVAL

A new correlation for oil formation volume factor of oil and gas mixture using Group Method of Data Handling; an empirical approach

By:

Ahmad Azhari Elhadi Eljezooly (15790)

A project dissertation submitted to the
Petroleum Engineering Department
Universiti Teknologi PETRONAS (UTP)
In partial fulfillment requirement for the
Bachelor of Engineering (Hons.)
Petroleum Engineering

Approved

____________________________ (Dr. MOHAMMED ABDALLA AYOUB)

UNIVERSITI TEKNOLOGI PETRONAS

TRONOH, PERAK

May 2014

CERTIFICATION OF ORIGINALITY

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.

(Ahmad Azhari Elhadi Eljezooly)

Table of Contents

List of Table

List of Figures

ACKNOWLEDGEMENTS

First and foremost, I would like to thank God who through his guidance has helped me through my final year project.

Secondly, upon completing 28 weeks of my FYP1 and FYP2 in University Teknologi Petronas , I am indebted to all personnel in making this final year project a success project; especially all the staff in Petroleum Engineering department for their advice and support throughout my FYP period.

I would also like to take this exceptional opportunity to honorably thank my supervisor **Dr. Mohammed AbdAllah Ayoub** for his continuous support, guidance and supervision on many technical and theoretical issues concerning my final year project.

I would never forget to gives grateful thanks to my family who always stood beside me throughout this period, especially to my father **Eng. Azhari Elhadi** who always presented better advice and guides.

Finally, Thank you very much, may almighty God repay you for your kindness.

ABSTRACT

Numerous production engineering and reservoir analysis problems have been solved by estimating formation volume factor (B_o) for oil-gas mixture. It is considered as one of most important factor of physical properties of hydrocarbon that use in evaluation and enhancement of the reservoirs. In addition to, evaluation of formation volume factor for oil and gas mixture is considered as an important tool in any field project development as well as in reservoir performance evaluation, because both of reservoir engineering and production design operations usually evaluate the changes of the fluid properties, and one of these properties is the FVF.

This project aims to construct a new correlation that can estimate the formation volume factor (B_o) for oil and gas mixture with much more accuracy (less errors) than the current one utilized by the industry. In order to develop a new correlation for B_o for oil and gas mixture, Group Method of Data Handling (GMDH) algorithms has been used in this study to generate an optimum model. GMDH Approach will be utilized for the first time to predict this property.

Total of 268 data sets have been collected from different regions. In addition to, different statistical and graphical tools have been used to assess model accuracy after collecting the required data. The performance of GMDH model is compared against the best correlating adopted by the industry currently. Small range of absolute average relative errors (1.53%) has been obtained whereas the correlation coefficient has been calculated as 0.993. Moreover, the standard deviation for the new Bo model has been calculated as 0.0271% with 0.00229% of minimum error for this correlation. Trend analyses have confirmed that this new model for oil formation volume factor at bubble point pressure is physically correct.

CHAPTER 1

Introduction

1.1 Project Background

The simple definition of formation volume factor is the ratio of oil volume in its natural resource to the oil at surface condition. (Ahmad, 2000). From this definition, we need to convert measured surface volume to reservoir volume since most of produced oil and gas measurements are made at surface. B_o is always greater than one because produced oil usually contains dissolved gas.

When the crude oil being produced there will be a reduction in the oil volume that being produced. This phenomenon is known as shrinkage. The Shrinkage of produced crude can be estimated by using the formation volume factor Bo. This factor is considered one of the most important physical properties of the crude oil, because it is related directly to the calculation of the oil which called stock tank oil initial in place (STOIIP).

The mathematical expression of the oil formation volume factor can be shown as the following;

$$B_o = (V_o)_{p,t} / (V_o)_{sc} \quad \text{(Equation 1)}$$

$(V_o)_{p,t}$ = Oil at reservoir condition

$(V_o)_{sc}$ = oil at surface condition

The following graph explains the relationship of the oil formation volume factor with the pressure in the reservoir

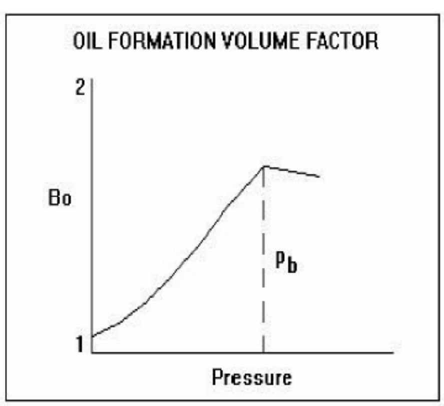


Figure 1: Oil Formation Volume Factor Curve, from petrowiki.org/Oil_formation_volume_factor

Due to expansion of the oil, the oil volume will increase due to reduction of the pressure below the initial reservoir pressure. As a result, the oil formation volume factor will increase. The phenomenon of increasing of the oil formation volume factor will be increased until reaching the bubble point pressure. The maximum value of the oil formation volume factor is obtained at bubble point pressure because the maximum expansion of the oil is reached at this point.

When the pressure is reduced below the bubble point pressure, there will be a reduction in oil volume. This phenomenon happens due to releasing of the solution gas from the oil being produced.

Several methods have been developed to estimate and evaluate oil formation volume factor (B_o) for oil and gas mixture.

Since the middle of the 1940's, majority of the researchers in US have presented the importance of using empirical approach to estimate the PVT properties. Several studies carried out in order to estimate those properties which led to enhance the researches in developing new correlations for PVT properties. Numerous studies for such kind of properties were estimated by Katz, Standing and Conquist. But unfortunately those correlations couldn't give high accuracy in estimating those PVT properties; because the experimental data to develop such correlations were difficult to be collected. Recent studies have been carried out by several researchers in the last thirty years in different regions all over the word. Among those researchers, Vazquez & Beggs, Glaso, Mohammed Al-Marhoun, Farshad & Leblane, and Abdul-Majeed & Salman.

By using United State mid-continent oil, Katz (1942) proposed a new graphical form for B_o. The main parameters that have been used in this study were specific gravity of oil and gas, oil gas ratio and pressure & temperature of targeted reservoir. The difficulty in this correlation was the requirement of using the calculations and graphs in order to get the final value for B_o.

Standing (1947) suggested a new graphical correlation including gas gravity, solubility of the gas, and oil specific gravity in addition to reservoir temperature as main parameters for his correlation. To develop this graphical correlation, Standing utilized more than 105 experimental data point from different California oil fields with 1.2% average error.

A new relationship for estimating formation volume factor by using specific gravity of oil, gas solubility, gas gravity, and temperature of the reservoir has been proposed by Vasquez and Beggs in 1980. By using the regression analysis for 6,000 measurements of B_o, Vazquez and Beggs have developed their own correlation for oil formation volume factor.

Glaso (1980) has developed a new correlation to estimate B_o from PVT analysis on 45 oil samples with reported average error -0.43% and standard deviation equal to 2.18%.

In Nigeria 1987, a total number of 503 data sets are collected from various reservoirs that located on the Niger Delta Basin were available for oil formation volume factor correlation by Obomanu and Okpobori. They utilized Al-Marhoun B_o correlation and then modified the B_o correlation form that was presented by Standing that time. As a result, a new correlation coefficient for oil formation volume factor is been developed for Nigerian crude oils.

Al-Marhoun (1988) suggested a new correlation for oil formation volume factor. Gas gravity, Gas solubility, oil-tank gravity and temperature were the main factors in his correlation. By using nonlinear-multiple regression analysis for more than 160 experimental sets of data point an empirical equation for B_o has been developed.

In the middle of 1988, Abdul-Majeed and Salman developed a correlation for oil formation volume factor. The correlation treated the B_o based on around 420 data points from general sources (unpublished sources). New correlation coefficient based on Al-Marhoun oil formation volume factor correlation was developed by their correlation. However, Alfattah and Al-Marhoun argued that a total number of 259 data points utilized by Abdul-Majeed and Salman were from Vazquez's work.

Asgapur, Cheun, wong and Mc Lauchlin (1989) developed another correlation for oil formation volume factor at and below bubble point pressure. This correlation carried out in western Canadian crude oils and gases at four reservoirs. Also this correlation based on Al-Marhoun bubble point pressure correlation in order to develop unprecedented correlation for B_o. Trend analysis were applied to check the correctness of the developed model, the new form showed less average error than that in Stading (1947) and Vazquez & Beggs (1980) correlations.

Labedi (1990) suggested one more correlation for oil formation volume factor from various reservoirs in Africa. 129 data sets used in this correlation, 97 data sets were form different reservoirs in Libya, 28 data sets were from Nigeria and just 4 data sets were from Angola. This correlation treated the B_o by using only the separator pressure & temperature and separation gas oil ratio due to the difficulty in measuring gas gravity (γ_g) and gas oil ratio (GOR) in previously mentioned oil fields.

In 1992, 51 data sets from UAE crudes have been used by Dokla and Osman in their developed another correlation for oil formation volume factor. Again they used Al-Mahroun correlation as a base to develop a new correlation coefficient from Al-Mahroun (1988) oil formation volume factor correlation.

In 1992, another correlation for oil formation volume factor is being estimated by Osorio, Garber, Leblance, and Farshad. Their main correlation feature was using surface separator stages as a criterion for developing new correlation for B_o. Also solution gas oil ratio was included in this correlation. About 98 reservoir samples from Colombian reservoirs were collected to contribute in developing of this correlation. New calculated B_o coefficient has been developed based on Standing (1947) and Glaso (1980) B_o correlations forms.

90 data sets have been collected from various reservoirs from Suez Gulf by Macary and El-Batanoney (1992) in order to develop a correlation for oil formation volume factor. To check the accuracy of the new model, this model was tested against another correlations form Egyptian data sets. Clearly the new model showed better improvement over tested one.

Based on Standing (1947) work, Omar and Todd (1993) proposed a new relation for oil formation volume factor. In order to construct the new model for B_o, 93 data sets have been collected from various reservoirs in Malaysia. The correlation treated the B_o as a function of gas gravity, oil gravity, gas oil ratio, and reservoir temperature. Using the new B_o correlation in bubble point prediction was the most valuable feature in this correlation.

In 1993, Petrosky and Farshad produced a new correlation for oil formation volume factor. By using 90 data sets from Gulf of Mexico, Standing B_o correlation was the basis for producing the new correlation coefficient. In order to test the validity and correctness of the produced model, nonlinear regression analysis were used along with maximum flexibility of the data to reach best result using available data sets.

Kartoatmodjo and Schmidt (1994) presented another correlation for oil formation volume factor by using total number of 5392 data sets for different global locations worldwide from various crude oil reservoirs. So from this point we can conclude that, Al-Marhoun B_o correlation and this correlation were the only correlations that used the different data sets from various sources all over the world. Vazquez & Beggs (1980) B_o were taken as a basis for the new developed correlation. In addition to, Petrosky and Farshad's approach was considered as the main approach for the new oil formation volume factor to provide maximum flexibility; hence best empirical model is been produced.

In 1997, a number of 62 data points collected from several UAE's oil fields were used by Almehaideb in order to produce new correlation for oil formation volume factor. Different parameters were considered to develop this correlation such as gas gravity, oil gravity, gas oil ratio and temperature of the reservoir. The main feature of this correlation was the improvement in evaluating B_o over existing correlations that time.

1.2 Group methods of data handling (GMDH) Algorithms

In this project, Group Method of Data Handling (GMDH) algorithm has been used to develop a new correlation for formation volume factor for oil and gas mixture. This method was developed by Prof. Alexey G. Ivakhnenko (1968) in the Institute of Cybernetics in Kiev (Ukraine). As was shown from different studies, this method was a computer-based method since a set of computer programs and algorithm were used to develop this approach with theoretical principles. Hereby, it's mostly appreciated that author for gave this opportunity to use this method and open the code to develop a new B_o correlation. This method has been quickly settled in various scientific laboratories worldwide.

The basic GMDH algorithm is a procedure of constructing a high-order polynomial of the form:

$$y = a + \sum_{i=1}^{m} b_i\, x_i + \sum_{i=1}^{m}\sum_{j=1}^{m} c_{ij}\, x_i x_j + \sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m} d\, x_i x_j x_k + \sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m}\sum_{l=1}^{m} e\, x_i x_j x_k x_l + \cdots$$

the important features of using GMDH in this project are, first by using this approach the analysis range of the new results will be minimized and hence save a lot of time. Secondly, by using the new approach of GMDH, it will ensure the generated model will not be affected by biases of humans as well as the misjudgments. GMDH approach includes automatically selections of influential input data (Parameters that have been mentioned above).

Majority of the previous study were experimental methods thus the new model of GMDH can be an alternative to that studies with much accurate results in acceptable and short time. In addition to, this new model aids in overcoming of several limitation that have been associated with the previous study as well as using less parameters as possible which was not been used by the previous studies and this because the GMDH model is developed by using self organizing approaches. By using this technique, GMDH can guarantee optimum model since well-proven optimization criteria is being used in this model.

This model is self-organized mainly there are network size, connectivity, element types and coefficient. The most valuable benefits from using the GMDH approach is the reduction of the modeling effects, therefore the biases misjudgments of humans can be avoided. Mathematically, GMDH approach is developed by a polynomial form to establish a relationship between the input data that are selected by a user and the output parameter that wants to be much accurate and acceptable in any specific field.

GMDH approach can provide valuable techniques such as capability of explanation and capability of comparing the obtained results by using data-based machine with empirical model principles. Mathematical approaches have been used in GMDH model in order to evolve a new correlation for formation volume factor for oil and gas mixture. The main reason to construct this new model is an attempt to develop a new correlation with much accuracy than the previous correlations (Experimental correlations or Empirical correlations).

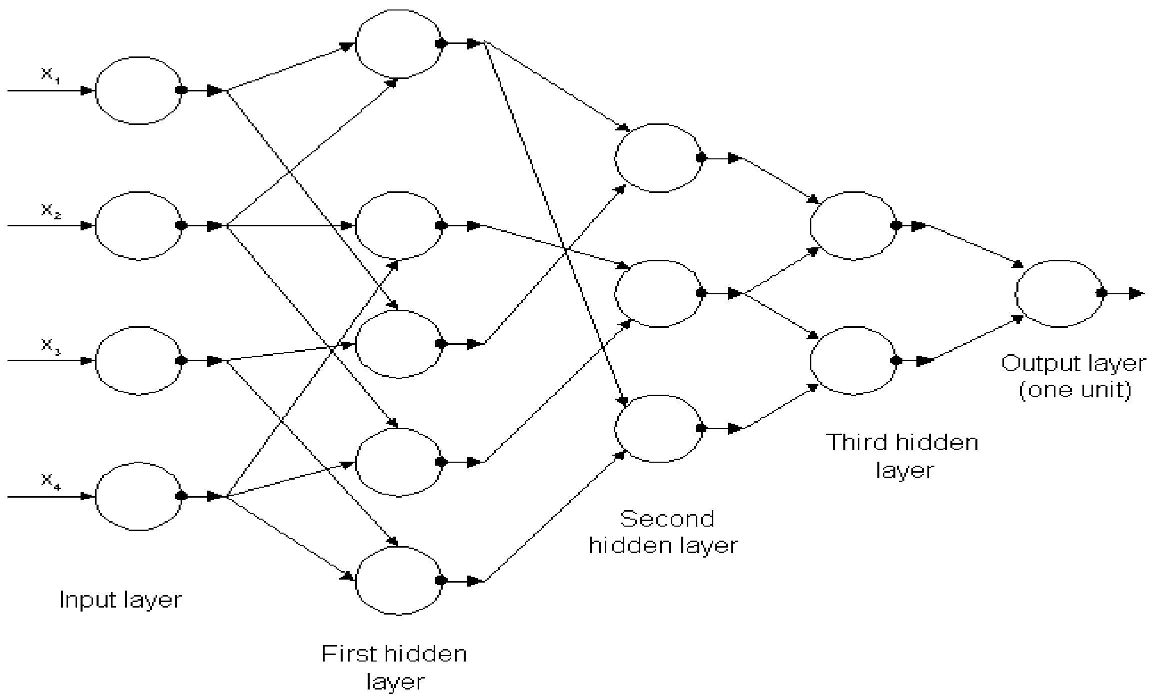


Figure 2: GMDH Network, by Ivan Galkin, U.Mass Lowell (Materials for UML 91.550 Data Mining course)

1.3 Problem statement

Many methods have been proposed to estimate the formation volume factor for oil and gas mixture in the most recent decades. The accuracy of estimating formation volume factor has been discussed frequently by many researchers. However, many correlations have been developed but they still can't be risen to a level of the most accurate correlation.

As measuring the formation volume factor experimentally is not an optimal option due to a high cost and because of difficulty in finding enough experimental data sets. In addition to, the variation of well conditions from one to another is an obstacle to have general correlation with acceptable range of error. As a result, the empirical approach is the most widely use nowadays in order to develop a new correlation for such correlations. morover, GMDH approach will be a new approach to estimate a new correlation for formation volume factor for oil and gas mixture with expected much accurate results than the previously published correlations.

The parameters affecting the formation volume factor for oil and gas mixture such as specific gravity of the oil, gas gravity and temperature are very important for the model generation.

1.4 Project Objectives

This project aims at developing a new correlation for formation volume factor (B_o) for oil and gas mixture as an empirical approach, this correlation has been validated and tested for oil formation volume factor at bubble point pressure due to difficulty in finding the required data to develop correlations for B_o at below and above bubble point pressure. The new correlation accuracy has been compared against the previously produced correlations namely Al-Marhoun, Standing and Alshammasi four and three paramaeters B_o corellations. Therefore the main objectives of this project can be stated as following:

I. Revising the best available correlations and defining their parameters.
II. Understand the factors influence the formation volume factor for oil and gas mixture.
III. Generate a new model using Group Methods of Data Handling (GMDH) to estimate the B_o.
IV. Investigate the effect of reducing the curse of dimensionality (reducing the number of correlation parameters).
V. Evaluate the model performance by comparing the gained outcomes against the published ones.

1.5 The Relevancy of the project:

Since this project aims at solving numerous reservoir performance and production problems, it is quite important to develop such a project in order to enhance oil wells performance. In addition to, Formation volume factor for oil and gas mixture is considered as one of the most important physical properties of the reservoir. More accurate correlations for PVT properties will give high estimation of different parameters in the well as well in the reservoir.

1.6 Feasibility of the study

To develop a new correlation for formation volume factor for oil and gas mixture, this project requires modeling software in order to conduct a successful study. By using Group Method of Data Handling - using MATLAB Software which is available in UTP, new model for B_o at bubble point pressure has been successfully obtained.

Another main part of this study is the field data; the field data are needed for this project and can be collected either from published paper or requested from Oil & Gas Industry. Hence, the project is clearly feasible to be implemented and results have been successfully obtained within the proposed time frame of the project.

Chapter 2

LITERATURE REVIEW

To get best reservoir performance and solve reservoir- performance problems, physical properties of the reservoir fluids should be well known and this could be determined in the laboratory from different samples usually collected from bottom-hole or by doing proper recoinbnation of surface tap samples and this in most of the cases that the results of the laboratory tests are not available, Standing (1947).

In addition to, evaluation of these physical properties of oil-gas mixtures considered as a most important part in any field development plan because both of reservoir engineering and production design operations usually evaluate the changes of temperature and pressure which directly affect the fluid properties, Vazquez (1980).

Different approaches have been proposed to evaluate formation volume factor at, above and below bubble-point pressure of oil-gas mixture which was considered as important tool in reservoir performance calculations in order to design the best various stages of development and production operation. Various studies have been issued to evaluate B_o at, above and below bubble-point pressure either by using large base of laboratory measured PVT data or by using empirical correlations to replace those commonly used. The used correlations of B_o have been developed many years ago from different field all over the world.

Standing (1947) suggested a new correlation for formation volumes by using PVT correlation for mixtures of California Oil and Gases. Formation volume factor of saturated liquid is a required factor in reservoir calculations because it used in computing the shrinkage of the oil produced from the reservoir when it passes to the surface (stock tank). Based on Katz's correlation of California crudes, Standing used a total of 105 bubble-point liquids in his study in terms of oil gravity (Υ_o), gas gravity (Υ_g), gas-oil ratio (GOR) and temperature of the targeted reservoir to come up with following correlating equation:

$$V_b = \phi\left\{GOR\left(\frac{\gamma_g}{\gamma_o}\right)^{0.5} + 125T\right\} \text{……………………….(Equation 2)}$$

Where

Vb = formation volume of bubble-point liquid, bbl

GOR = gas-oil ratio, cu ft per bbl.

γ_g = gravity of dissolved gas (air = 1).

γ_g = specific gravity of tank oil at 60 deg F.

T = temperature, deg F

Because of bubble-point pressure not common to have gas-oil ratios in excess 2000 cu ft per barrel, Standing develop a new equation for formation-volume of gas plus liquid phases, a correlated equation as following;

$$Vf = \phi\left\{(P)\left[(GOR)\frac{(T)^{0.5}}{(\gamma_g)^{0.5}}(\gamma_o)^{29\times10^{-0.00027GOR}}\right]\right\} \text{…………………… (Equation 3)}$$

Where,

VF = formation volume of gas plus liquid phases, bbl

Per bbl of tank oil

P = pressure, psi, absolute

GOR = gas-oil ratio, cu ft per bbl

T = temperature, deg F

γ_g = gas gravity (air = 1)

γ_o = specific gravity of tank oil at 60 deg F.

Due to importance of formation volume factor of hydrocarbon mixtures in reservoir performance, researchers have developed many experimentally and empirical correlation for formation volume factor of oil-gas mixtures. Total of 600 laboratory PVT analysis from different field all over the word discussed by Vazquez in 1980. The study discussed oil FVF's as empirical function and the factors that being considered in that study were GOR, oil gravity, gas gravity pressure and lastly temperature. Although Vazquez developed his correlation of B_o below and above bubble-point pressure from limited data, his method was most widely used in petroleum industry because he managed to use regression analysis techniques to correlate the laboratory data.

Vazquez (1980) concluded that for B_o below P_b (bubble-point pressure) the following equation was found to represent the measured laboratory data as a function of dissolved gas, oil gravity, gas gravity and temperature of specific reservoir.

$$B_o = 1 + C_1 R_s + C_2 (T-60)\left(\frac{\gamma_o}{\gamma_{gs}}\right) + C_3 R_s (T-60)\left(\frac{\gamma_o}{\gamma_{gs}}\right) \text{……… (Equation 4)}$$

The value of the coefficient depend on oil gravity and given by the following;

Table no (1): oil gravity coefficients

Coefficient	**$\gamma_o < 30$**	**$\gamma_o > 30$**
C_1	4.677×10^{-4}	4.670×10^{-4}
C_2	1.751×10^{-5}	1.100×10^{-5}
C_3	-1.811×10^{-8}	1.337×10^{-9}

Due to the change in the volume of unsaturated liquids above bubble point, isothermal compressibility usually affects B_o, (Vazquez ,1980).

The following equation stated by (Vazquez,1980) for Formation volume factor above bubble-point pressure,

$$B_o = B_{ob} \exp\left[C_o\left(P - P_b\right)\right] \quad \text{.......................................} \text{(Equation 5)}$$

Vazquez utilized more than 4,036 data point in a linear regression model to develop a correlation for compressibility used in the previous equation as following;

$$C_o = \frac{\left(a_1 + a_2 R_s + a_3 T + a_4 \gamma_{gs} + a_5 \gamma_o\right)}{a_6 P}$$

Where

a_1 = -1433,

a_2 = 5,

a_3 = 17.2,

$_4$ = -1180,

a_5 = 12.61, and

$a_6 = 10^5$

Based on Standing's B_o Correlation for California crude oil, Glaso presented a new correlation for B_o based on the point of not using other field data all over the word in Standing's correlation which was considered as the most widely used at that time (Glaso, 1988). PVT relations were developed from different fields for oils, and the main differences from Standing's work could be summarized in these two following factors:

1) different paraffinicty for crude oils from different oil field.

2) Considerable amount of nonhydrocarbon could be existed in surface gases from various reservoirs.

Generalized PVT correlations for B_o at and below bubble-point pressure were being developed by Glaso by considering variation in laboratory data from North Sea oils.

$$B_{ob}=\phi_2\left\{R\left(\frac{\gamma'_g}{\gamma'_o}\right)^{0.526}+0.968T\right\} \quad \text{.....................................(Equation 6)}$$

B_{ob}= oil formation volume factor at bubble-point (saturation) pressure, RB/STB(res m^3/stock-tank m^3).

$$B_t=\phi_3\left\{R\left[\frac{T^{0.5}}{\gamma'^{0.3}_g}\times\gamma_o^{2.9\times10^{-0.00027\times R}}\times P^{-1.1089}\right]\right\} \quad \text{....................(Equation 7)}$$

B^*_{ob} = correlating coefficient in order to calculate **B_{ob}.**

B_t = total oil formation volume factor below saturation pressure, RB/ STB (res m^3/stock-tank m^3).

Based on these correlations, Glaso stated that the shrinkage is the main phenomenon when the oil produced from the reservoir and hence B_o at saturation (bubble-point) used to evaluate that. While more than 3000 scf/STB of gas-oil ratio is produced when the oils at bubble point pressure. Regression analysis were used in this study to be more accurate in addition to the constants were determined to be as a= 0.526 and b= 0.968.

Al-Marhoun (1988) proposed a new correlation to determine formation volume factor at, below and above bubble-point pressure. Oil gravity, gas solubility, gas gravity and temperature were the main factors that included in Al-Mahroun's correlation. He used nonlinear multiple regression analysis with a view to evolve an empirical equation by using around 160 experimental data sets exclusively from Middle Eastern oil fields. Al-Mahroun proposed the following equation:

$$B_o=0.497069+0.862963\times10^{-3}T+0.182594\times10^{-2}F+0.318099\times10^{-5}F \quad \text{..(Equation 8)}$$

Where the parameter F defined as;

$$F = R_s^a \gamma_g^b \gamma_o^c$$

And the values of the coefficient a,b and c are as follows;

a = 0.742390

b = 0.323294

c = -1.202040

In 1992 Al-Mahroun developed a new correlation to estimate formation volume factors at, below and above bubble point for oil-gas mixtures. Empirical equations were develpoed by Al-Mahroun as functions of gas relative densities, oil relative densities, solution gas-oil ratios, temperature and reservoir pressure. A total of 11,728 experimental data point collected from various fields all over the world to obtain B_o at, below and above bubble-point pressure.

Based on Standing (1947), Vazquez and Beggs (1980), Glaso (1980) and Al-Marhoun (1988), Al-Marhoun (1992) used around 700 bottom-hole fluid samples from all over the world to develop his analysis. However, the majority of these samples were from Middle Eastern and North America regions. To enhance the accuracy of this study, least square methods and satistical analysises were used to develop a new correlation of B_o as follows:

B_o at bubble-point pressure has been correalted as a function of gas relative density, oil relative density, dissolved gas, temperature and pressure as following;

$$B_{ob} = f\left(R_s, \gamma_g, \gamma_o, T\right) \text{..........................(Equation 9)}$$

Where

R_s= gas oil ratio of the solution, SCF/STB

γ_g = relative density of the gas, (air=1)

γ_0 = relative density of the oil, (water = 1) and

T = temperature, ^{o}F

To reduce the deviation in measured date, least square linear regression used in previous equation and the following equation obtained to be the best form;

$$B_{ob} = 1 + a_1 R_s + a_2 R_s \left(\frac{\gamma_g}{\gamma_o} \right) + a_3 R_s (T - 60)(1 - \gamma_o) + a_4 (T - 60) \text{........(Equation 10)}$$

Where

$a_1 = 0.177342 * 10^{-3}$

$a_2 = 0.220163 * 10^{-3}$

$a_3 = 4.2925580 * 10^{-6}$

$a_4 = 0.528707 * 10^{-3}$

B_o on top of bubble-point pressure is expressed by Al-Marhoun in the following equation;

$$B_o = B_{ob} \left(\frac{P}{P_b} \right) \text{...(Equation 11)}$$

Where

Bo= oil formation volume factor above bubble-point pressure, RB/STB.

P= pressure. psia.

Pb= bubble-point pressure, psia.

For oil formation volume factor below bubble-point pressure can be expressed by the following form;

$B_t = B_{ob}(P/P_b)^d$

Where

B_t = total Formation volume factor below bubble-point pressure, RB/ STB.

$$d = a_9 (T + 460) + a_{10} \ln \gamma_g + a_{11} \gamma_o + a_{12} \ln \left(\frac{P}{P_b} \right) + a_{13} \left(\frac{P}{P_b} \right) + a_{14} \ln \gamma_o \text{..........(Equation 12)}$$

Where

$a_9 = -0.35279600 * 10^{-3}$
$a_{10} = -0.35328914$
$a_{11} = -0.24964270$
$a_{12} = 0.08685097$
$a_{13} = 0.36432305$
$a_{14} = 1.64925964$

Petrosky and Farshad (1993) developed new correlation for formation volume factor at bubble point pressure by using PVT correlation. They used SAS® software with nonlinear multiple regression analysis for 81 laboratory PVT analyses to develop their B_o correlation. The data required to develop their correlation were collected from Gulf of Mexico crude oil with applying two stages laboratory separator tests in order to construct their model. As a result, new correlation for formation volume factor at bubble point pressure was obtained with much accurate results than the published correlations. In this study both of the average absolute error and standard deviation were reduced, the average relative error was -0.01% and the absolute error was 0.64% with corresponding standard deviation of 0.86% and 0.58% which is much accurate than the previous studies.

Two correlations for formation volume factor have been developed by Alshammasi (1999). By using different parameters, Alshammasi published his paper for Bo. His first correlation was including solution gas solubility (Rs), specific gravity of the oil, specific gravity of the gas and the temperature of the reservoir. The correlation coefficient of this correlation was 0.9987 with 17.85% as average absolute error. By excluding the gas gravity to reduce the parameters from 4 to 3 parameters, Alshammasi developed his second correlation for formation volume factor. With reducing the parameters, the new average absolute error for his correlation was 19.86%.

$$Bo = 1+5.53*10^{-7} (Rs*(T-60)) + 0.000181*(Rs/\Upsilon o) + 0.000449*(T-60)/\Upsilon o) + 0.000206*(Rs*\Upsilon g/\Upsilon o) \quad \text{(Equation 13)}$$

Al-Shammasi formation volume factor equation with four variables

$$Bo = 1+0.000412*(Rs/\Upsilon o) + 0.000650*((T-60)/\Upsilon o) \quad \text{(Equation 14)}$$

Al-Shammasi formation volume factor equation with three variables

CHAPTER 3

Methodology

3.1 Research methodology

The qualitative method has been used in this project to generate the desired outputs. Intensive research has been done to prepare a high quality literature review in order to assist the author in analyzing the data and eventually in the main part which is the discussion part. Several published papers have been studied so as to prepare a wide set of information. As a result, the author found much ease in analyzing the data and compares the developed correlation with previous published correlations. Therefore, newly developed correlation for formation volume factor at bubble point pressure has been developed.

The data gathering process was the major challenge part in this project, because it is quite difficult to find enough data for PVT at below and above bubble point pressure. The GMDH approach forms the basis of this study, in which may depend on the long term accuracy and quality of the data selected. Through the MATLAB software, selected data has been calibrated with regression analysis method.

In order to solve any engineering problem, one of the following approaches must be used. Which can be classified as;

- Exact or rigorous approach.
- Modelling approach.
- Mechanistic approach.
- Experimental approach.

GMDH approach which has been used in this study is classified as “modelling approach”. This approach has not been used in the literature for estimating formation volume factor for oil and gas mixture. Standing, Al-Marhoun and Alshamassi correlations have been chosen to be the main correlations to be compared against this study, because until now these correaltions considered as the most accurate correlations with lowest errors and standard deviations. In addition to, they obtained the highest accuracy of correlation coefficient compared to other correlations.

The following flow chart summarizes the main method that has been used in this project;

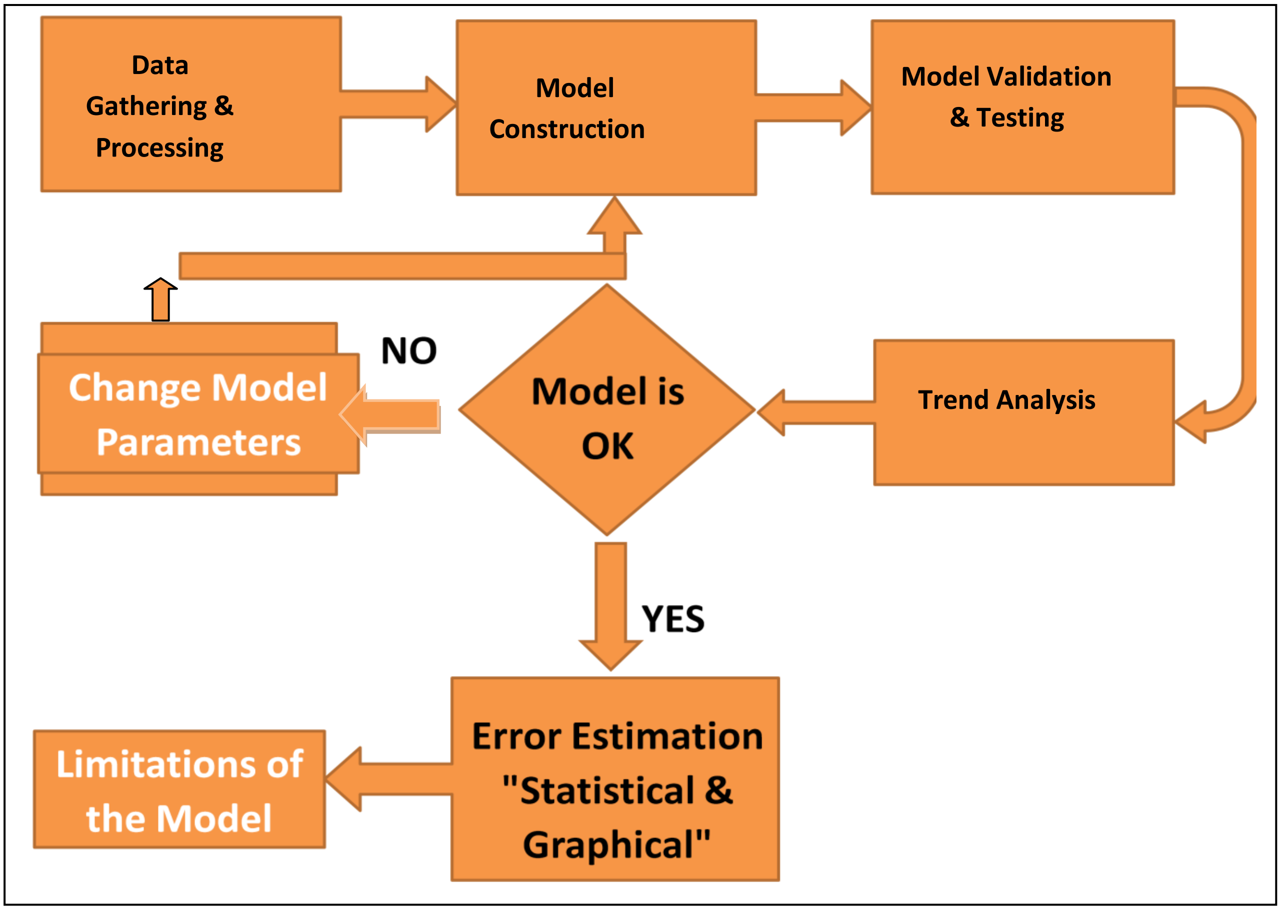


Figure no (3): Methodology flow chart

3.2 GMDH Model

First Step: Gathering of the required data

In order to develop an optimum model to correlate formation volume factor of oil and gas mixture, appropriate data should be collected and tested. These data can be collected either from published papers by doing an intensive research or from oil and gas industry. In this project data from a published resource has been used. Not any kind of data can be used in generating a new model unless it meets the requirements of the needed data.

The required data which will be the inputs should be well known in order to generate the coveted outputs. Furthermore, the number of data that required in building the new model should be big enough so as to best and acceptable results can be obtained. As a result, an improvement in the previous correlations can be achieved.

Second Step: Preprocessing of the data

One of the most important steps in developing the new model is by accurately cleaning and integrates the inputs gathered so that the objectives of the study can be obtained successfully. Usually there are two main stages in this step which are database consolidate and data filtration.

In the first step which is database consolidate, the data that has been collected in the first step will be tabulated so the inputs will be arranged, but not dispersed. In order for the author to discover any discrepancies, anomalies, repetitive, or any missing entries of the inputs, organization of the data is required.

The second step is to filtrate the data which is usually done to take away the inputs outliers, every extraordinary distribution and other defects within the inputs. Another objective of this step is to find invisible correlations within the inputs and choosing the best inputs to be developed, which indicate that this filtration of the data is not just about removing bad data. Moreover, this step aims at interpolating the missing inputs and selecting the most accurate columns to be analyzed after developing the new model.

Third step: Data Handling

This step is just about dividing the inputs into 3 different sets namely; training set, validation set and test set. The main function of data handling is to measure the degree of accuracy of produced outputs which in this case is the new correlation for formation volume factor for oil and gas mixture.

The first set is the training set which cares about the inputs and outputs. Group Method of Data Handling (GMDH) approach is used with this step to train a knowledge database. In this step, quite big number of data is required in order to ensure an optimum model will be successfully produced, thus as the number of the data increase much accurate results can be generated.

The second set is the validation set which also forms the basis of the developed model. During the training process the validation process is applied to check and test the model performance in terms of its accuracy, sensitivity. The validation set is to some extent a model checker to guarantee balance while developing the new model.

The last set is an independent set which is the test set. However, this procedure follows same probability distribution of training inputs. Focusing more in the test set, the author can test the final performance of the developed model. The more precisely the input data, the better outputs expected.

2:1:1 ratio has been chosen to be used in this project. Which means half of the data will be utilized in training set, ¼ for validation and the rests of the data will be utilized in the test set.

Fourth Step: Model Development

In order to develop a new correlation for formation volume factor for oil and gas mixture, software is required to build the new model. MATLAB software has been chosen as the main software in building the model, because it gives high range of flexibility comparing to others software. In addition to, MATLAB software makes the graphs more visualize. In this study, three sets of data has been used therefore MATLAB software is more perfect to be used in this study in terms of performance analysis comparing to other software. To meet the essential objective of this project, a code has been developed so as to provide the training, validation and data sets of the developed model.

Fifth step: Checking the Performance of developed model

Graphical tool aids is used to represent the graphical error analysis as well as testing the accuracy and the performance of the new correlation. The cross plot technique would be the main technique for this study.

3.3 Project activities

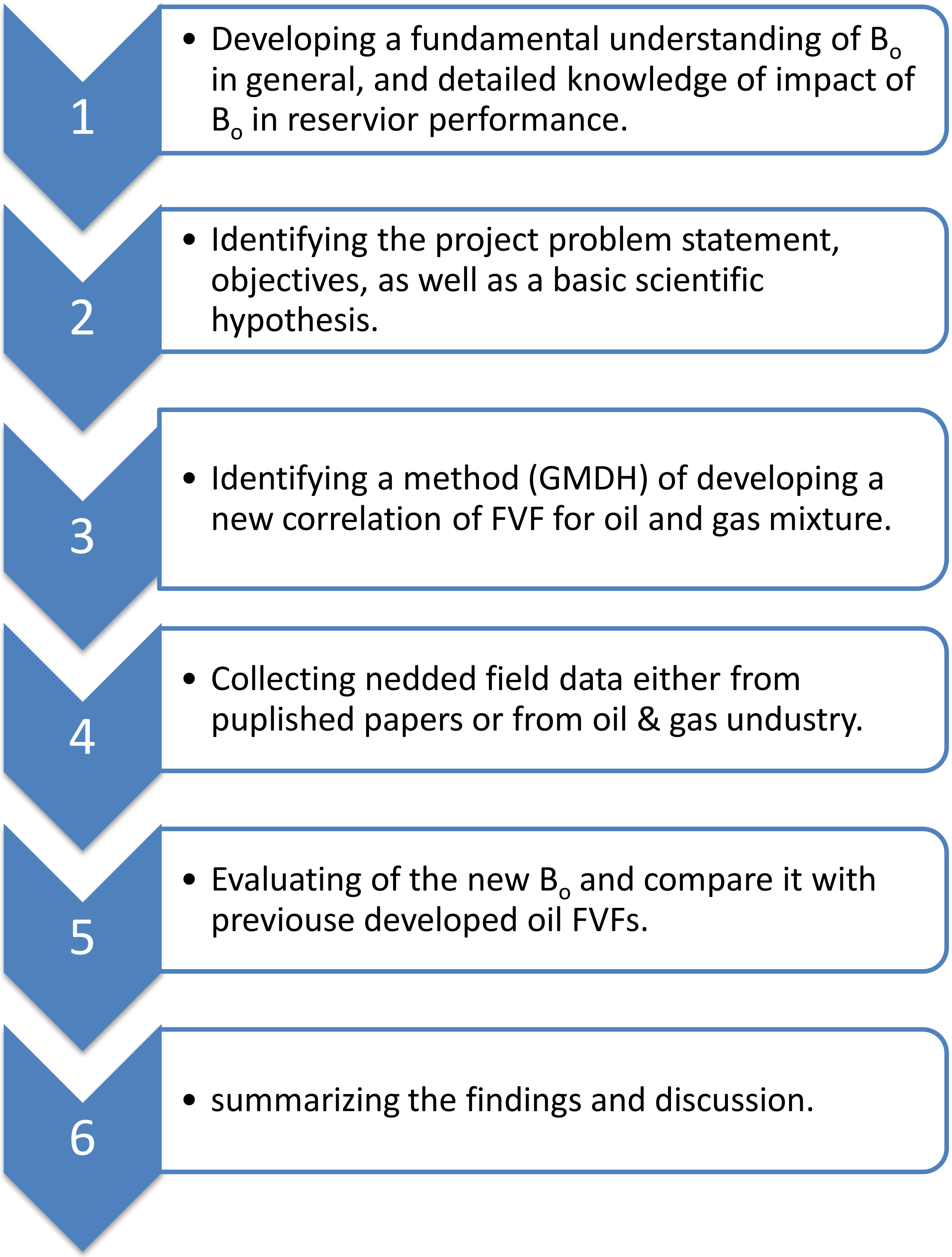
Developing an unprecedented correlation for formation volume factor for oil and gas mixture is the main activity of this project. By using Group Method of Data Handling (GMDH) method and MATLAB software the new model for formation volume factor at bubble point pressure has been successfully produced. Regression analysis techniques also have been used to compare the obtained results against published one after collecting the required data from a published paper.

3.4 Key Milestone

Table (2):Key Milestone

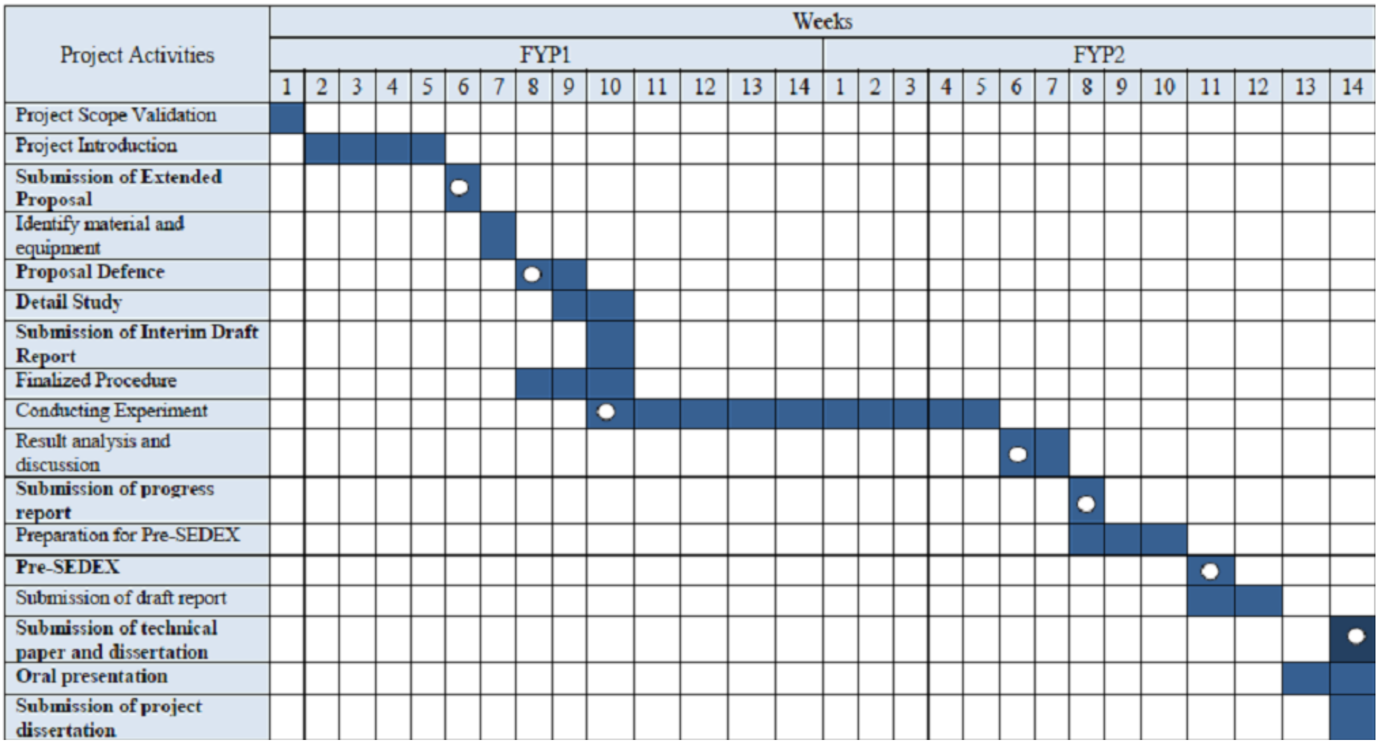
No	Activities	Date
1	Submission of Progress Report	Week 8
2	Poster Presentation (Pre-SEDEX)	Week 10
3	Submission of final report draft & Technical paper	Week 13
4	Oral Presentation (VIVA)	Week 15
5	Submission of Report Dissertation	Week 16

3.5 Methodology workflow



3.6 Gantt chart:

Table (3): Gantt chart

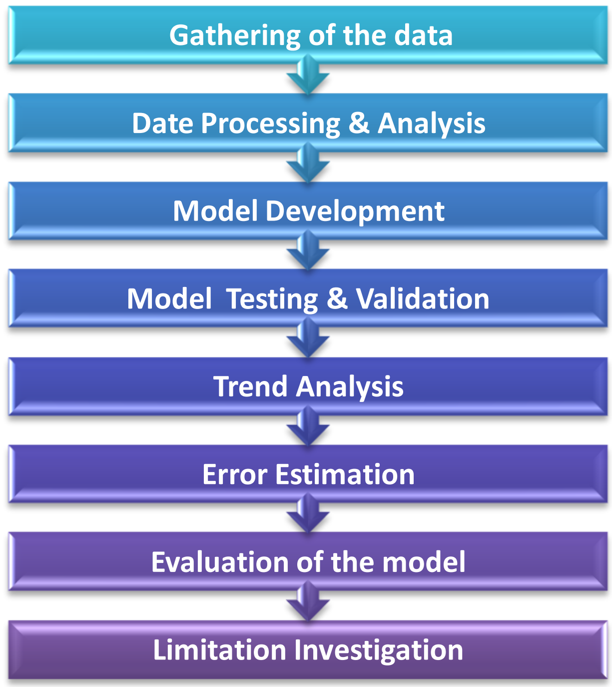


3.7 Tools

The modelling in this project has been carried out by using GMDH Algorithms which has been explained in the introduction section. The following tools have been also used in this study;

- Microsoft Office Word: To write the reports
- Microsoft Office Excel: To prepare data sheets and calculations
- Prezi online website: To prepare presentations
- MATLAB Software: To develop GMDH modelling approach for new oil formation volume factor correlation

3.8 Project activities framework:



3.9 Error analysis "Evaluation Techniques":

Each of the empirical correlations and experimental factors affecting the B_o calculations have been included in previous studies for such a project. The errors analysis in identifying any new correlation is considered as one of the most important factors that lead to the acceptance of any new model compared against models that previously discovered, especially if these correlations did not rise up to the level of the most accurate correlation. From this point we can say that the errors analysis is very important and must be implemented, and by using the calculations of B_o we could face some errors that we might get to find the new correlation of B_o and compare it against what has been developed before.

3.10 Statistical Error Analysis

As has been mentioned above it is very necessary to perform the errors analysis in developing any new project in order to check the performance of that produced project. In general, there are many techniques that could be used to test the correctness of FVF correlation of oil and gas mixture. According to the literature review, the statistical error analysis that were used are, the average percentage relative error (Er), the average absolute percentage relative error (Ea) , maximum absolute percentage relative error (Emax), standard deviation and lastly the coefficient of the correlation. Any parameter has its own equation as shown below:

1. Average Percent Relative Error:

$$E_r = \left({}^{1}/_{n_d}\right)\sum_{i=1}^{n_d} E_1$$

Where

$$E_1 = \left\{\left(x_{exp} - x_{est}\right)/x_{est}\right\}_i .100, \qquad i = 1,2,\ldots..n_d$$

2. Average Absolute Percent Relative Error:

$$E_a = \left({}^{1}/_{n_d}\right)\sum_{i=1}^{n_d} |E_1|$$

3. Maximum absolute percentage average error :

$$E_{max} = max_{i=1}^{nd}\ |E_i|$$

4**. Standard deviation:**

$$s = \sqrt{\frac{1}{n_d - 1}\sum_{i=1}^{n_d}(E_i - E_r)^2}$$

5. Coefficient of the correlation:

$$r = 1 - \sqrt{\sum_{i=1}^{n_d}\left(x_{exp} - x_{est}\right)_1^2 \Big/ \sum_{i=1}^{n_d}\left(x_{exp} - x^{-}\right)_1^2}$$

3.11 Graphical Error analysis:

It is very important to visualize the performance of any nwe model, that is why graphical analysis of obtained errors is required. This is helpful in determining the distribution of the error with aid of cross-plots in addition to other analysis can be performed.

CHAPTER 4

Results & Discussion

4.1 Data Gathering & Processing:

When it comes to develop an unprecedented correlation for FVF for oil and gas mixture, there are several parameters that contribute in constructing such a correlation. The quality and quantity of these parameters should be examined well to ensure right information have been used in developing the new model. During the gathering of the data, it is obvious there are many parameters that contribute in constructing the correlation of FVF of oil and gas mixture. Nevertheless, by using GMDH model not all of these parameters have been considered as main input data to develop the final output; because it is rarely to find all those parameters when it comes to data collection process due to time limit and unavailability of enough published papers.. Even though, this GMDH model has been develop by using less parameters as possible which in this study three parameters have been used namely gas solubility and reservoir temperature.

From the literature review, in order to construct the new model by using GMDH; not less than 200 data points should be considered in constructing the mathematical approach. Referring to most common parameters been used in the previous studies regardless it was an experimental or empirical methods, the input parameters have been selected.

4.2 OBTAINED RESULTS

Upon successful of constructing a new model to estimate a new correlation of B_o at bubble point pressure using GMDH method, as has been expected to use the new result in developing of new correlation for oil FVF which will led to decrease some of the problems that related to reservoir development, production engineering as well as it will fostering the capability to increase the oil productivity when EOR techniques are used.

To evaluate the accuracy and to check the development of the new model, trend analysis is required. This will draw a clear picture of obtained model with comparison of its performance against previous models. Thus by using GMDH different input parameters have been used for each set. For example every single set will include just one parameter and maintain the rest of the parameters constant.

It is important to study the effects of each parameter in developing of the new model such as; specific gravity of the oil, specific gravity of the gas, Gas oil Ratio (GOR) and temperature and pressure of the reservoir. The created trends of the new model were shown much accurate results than previous one.

As has been mentioned earlier in the literature review, the new predicted model of the oil formation volume factor will be compared against existing FVF correlations in term of its accuracy and performance. In addition to, comparison table of all B_o correlations as well as statistical parameters have been developed. By using GMDH in developing a new correlation for FVF for oil and gas mixture, the new developed corellation for oil formation volume factor outperforms over all previously obtained correlations.

4.3THE GATHERING AND SORTING OF PVT DATAPOINTS

A set of data has been selected to represent the real data to be developed using GMDH approach; these data include API, specific gravity of oil, specific gravity of gas, gas solubility and pressure & temperature of the reservoir.

By using data set that is available in the previous published papers, the author has attempted to develop a new correlation for oil formation volume factor at bubble point pressure as an empirical correlation. The B_o at below and above bubble point pressure also can be developed by using GMDH approach. But due to not enough data that required, the author couldn’t get the chance to develop B_o at below and above bubble point pressure. The PVT data that have been used in this study were oil formation volume factor at bubble point pressure, gas solubility, oil specific gravity, gas specific gravity and pressure & temperature of the reservoir. These data were among the data that has been used by many researchers to develop such a correlation.

Based on the selected data, the author use a total of 268 data point collected from different regions. 92 were from Malaysia oil fields and 125 from Middle East and about 51 data sets from UAE's reservoirs. After selecting the required data, the process of duplicate screening as well as crosschecking for the entered groups were applied. Thus the repentance of the inputs was successfully avoided.

By using Microsoft Excel Spreadsheet, the author got to group the data sets randomly. This process was following by selecting appropriate ratios for data training, data validation and data test processes. 2:1:1 has been selected in order to produce the desirable results. Different numbers of data were used in developing each of mentioned processes in details as following:

Table (3): Data set

	Data for Training	**Data for Validation**	**Data for Testing**
Number of data	134	67	67

Table (4): Maximum and minimum value for selected data

Parameters	**Maximum**	**Minimum**
API	**53.2**	**21.9**
Oil Specific gravity	**0.9224**	**0.7661**
Gas Specific gravity	**1.315**	**0.612**
Gas Solubility scf/stb	**2266**	**127**
Reservoir Temperature, degree F	**280**	**74**
Oil Density, API	**53.92**	**30.95**
Bubble point pressure, PSI	**4640**	**508**

4.4Parameters Reduction:

Reducing the parameters that have been used in the developing the new correlation for oil formation volume factor (B_o) was the most significant objective in this study. Focusing more on selecting the much appropriate inputs in order to obtain much accurate outputs.

Erasing scheme is to some extend is a process of reducing the entered parameters so as to see the effects of each parameters in developing the GMDH model. These parameters include API, specific gravity of the oil, specific gravity of the gas, gas solubility (Rs) and lastly the reservoir pressure & temperature. Moreover, the statistical error analysis can be used to evaluate the effects of reducing each parameter and visually display these impacts on graphical cross plots. The following figure illustrates the different parameters that have been in different oil formation volume factor corellations; only two parameters have been used in this study

Table (5): Main parameters for diffierent studies

Correlation	This Study	Al-Marhoun	Standing	Alshmmasi (3 par.)	Alshammasi (4 par.)
Oil SG	✗	✓	✓	✓	✓
Gas SG	✗	✓	✓	✗	✓
Gas Solubility	✓	✓	✓	✓	✓
API	✗	✗	✗	✗	✗
Reservoir Temperature	✓	✓	✓	✓	✓
Density	✗	✗	✗	✗	✗

4.5 GMDH MODEL FOR OIL FORMATION VOLUME FACTOR, B_o AT BUBBLE POINT PRESSURE

In order to develop new correlation for oil formation volume factor at bubble point pressure, several equations should be develop using GMDH approach. Different inputs will give different results. Therefore, this procedure required careful and appropriate selection of the inputs. By trying different inputs, the best correlation to evaluate the B_o at bubble point pressure had been obtained by using only 2 parameters out of 7 parameters from the gathered data. GMDH approach was used with the following selected data:

- Solution Gas-Oil Ratio, R_s
- Reservoir Temperature

The author developed many GMDH models, each model includes different input parameters but the optimum result was obtained by using two parameters as has been mentioned above. From the literature review, we can notice that these two parameters have been used in all Bo correlations.

```
Building GMDH-type neural network...

Building layer #1...

Neurons tried in this layer: 15

Neurons included in this layer: 1

RMSE in the validation data of the best neuron: 0.032700

Done.

Number of layers: 1

Number of used input variables: 2

Execution time: 0.15 seconds

Number of layers: 1

Layer #1

Number of neurons: 1

y = 1.03 -0.000506*x6 +0.000356*x5 +5.75e-007*x5*x6 +3.9e-006*x6*x6
+4.04e-008*x5*x5
```

There are only 6 parameters as inputs which are bubble point pressure, specific gravity of oil, specific gravity of gas, gas solubility, API, and temperature of the reservoir. By reducing the number of the input parameters, the best performance of the model can be obtained with high degree of accuracy. The previous box shows the outputs that have been obtained from GMDH model. the following equation represents the final equation that has been developed for oil formation volume factor at bubble point pressure;

$$y = a_0 - a_1 * x6 + a_2 * x5 + a_3\,e{-007} * x5 * x6 + a_4 e{-006} * x6 * x6 + a_5 e{-008} * x5 * x5 \quad \text{.....(Equation 15)}$$

y= oil formation volume factor at bubble point pressure, rb/stb

X_5 = gas solubility,scf/stb

X_6 = reservoir temperature, F

a_0 = 1.03

a_1= 0.000506

a_2= 0.000356

a_3= 5.75

a_4= 3.9

a_5= 4.04

The following Schematic Diagram illustrates the basic concept behind the developed GMDH model;

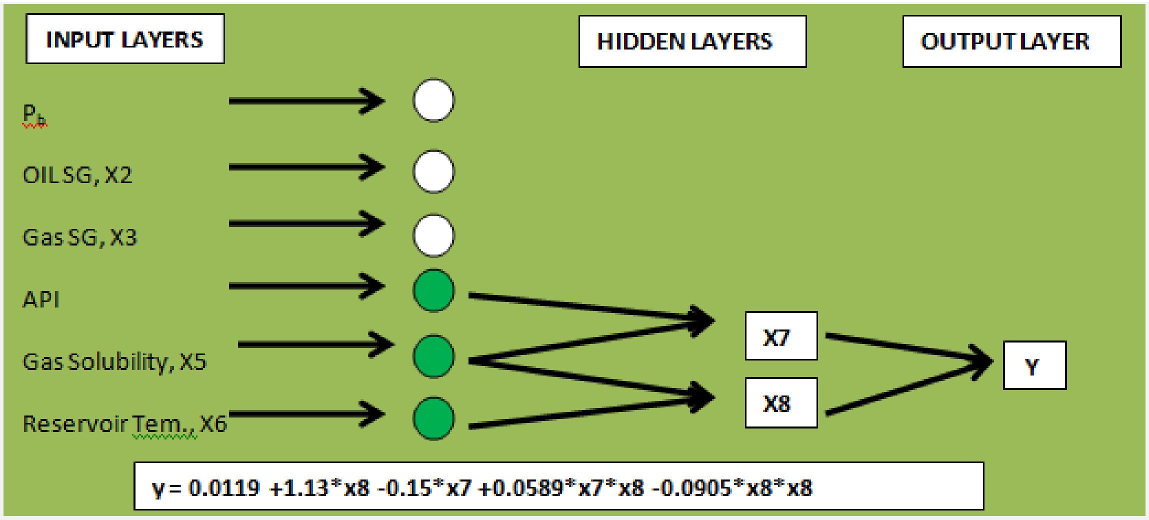


Figure 3: Schematic Diagram Of The Proposed GMDH Model Topology

But API has been excluded from this model because it doesn't match the real trend analyses of oil formation volume factor. So the last model has been developed by using only gas solubility and reservoir temperature. And the new equation has been mentioned in the previous section.

4.6 B_o STATISTICAL AND GRAPHICAL ERROR ANALYSIS

The following table tabulates the statistical error analysis of the developed GMDH model (Bo model)

Table (6) Statistical Error Analysis

Statistical Er. Anal.	This study	Standing	Al-Marhoun	Alshammasi 3 Par.	Alshammasi 4 Par.
AAPE	1.547	2.561	1.947	2.839	1.692
R^2	0.993	0.9799	.984	0.9838	0.9884
Standard Deviation	0.0271	2.182	2.099	1.684	1.648
E_{max}	5.7807	13.465	12.505	10.180	9.313
E_{min}	0.00229	0.0138	.0018	0.0329	0.0297

From the previous table it is obvious that there is small domain of average absolute relative errors (1.543%) whereas the correlation coefficient has been calculated as 0.993 . Moreover, the standard deviation for the new B_o model has been calculated as 0.0271% with 5.78% of suggested maximum error for this correlation; this percentage points out a high accuracy of the measurements. The following diagram illustrates the statistical error analyses of the new B_o comparing to the predicted B_o. As a result, this study generated unprecedented values for formation volume factor at bubble point pressure.

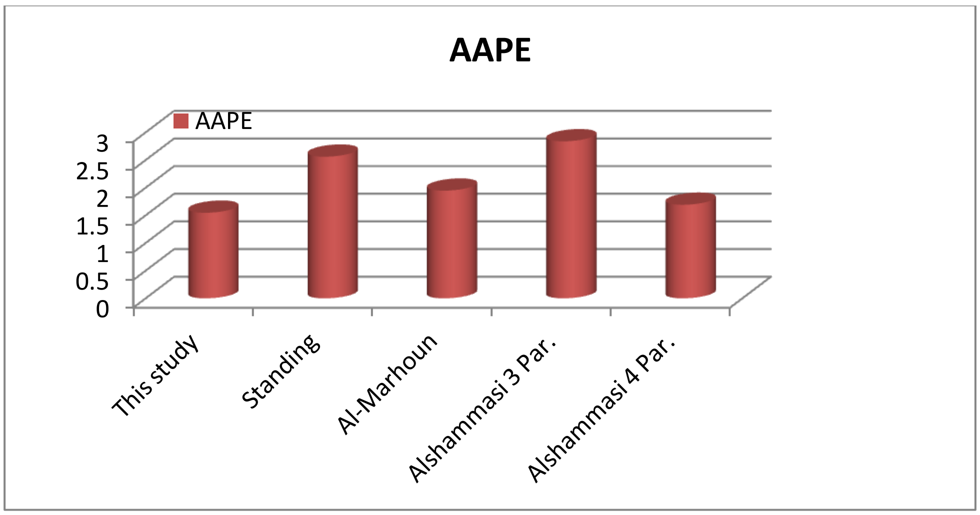


Figure 4: Bo average absolute relative errors Vs. other correlations

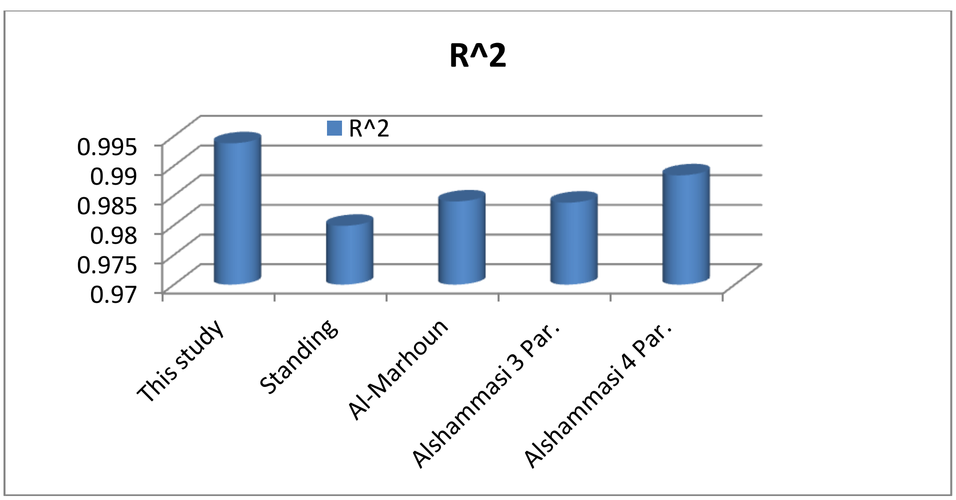


Figure 5: Bo correlation coefficient Vs other correlations

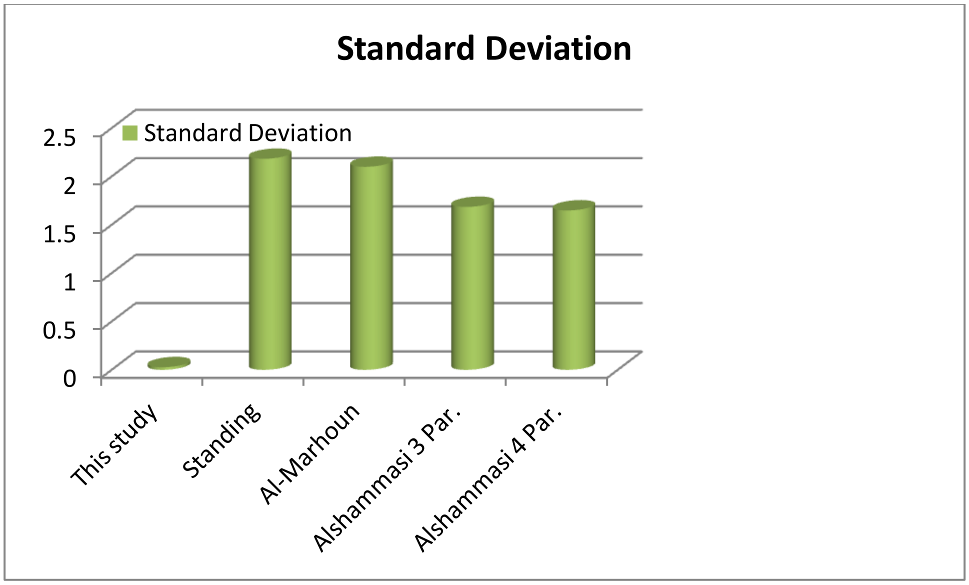


Figure 6: Bo Standard Deviation Vs Other correlations

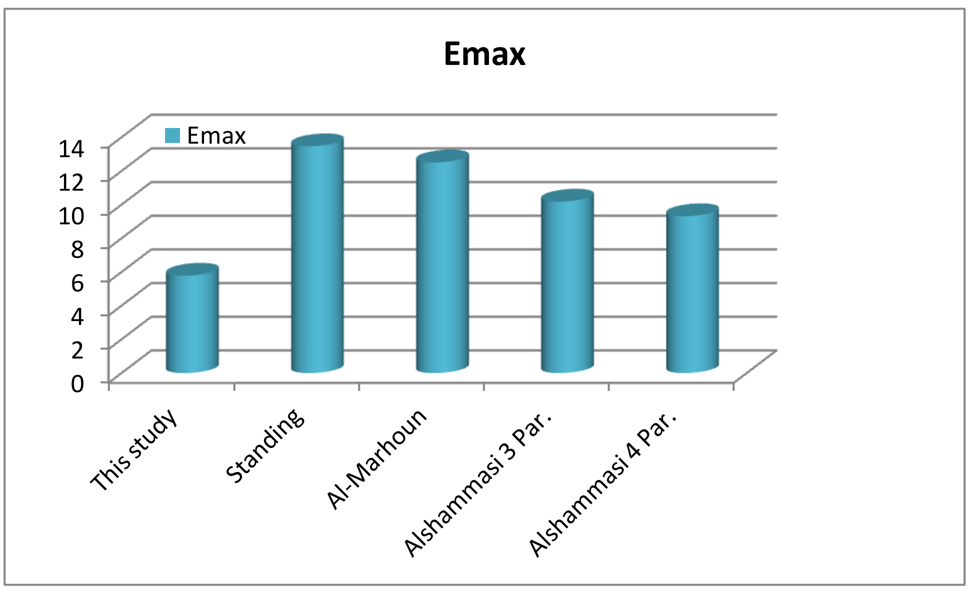


Figure 7: Bo Maximum absolute percentage average error Vs other correlations

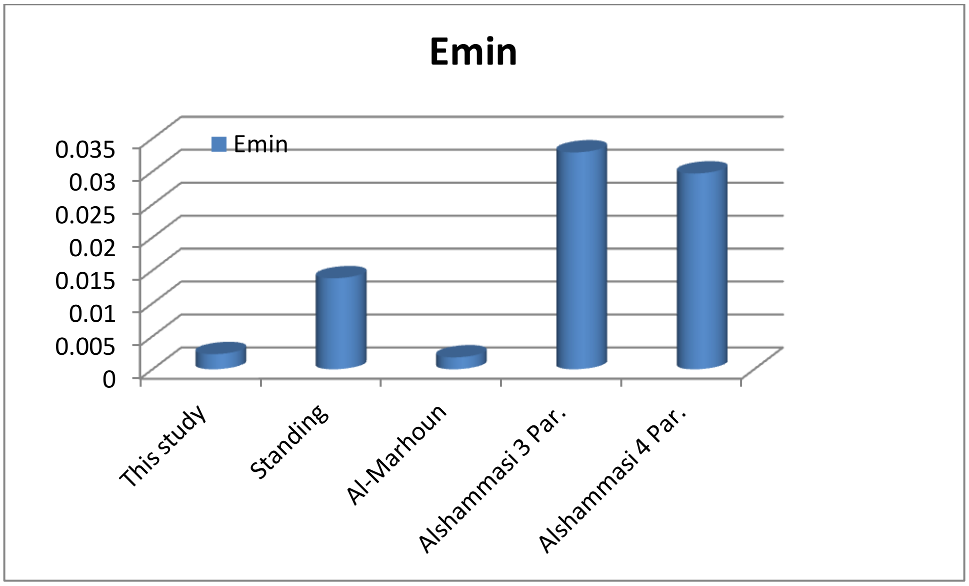


Figure 8: Bo Minimum absolute percentage average error Vs other correlations

4.7 Scatter graphs

The scatter graphs of the measured Bo vs. expected Bo are displayed in the following figures. Usually the cross plots shows the degree of Compatibility of two prospective values. By checking the cross plots between the measured and the predicted values, the author can point out the accuracy of this study. If the points lay on the line that means it is perfect. The scatter diagram of this study shows accurate results in other hand the standing's correlation demonstrates much dispersion.

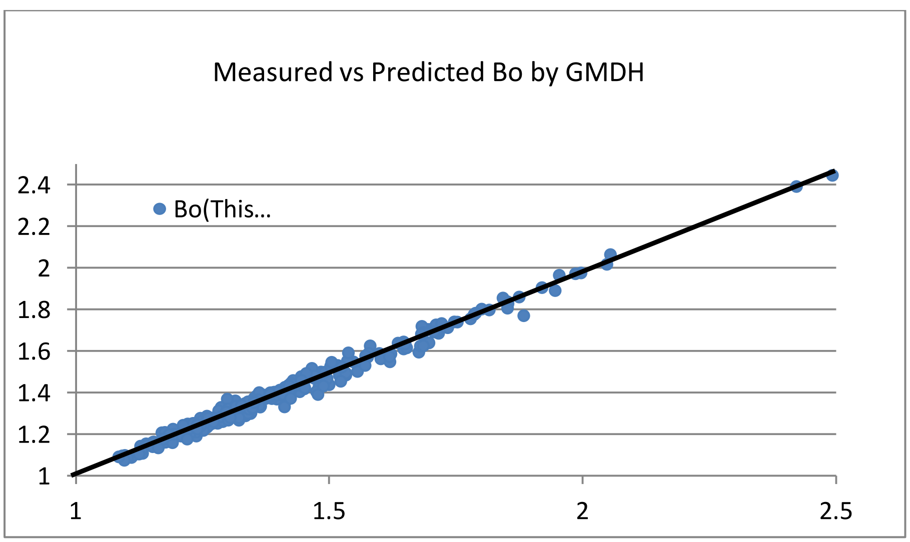


Figure 9: scatter graphs of the measured Bo vs actual Bo

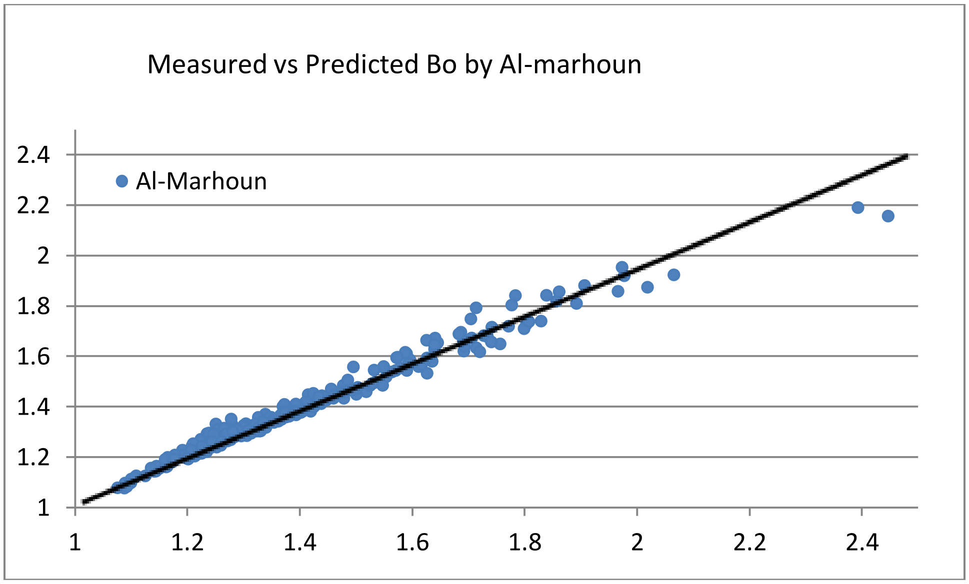


Figure 10: Measured vs. Predicted Bo by Al-Marhoun Correlation

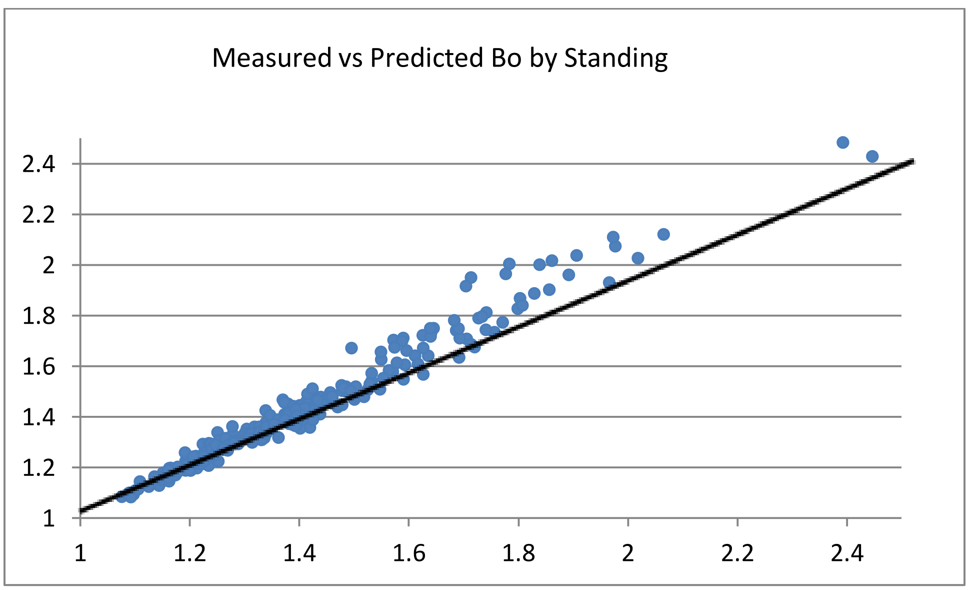


Figure 11: Measured vs. Predicted Bo by Standing Correlation

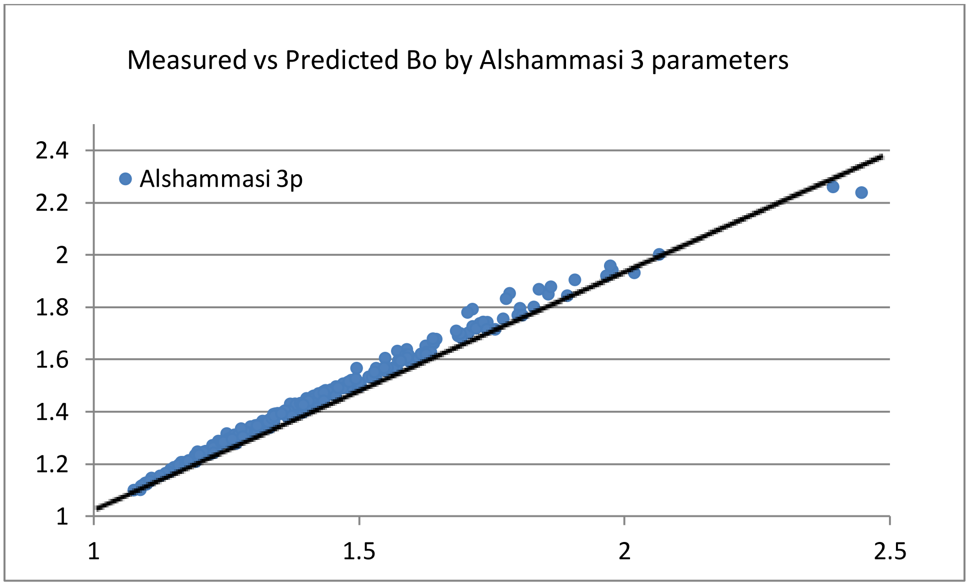


Figure 12: Measured vs. Predicted Bo by Al Shammasi (3 par.) Correlation

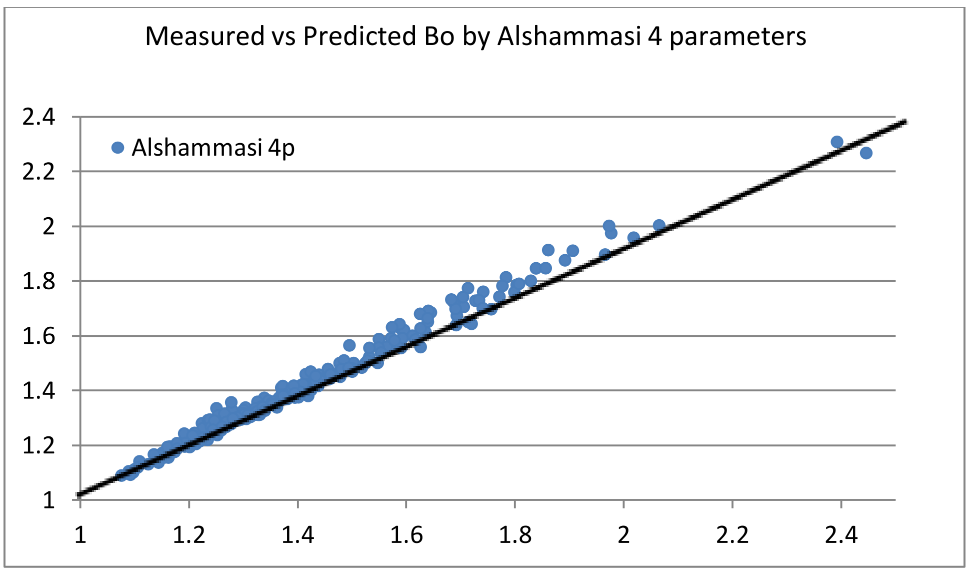


Figure 13: Measured vs. Predicted Bo by Al Shammasi (4 par.) Correlation

The following figures presents the cross-plot of estimated formation volume factor at bubble point versus measured formation volume factor for the proposed GMDH model; Training, Validation and testing sets. The coefficient that has been obtained for training set was 0.99294, while for the validation set was 0.97794 and for the testing set the value that obtained is 0.99389.

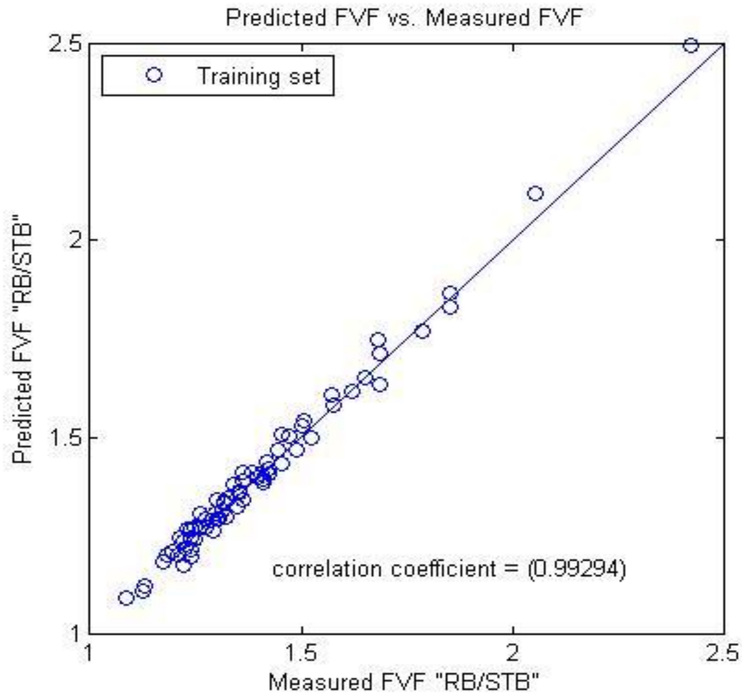


Figure 14: Scatter graph for Training set

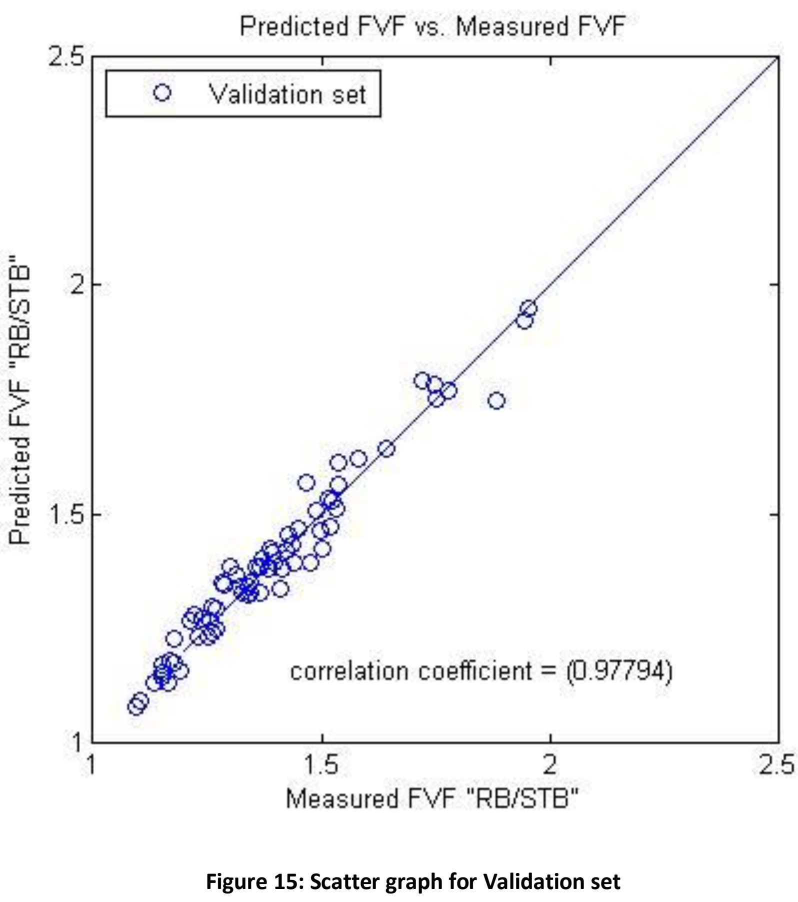


Figure 15: Scatter graph for Validation set

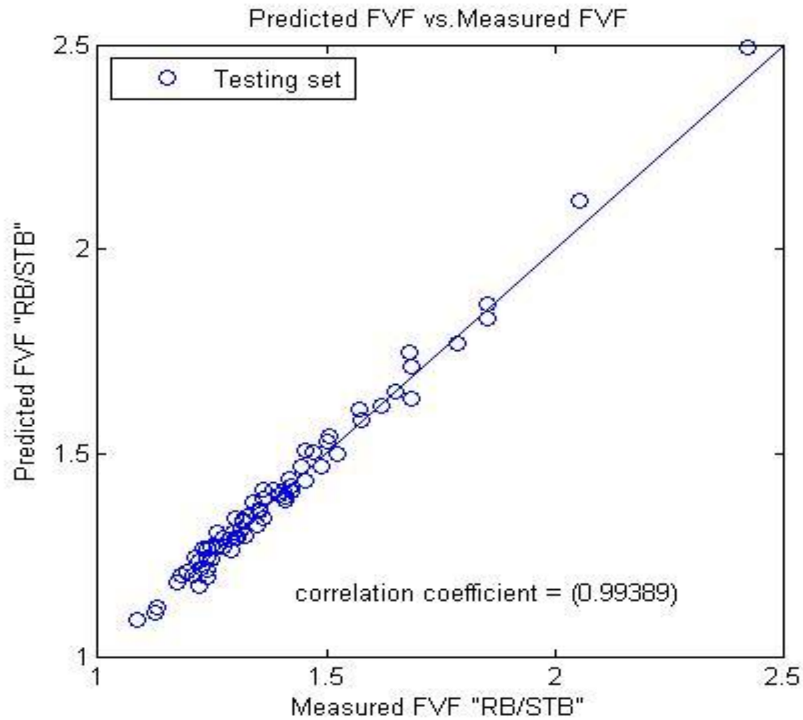


Figure 16: Scatter graph for Testing set

4.8 Trend Analysis

Trend Analysis is the practice of collecting information and attempting to spot a pattern to see the relationship between prospective inputs against physical lows.

Trend analyses have been discussed by several authors. These trends are considered as basic part in the acceptance of the developed empirical PVT correlations. Usually the trend analyses conduct to check the trend of specific parameters against the physical low without any contradiction. Moreover, it is use to check the GMDH developed model against the physical low or to check whether the new model is physically correct or not. The following figures visualize the relationship between the oil formation volume factor with the model main input parameters which are the reservoir temperature and gas solubility. As has been expected the new model obtained truthful trends that match the real oil formation volume factor trend. The reservoir temperature is considered as the main parameter to check the oil FVF trends, so as reservoir temperature increases the FVF will increase. When solution gas oil ratio increase it will cause increase in the total volume due to more amount of gas being in the oil which will lead to reduce the oil density and increase the total volume. the following diagrams indicate the relationship between Bo and the main input data that have been used in this study.

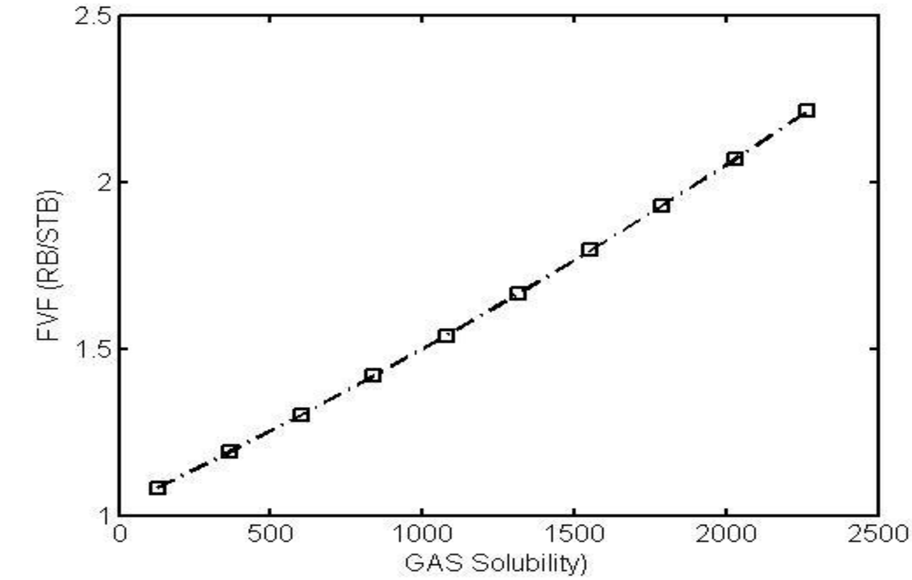


Figure 17: oil FVF Vs. Gas solubility

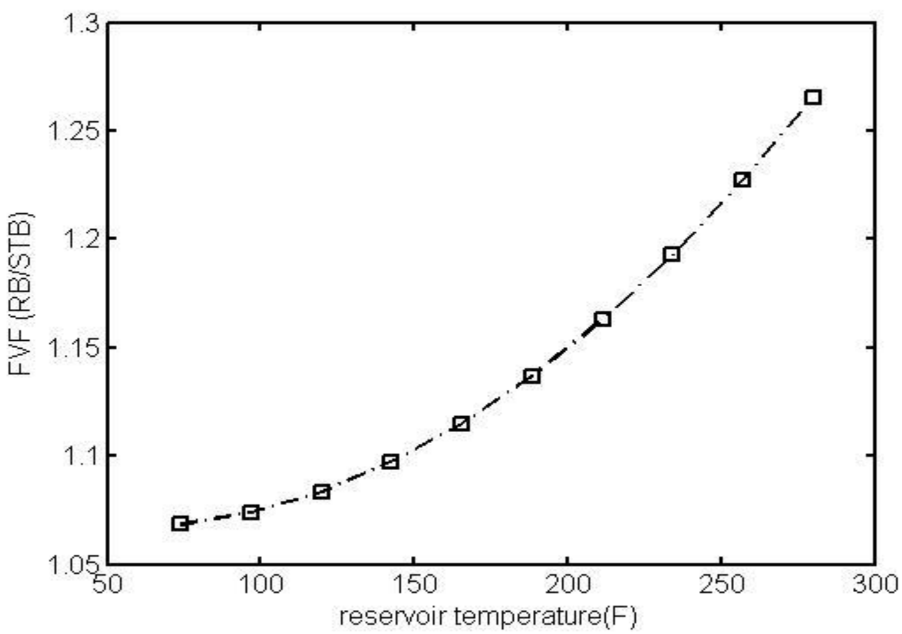


Figure 18: oil FVF vs. Reservoir Temperature

API trend doesn't show the real trend analysis with FVF; That is why API has been excluded manually form the new model and the new proposed mode has been developed by using on gas solubility and reservoir temperature which gave the optimum results.

4.9 Group Error Analysis:

Group error analysis is other technique that has been used in order to check the accuracy and the performance of the new model. the main input parameters have been divided into three different ranges. By estimating the average absolute error for each range, the new model achieved better performance than the previous studies. The following figure show the group error anaysis for both gas solubility and resrvoir temperature;

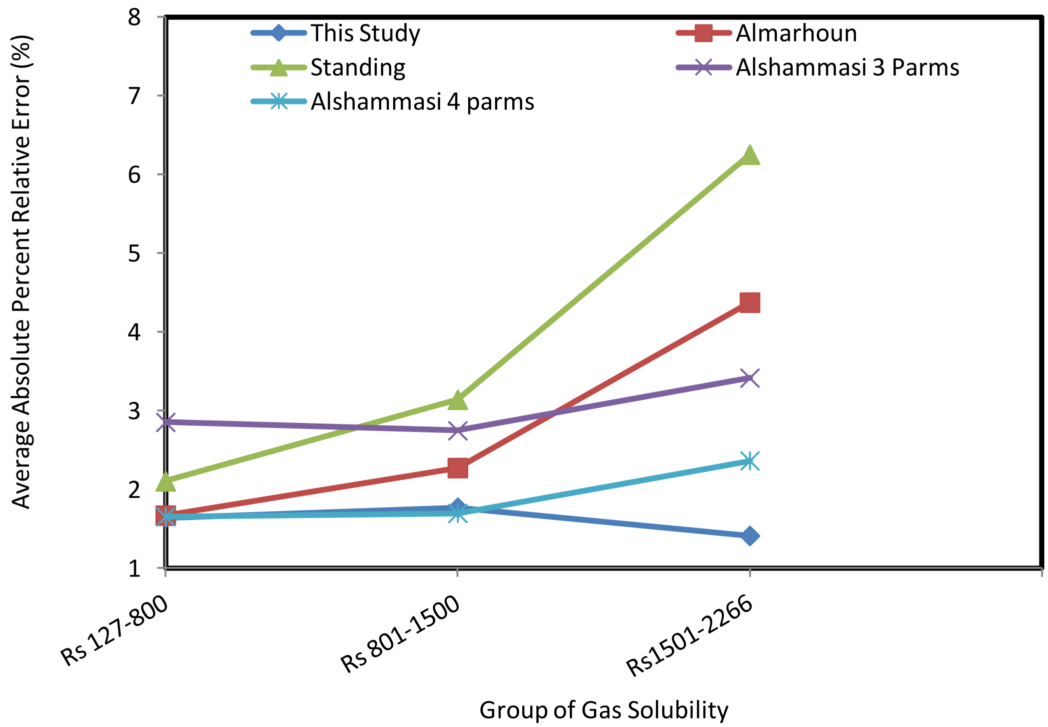


Figure 19: Group error analysis for gas solubility

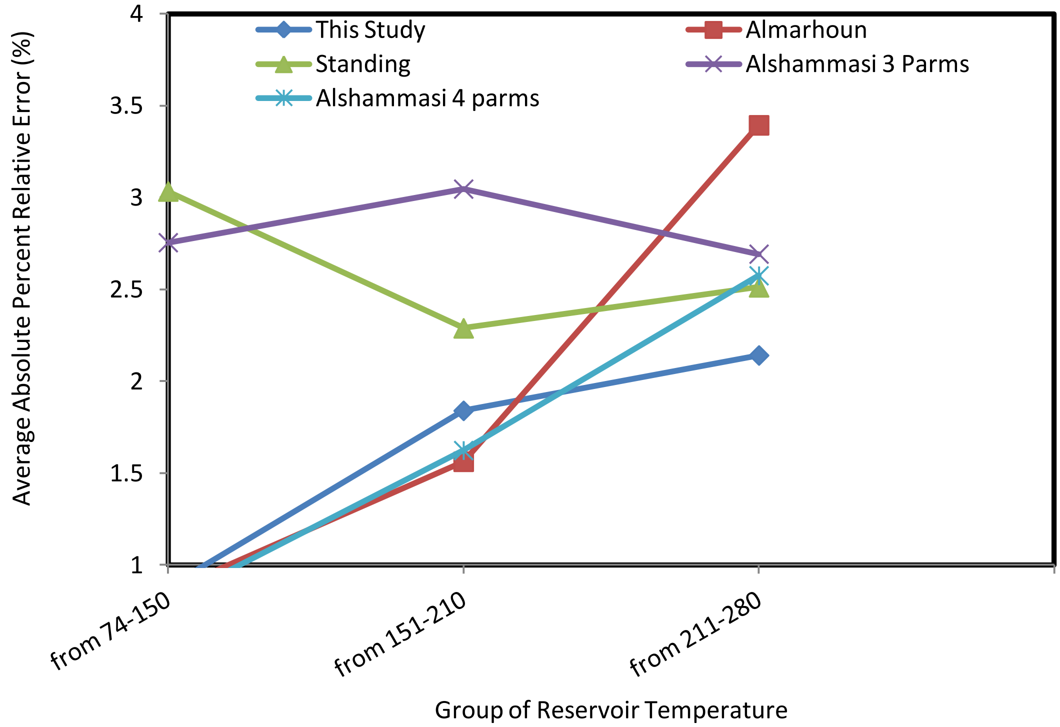


Figure 20: Group error analysis for reservoir temperature

Chapter 5
Conclusion and Recommendations

5.1 Conclusion

It is clear that there are many studies that have been carrioud aout by several researchers to develop correlations for oil formation voulme factor for oil and gas mixture. Each developed correlation has its own feature and it is developed by using different approachs. In order to develop an unprecedented correlation for FVF for oil and gas mixture, different parameters should be considered in constructing the new model using GMDH approch; for example oil specific gfavity, oil specific gravit, gas oil ratio (GOR) and temperature & pressure of the reservoir. In this study gas solubility and reservoir temperature have been used as the main input for GMDH model. Thus, it is quit important to find appropriate data sets as input data for the developed model. total number of 268 data sets have been utilized to develp the new B_o at bubble point pressure. In addtion to, validity and accurateness of the new model have been checked against the most accurate published correlation.

The new developed model has achieved its objectives that have been setted in the earlier chapter of this report. The new correlations for oil formation volume factor outperforms other tested empirical correlations (Standing, Al-Marhoun and Alshammasi). On top of that, this developed model also successfully manages to study the effect of reducing the parameters used for the GMDH build correlation.

Small range of absolute average relative errors (1.53%) has been obtained whereas the correlation coefficient has been calculated as 0.993. Moreover, the standard deviation for the new Bo model has been calculated as 0.0271% with 0.00229% of minimum error for this correlation. Trend analyses have confirmed that this new model for oil formation volume factor at bubble point pressure is physically correct.

5.2 Recommendations

Based on the previous conclusion, there are many recommendations that can be suggested for this project in order to enhance the project performance as well as obtaining much accurate results:

- GMDH model can be more accurate by collecting wide range of data from different fields with additional inputs.
- The code of GMDH for oil formation volume factor at bubble point pressure can be improved which will definitely lead to more accuracy in the outputs of the developed model in the future. Therefore, all researches are highly recommended to focus in this point.
- Smart Simulator can be used to double check the performance of the developed model.
- Experimental work may be required to obtain data sets in order to predict the oil formation volume factor at below and above bubble point pressure since the author find difficulty in gathering enough data at below and above bubble point pressure.

References

[1] Ahmed, T.," Reservoir Engineering Handbook"; second Edition, Gulf publishing Company, Houston, Texas, 2000.

[2] Katz, D. L.: ''Prediction of shrinkage of crude oils,'' Dril. Prod. Prac. Am. Pet. Inst. Pp. 137-147 (1942).

[3] Standing, M.B., "A Pressure-Volume-Temperature Correlation for Mixtures of California Oils and Gases"; Drill and Prod., API, pp 275-286, 1947.

[4] Standing, M. B.: ''Oil-System Correlation.'' Petroleum Production Handbook T. C. Frick (ed), SPE Richardson. TX (1962) 2. Chap. 19.

[5] Calhoun, J.C.,JR.," Fundamentals of Reservoir Engineering"; University of Oklahoma Press, Norman, Oklahoma, 35, 1947.

[6] Vazquez, M.E., and Beggs, H.D., "Correlations for Fliud Physical Property Prediction"; JPT, pp. 986-970, June 1980.

[7] Glaso, O.," Generalized Pressure-Volume-Temperature Correlations", JPT, pp. 785-795, May 1980.

[8] Al-Marhoun, M. A.: ''PVT Correlations for Saudi Crude Oils,'' SPE 13718, SPE Middle East Oil TECH CONF and EXHB (Manamah, Bahrain, 11-14 – March 1985)..

[9] Obomanu, D. A. and Okpobory, G. A.: ''Correlating the PVT Properties of Nigerian Crudes," Tran ASME (1987) Vol. 109, pp 214-14.

[10] Al-Marhoun, M.A.," PVT Correlatons for Middle East Crude Oils", JPT, pp. 650-666, May 1988.

[11] Abdul-Majeed, G. H. A. and Salman, N. H.: ''An Empirical Correlation for FVF Prediction," J Can. Pet. Tech., 27(6): 118-122 (1988).

[12] Al-Marhoun, M. A.,"New Coreelations Formation Volume Factors of Oils and Gas Mixtures", JCPT, pp. 22-26, March 1992.

[13] Farshad, F. F, Leblane, J. L, Garber, J. D. and Osorio, J. G.:"Empirical Correlation for Colombian Crude Oils," SPE 24538 (1992)

[14] Dokla, M. and Osman, M.: ''Authors' Reply to Discussion of Correlation of PVT Properties for UAE crudes," SPE Formation Evaluation (March 1993) pp 82; SPE paper 26316.

[15] Al-Fattah, Saud Mohammed and Al-Marhoun, M. A.: ‘’Evaluation of empirical correlation for bubble point oil formation volume factor,” Journal of Petroleum Science and Engineering. 11(1994) 341-350.

[16] 1A.A.Al-Shammasi, "Bubble point pressure and oil formation volume factor correlations," 1999.

[17] Ivakhnenko, A. G. (1966). Group Method of Data Handling a Rival of the Method of Stochastic Approximation. Soviet Automatic Control, 13, 43-71.

[18] Farlow, S. J. (1981). The GMDH algorith of ivkhnenko. The American Statistian, Vol.35, No.4, 210-215.

[19] Farlow, S. J. (1984). The GMDH algorithm,” in Self-Organizing Methods in Modeling: GMDH Type Algorithms. New York: Marcel-Dekker.

[20] "GMDH," [Online]. Available: http://www.gmdh.net/GMDH_his.htm. [Accessed 2013].

[21] "GMDH.net," [Online]. Available: http://www.gmdh.net/index.html. [Accessed 2013].

[22] Azubuike, I. I. and Ikiensikimama, S. S., ” Prediction of Oil Formation Volume Factor Using an Intelligent Tool: Artificial Neural Network”.

Appendixes

1. GMDH Script that have been used in this study

```
clc;
% 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('main_data.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=1:134
    atr(i,:)=ndata(i,:);
end
for i=135:201
    aval(i-134,:)=ndata(i,:);
end
%
for i=202:length(ndata)
    atest(i-201,:)=ndata(i,:);
end
Ytr=atr(:,1);
Xtr=atr(:,2:7);
Xtst=atest(:,2:7);
Ytst=atest(:,1);
Yv=aval(:,1);
Xv=aval(:,2:7);
[model, time] = gmdhbuild(Xtr, Ytr, 2, 0, 2, 0, 2, 1, 0.9, Xv, Yv,1);
gmdheq(model, 3);
[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(Ytst,Yqtst,'o')
grid off
set(gcf, 'color', 'white')
axis square

title('Predicted FVF vs. Measured FVF');
xlabel('Measured FVF "RB/STB"');
ylabel('Predicted FVF "RB/STB"')
legend('Training set', 'location', 'Northwest')
% Addding Reference Line with 45 degree slope
line([1.2 ; 2.5],[1.2 ; 2.5])
%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

% Addding Reference Line with 45 degree slope
line([1.2 ; 2.5],[1.2 ; 2.5])
%HINT: Select the y-value based on your data limits

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

plot(Yv,Yqval,'o')
grid off
set(gcf, 'color', 'white')
axis square
title('Predicted FVF vs. Measured FVF');
xlabel('Measured FVF "RB/STB"');
ylabel('Predicted FVF "RB/STB"')
legend('Validation set', 'location', 'Northwest')
% Addding Reference Line with 45 degree slope
line([1.2 ; 2.5],[1.2 ; 2.5])
%HINT: Select the y-value based on your data limits

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

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

title('Predicted FVF vs.Measured FVF');
xlabel('Measured FVF "RB/STB"');
ylabel('Predicted FVF "RB/STB"')
legend('Testing set', 'location', 'Northwest')
% Addding Reference Line with 45 degree slope
line([1.2 ; 2.5],[1.2 ; 2.5])
%HINT: Select the y-value based on your data limits

% Evaluating the correlation coefficient for testing set:
% ======================================================
Rtt1=corrcoef(Yqtst,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
Errort1 = Yqtr-Ytr;
plot(Errort1,':ro');
grid off
set(gcf, 'color', 'white')
title('Error Distribution 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/n*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(Errortt1);

% 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 of GAS Solubility while fixing the other parameters
% ------------GAS Solubility variation------------------------------------

ps5=[linspace(508,508,10); %Bubble Point Pressure [min=508    max=4640
mean=2137.420074]
linspace(27.5,27.5,10); %API [min=21.9    max=53.2  mean=36.23596]
linspace(0.889937107,0.889937107,10);%OIL Specific Gravity [min=0.766107
max=0.922425    mean=0.844612]
linspace(1.072,1.072,10);%GAS Specific Gravity [min=0.612  max=1.315
mean=0.892736059]
linspace(127,2266,10);%GAS Solubility [min=127    max=2266    mean=689.9442379]
linspace(130,130,10)]';%Reservoir Temperature [min=74    max=280
mean=177.513]


% Now simulate
[Yq_Rs] = gmdhpredict(model, ps5);
% Plot Figures for GAS Solubility variation
figure
px5=plot(ps5(:,5),Yq_Rs(:,1),'-rs');
set(gca,'YGrid','off','XGrid','off')
set(gca,'FontSize',12,'LineWidth',2);
set(px5,'LineStyle','-.','LineWidth',1.5,'Color','k','MarkerSize',6)
xlabel('GAS Solubility)','FontSize',12)
ylabel('FVF (RB/STB)', 'fontsize',12)

%------------------------------------------------------------------------------
% Simulation: Variation of resevoir temperature while fixing the other
parameters
% ------------resevoir temperature variation-----------------------------------
--

ps6=[linspace(508,508,10); %Bubble Point Pressure [min=508    max=4640
mean=2137.420074]
linspace(27.5,27.5,10); %API [min=21.9    max=53.2  mean=36.23596]
linspace(0.889937107,0.889937107,10);%OIL Specific Gravity [min=0.766107
max=0.922425    mean=0.844612]
linspace(1.072,1.072,10);%GAS Specific Gravity [min=0.612  max=1.315
mean=0.892736059]
linspace(141,141,10);%GAS Solubility [min=127    max=2266    mean=689.9442379]
```

```
linspace(74,280,10)]';%Reservoir Temperature [min=74    max=280    mean=177.513]

% Now simulate
[Yq_resevoirT] = gmdhpredict(model, ps6);
% Plot Figures for resevoir temperature variation
figure
px6=plot(ps6(:,6),Yq_resevoirT(:,1),'-rs');
set(gca,'YGrid','off','XGrid','off')
set(gca,'FontSize',12,'LineWidth',2);
set(px6,'LineStyle','-.','LineWidth',1.5,'Color','k','MarkerSize',6)
xlabel('reservoir temperature(F)','FontSize',12)
ylabel('FVF (RB/STB)', 'fontsize',12)
```

2. Other Empirical correlations for Oil Formation Volume Factor

Standing:

B_o =a_0+(a_1*10^-4)*(((R_s *(γ_g / γ_o)^0 .5)+(a_2* T))^1 .175) ....(Equation 16)

Al-Marhoun

Bo =a_0+((a_1 * 1 0^-3)*(T+460))+((a_2*10^-2)*((R_s ^0.74239)*(γ_o ^0.323294)*(γ_o ^-1 .20204)))+((0.3 18099*10^-5)*(((R_s^0.74239)*(γ_o ^0.323294)*(γ_g ^-1 .20204))^2))... ....(Equation 17)

Al Shammasi 3 parameters

Bo = 1+a_1*(R_s/ γ_g) +a_2*((F3-60)/ γ_g)..(Equation 18)

Al Shammasi 4 parameters

Bo =1+ (a_1 * 1 0^-7)*(R_s *(F3-60))+(a_2 *(R_s/ γ_g))+(a_3*(T - 60)/ γ_g)+(a_4*R_s* γ_o / γ_g)......

....(Equation 19)

3. Error Distribution for GMDH model

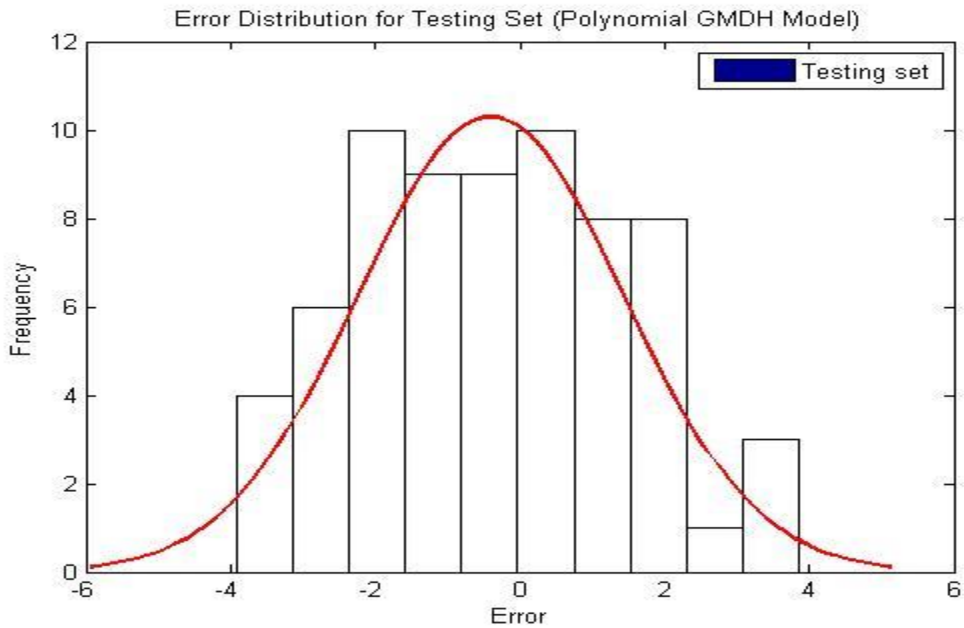


Figure 21: Error distribution for Training set

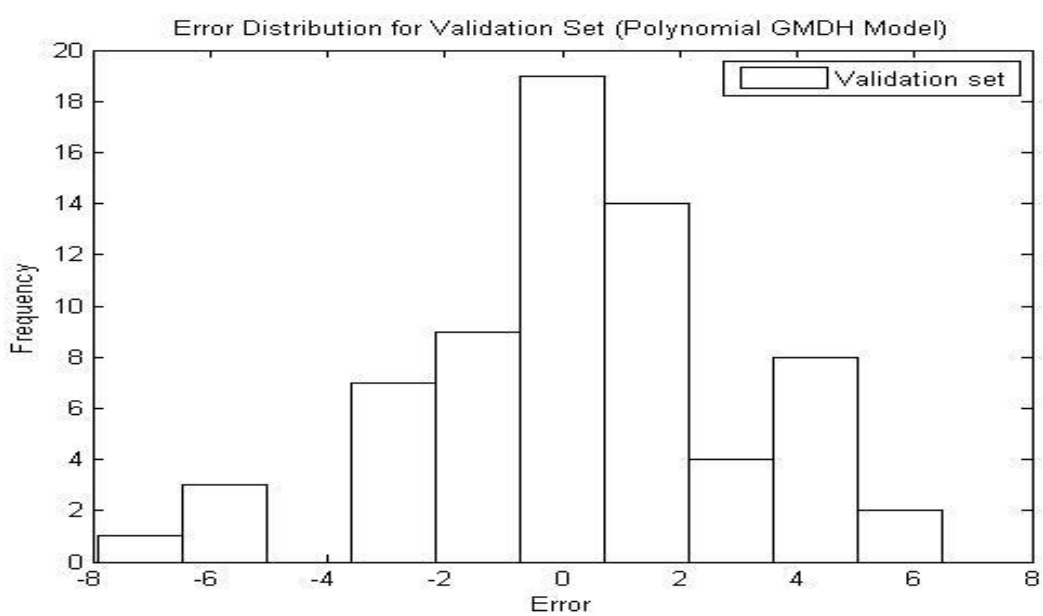


Figure 22:Error distribution for Validation set

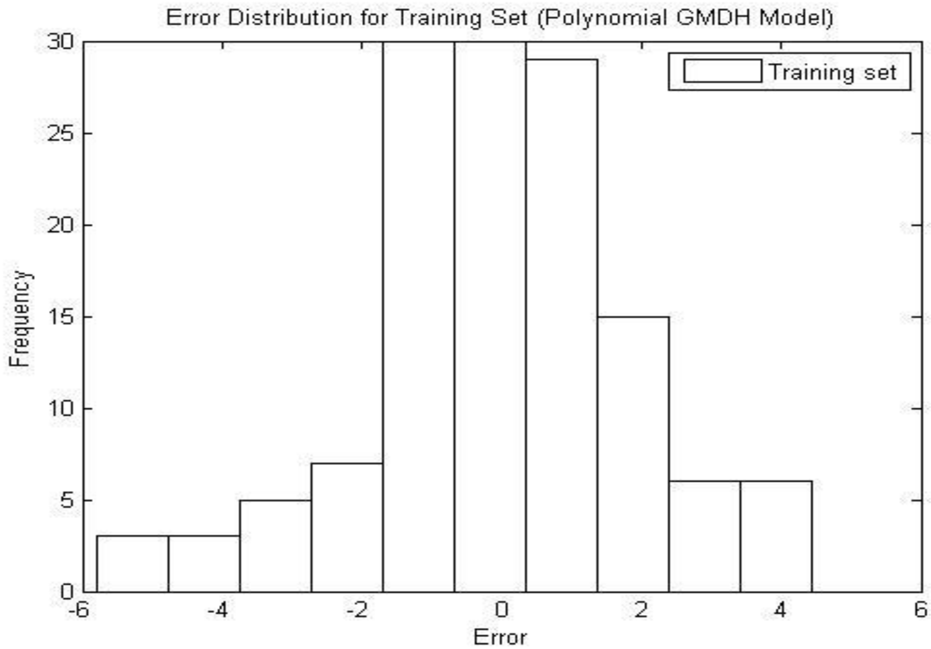


Figure 23: Error distribution for Testing set