Development of PVT Correlations for Heavy Oils Using Smart Regression Technique

By

Awab Moahmed Mahmod Soliman

17744

Dissertation submitted in partial fulfilment
of the requirements for the

Bachelor of Engineering (Hons)
(PETROLEUM ENGINEERING)

Jan 2015

Universiti Teknologi PETRONAS
Bandar Seri Iskandar
31750 Tronoh
Perak Darul Ridzuan

CERTIFICATION OF APPROVAL

Development of PVT Correlations for Heavy Oils Using Smart Regression Technique

By

Awab Moahmed Mahmod Soliman

17744

A project dissertation submitted to the
Petroleum Engineering Department
Universiti Teknologi PETRONAS
In partial fulfilment of the requirement for the
BACHELOR OF ENGINEERING (Hons)
(PETROLEUM ENGINEERING)

Approved by,

____

(Dr. Mohammed Abdalla Ayoub)

UNIVERSITI TEKNOLOGI PETRONAS
TRONOH, PERAK
Jan 2015

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.

Awab Moahmed Mahmod Soliman

Abstract

PVT analysis is an important part for most of reservoir and production engineering calculations. The most accurate data is usually provided by lab reports from samples taken from the field and send to the special PVT laboratories, this method is expensive. To find similar results, researchers were developing PVT correlations through the last 5o years, each correlation was generated utilizing specific field data, accordingly, the field characteristics has its influence in the accuracy of the output values, especially when these correlations has to be used for other field with or without the same characteristics. Through this project, measured PVT data will be collected and used to generate new models for oil physical fluid properties. Then the results will be compared with the most known published correlations. The technique to be used is Group Method of Data Handling which is a new approached used recently in oil and gas industry.

Acknowledgement

In the Name of Allah the most Merciful, All perfect praise to Allah, the Lord of the worlds, and may the peace and blessings be upon the final Prophet Muhamad, upon his household, his companions, and following them all in goodness till the day of judgment. Praise and thanks due to Allah, for his blessings and guidance throughout my life and the success He granted me in completing this industrial internship program in ease and success.

I would love to take this opportunity to thank all whom involved in making my final year project a countless educational session full of experiences.

With deep heart gratefulness to my mother who struggled to take me to where I am. I extremely appreciate her motivation and inspiration in my decisions, and her valuable advices that they provided me through all the study period.

My grateful and high appreciations towards my FYP supervisor Dr. Mohammed Abdalla Ayoub, for his dedication of his time and effort in instructing, guiding me in all my project-related tasks despite his other obligations.

My thankfulness is also extended to Universiti Teknologi PETRONAS providing such opportunities to accomplish this project.

Table of contents:

List of figures:

List of tables:

Abbreviations and Nomenclatures

API ° = stock tank oil gravity

GG, go, gm/cm3= specific gravity of gas

T_r (°F) = reservior tempreature

P_r(psia)= reservior pressure

P_b (psia)= bubble point pressure

μ_{od} (cp)= dead oil viscosity

μ_{ob}(cp)= saturated oil viscosity

μ_o(cp) = under-saturated oil viscosity

R_s(scf/STB)= solution gas oil ratio

OFVF, B_o= oil formation volume factor RB/STB

AAPRE = Average Absolute Percent Relative Error

Emin = Minimum Absolute Percent Relative Error

Emax = Maximum Absolute Percent Relative Error

R^2= Square of Correlation of Coefficient

Chapter 1

Introduction

1.1 Background:

Heavy oil characterized generally by its high viscosity which means high resistance to flow from the reservoir to the production facilities and accordingly through pipelines and transportation facilities. In oil and gas industry the heavy oil is classified under range of 10 to 22.3° API gravity, but oil with less than 10 ° API is considered as extra heavy oil. Despite the fact of the high viscosity of the oil, heavy crude oil nowadays is providing an interest field to oil and gas companies to develop, it certainly provide high income as related to the huge reservoir discovered recently all over the world, specially Canada and Venezuela. These reserves has to be calculated using physical properties of crude oil.

Knowledge of physical properties of petroleum fluids is essential for both reservoir and production engineers to perform their calculations. They are used for calculation of oil in place approximation and simulation, the necessity of fluid properties obtained from material balance equations, surface volumes and also important transportation parameters which will affect the fluid flow.

In reservoir scope the oil viscosity, dissolved gas content, density, and further factors along with how these properties contrast with temperature and pressure are obligatory for reservoir performance evaluation and for surface and subsurface facilities design purposes. Pressure Volume Temperature analysis (PVT) refer to the measured values of the physical fluid properties in correlation between their Pressure, volume and temperature. Concerning to reservoir and production systems, these physical properties must be measured at the reservoir temperature and altered reservoir pressures for reservoir performance studies.

With advancement in heavy oil extraction techniques, defining the actual reserves from those fields might need accurate properties determination. Properties such as FVF, GOR, viscosity and compressibility are ideally obtained from the laboratories tests. Reservoir fluids sample measured with special equipment for estimating the temperature, volume, and pressure of the sample are attained either by reservoir fluid sampling techniques or by combination of separator oil and gas at reservoir pressure, however it is not feasible to obtain them experimentally due to some technical difficulties like unavailability of the laboratories in the field and improper sampling technique which lead to contamination of the sample.

On the other hand, lately many modelling approaches were introduced in oil and gas industry. Group Method of Data Handling is one of those newly involving technique, which was built on the principles of self-organization as an inductive modelling method. This method was invented in the late 1960s by Prof. Alexey Grigorevich Ivahnenko, an academician from the Ukrainian Academy of Science, Institute of Cybernetics in Ukraine. Group Method of Data Handling (GMDH) method will be utilized through this project as it is recommended through the development of new correlations to provide more accurate results of the fluid properties than the published correlations.

1.2 Problem statement:

As the laboratories measured PVT data is consider more accurate and reliable than the correlation, but still not always obtainable. Because these laboratories facilities and operation funding is very expensive, therefore the sampling analysis and results provided by service companies remains expensive.

The most published PVT correlations are empirically derived correlations, they are geographically dependent on the specified field's fluid properties. Due to this limitation applying them directly to other fields is less reliable (low accuracy of any estimated fluid property as compared to its measured data).

In the development of these published correlations, some simplifying assumption and inappropriate or less accurate techniques has affected the accuracy of their results. Most of them cannot equitably estimate the properties of heavy oils especially at low temperature as they were derived utilizing lighter oils parameters.

1.3 Project objectives:

- To propose new set of PVT correlations from a wide range of field's data, utilizing smart regression technique which is Group Method of Data Handling.
- To evaluate the potential of using GMDH as smart regression technique in developing heavy oil correlations.
- To compare the performance of available models against the generated correlations by GMDH.

1.4 Scope of study:

- Throughout this study an experimentally measured PVT data from various field will be adopted to be utilized for developing new correlations.
- Cover heavy oil with defined gravity range of $10<API<22.3$
- Applying the most known PVT correlations published for heavy oils fluid properties.
- The new models would cover the following properties:
    - P_b Bubble point pressure, psia.
    - μ_{ob} Dead oil viscosity, cp.
    - μ_{od} Saturated oil viscosity, cp.
    - μ_o Under-saturated oil viscosity, cp.
    - B_o Oil formation volume factor, RB/STB.

Chapter 2

Literature Review

From the literature overview, PVT correlations are regularly act as alternation while such direct measurements of oil physical properties are not obtainable. Basically, the types of correlations are divided into two, Generic and geographical correlations. Generic correlations considered as the first correlations group developed using data selected in random manner. Second correlations group is established from a firm geographical area, this method way is to select one oil category or class, and they either develop new correlation for specific property or modify another published model that is generated using crude oil sample report with nearly or similar fluid characteristics.

2.1 Development of PVT correlations:

Literature search and review have shown the importance of developing PVT data correlations, they are usually acquire as much as data to be provided to qualify its results, each correlation model was adapted through the most significant parameters of the desired physical property.

Viscosity as the controlling property for heavy oil did not get much attention to develop its model, taking into account that most of the viscosity correlations has been developed from lighter oils. A reliability analysis of PVT correlations reported by De Ghetto, Paone and Villa [1] with conclusion that the dead oil viscosity property which was one estimated for the worst manner for all the correlations with average 36.2% errors. They related the justification due to that PTV correlations estimate the property by 2 input variables only, the API gravity and the reservoir temperature.

M.S, Hossain [2] discussed that the overall published correlations for oil viscosity developed empirically from light oil reports, therefore their prediction of heavy oil viscosity at low temperature is not reasonable, in his study the new empirically developed correlations for both saturated and also under saturated oil viscosity were generated to be applicable for higher API gravity of heavy oils within range of 10 to 22.3 API, the improvement in the accuracy of the viscosity models for dead oil was 3% to 50%, for saturated oil was 3% to 13% and for under saturated was 22% to 27% over the existing correlations.

The crude oil chemical composition were used by Lohrenz and Bary [12] in the development of another empirical dead oil viscosity correlation, the functions of the models are to be presented by the measured parameters like solution gas oil ratio, API Gravity, temperature and pressure, Later the pour point was added as an input parameter by Egbogah as reported by Egbogah and Ng3[3].

Birol and Peter [4] developed a different correlations in order to estimate viscosities of saturated and under saturated oils. They explained that they used the dead oil viscosity as an input for assessing the viscosity correlation of saturated oil, and then same criteria applied on under saturated oil viscosity correlation utilizing saturated oil viscosity as an input parameter. The model is mainly function of both oil API gravity and the reservoir emperature. Their anticipated model was a utility of additional linked parameters, connecting μ_{oD} to the P_{bp} and R_{sbp}. The Difference in bubble point pressure levels for both aromatic and paraffin oil could be caused by the same quantity of the solution gas, therefore their approach would be able to internment more aspects of the oil type. They claimed that (with consideration of the rang of the collected data and the field nature) their proposed model is more accurate and better than tested correlations of dead oil viscosity, the average relative error for μ_{oD} values of -2.86% and only 12.62% as an average absolute relative error.

Petrosky and farshad [11] developed their correlation through SAS software, which follow multiple nonlinear analysis technique, as consideration of that some functions of the model do not follow linear functional forms. The data were collected from Gulf of Mexico, 126 laboratories report of PVT analysis were used for the study.

For dead oil viscosity the correlation, which was a function of the oil gravity as API and the temperature of the reservoir, the obtained equation was:

$$\mu od = f\ (API, T)$$

$$\mu od = 2.3511*10^{7} * T^{(-2.10255)} * (\log API)^{X} \qquad (2.1)$$

where

$$X = 4.59388*(\log T) - 22.82792$$

Most of the data showed close value in between the measured data and the calculated values with average absolute relative error of -3.5 % to 12.4 %. Glaso dead oil viscosity correlation provided average absolute relative error of 25.4 % as founded by Sotton and Farshad, and the new correlation gives lower average absolute relative error of 13.5% than Glaso's correlation. This also compared to Beals and Kartoamodjo's correlation unpredicted the dead oil viscosity, and also over prediction by the modified Begs and Roinsons correlations.

For Saturated Oil Viscosity, the equation obtained was function of μod and Rs as following:

$$\mu_{ob} = A.(\mu \text{od})^{B}$$

$$where$$

$$A = 0.1651 + 0.6165 * 10^{-6.0866 * 10^{-4} * Rs} \tag{2.2}$$

$$and$$

$$B = 0.5131 + 0.5109 * 10^{-1.1831 * 10^{-3} * Rs}$$

Although the calculated values gave average absolute relative error of -3.1 % to 14.5 %, the Begs and Roinsons saturated oil viscosity model provided higher accuracy than this model which resulted with higher average absolute relative error and slightly lower standard deviation as founded by Sotton and Farshad. Another model were investigated by same data which is Chew and Connaly's model and also Kartoatmodjo's model and both overestimated the viscosity of saturated oils.

For Saturated Oil Viscosity, the equation obtained was function of μob and reservoir pressure and bubble point pressure as follow:

$$\mu_o = \mu_{\text{ob}} + 1.3449 * 10^{-3} (P - P_b) * 10^{A} \tag{2.3}$$

$$where$$

$$A = -1.0146 + 1.3322 * \log(\mu_{\text{ob}}) - 0.4876(\log(\mu_{\text{ob}}))^2 - 1.15036(\log(\mu_{\text{ob}}))^3$$

The estimated values of under-saturated oil viscosity gave average absolute relative error of -0.2 % to 2.9 %, all the data tested showed agreement between the experimental data and the estimated values. The most accurate model for this field was Vasques and Begs under-saturated oil viscosity model as concluded by Sotton and Farshad, un-prediction of the under-saturated oil viscosity was from both Beals and Kartoamodoj's models.

Oil formation volume factor below, at and above bubble point pressure is considered very important tool for reservoir engineering calculations, as far as concern of reservoir performance and management. Many of empirical correlations were also developed in this field and still continued.

Standing in [19] had developed correlation of B_o from 105 California's oil samples, the model depend on gas gravity, oil gravity, and reservoir temperature. His model is commonly used in oil and gas industry. Another correlation was published by Vasquez and Begs (1976) [16], they used 6000 laboratory measured data points under two groups defied by below or above 30 API gravity, and they related a reference pressure of the separator which was 100 psi to the gas gravity after they found that the gas gravity has direct effect on the correlation.

$$Bo = 0.972 + 0.00147\left[Rs\left(\frac{\gamma g}{\gamma o}\right)^{0.5} + 1.25T\right]^{1.175} \tag{2.4}$$

In 1990 Labedi [20] removed the effect of the total gas to oil ratio and gas gravity in his new correlation by using both pressure and temperature of the separator, for this model generating, 97 of data points were utilized from Libya crude oil with another 4 reports from Angola and 28 data set from Nigeria. The substitution of total gas oil ratio and gas gravity was because they are not likely to be estimated in the field.

For Middle East oil, Al-Marhoun [10] has published new correlation for Bo. 160 sample reports were used for the development of correlations which was the first correlation for the Middle East crude oil. Later by 1992 he published another correlation developed from global data from more than 700 reservoirs with 11728 experimental data points from all over the world.

$$B_o = 0.497069 + 8.62963\times10^{-4}(T+460) + 1.82594\times10^{-3}A + 3.18099\times10^{-6}A^2$$

$$A = R_s^{\,0.74239}\gamma_g^{\,0.323294}\gamma_o^{\,-1.20204} \tag{2.5}$$

Sulaimun, Ramli and Ademei [22] published new correlation developed by GMDH, they used 39 data set of PVT from Malaysian Crude oil. Their model provided the most accurate B_o values with AARE was 0.976%, the second estimation was provided by Petrosky correlation with AARE was 3.435 %, AARE was 14.654% from Standing, Glaso AARE was 22.767% where the minimum AARE for this data set was given by Al_Marhoun correlation with 26.342%, their developed model is:

$$\beta o = A_1 - (A_2.\gamma_o) + (A_3.\gamma_g) + (A_4.R_s) - (A_5.\gamma_g.\gamma_o) + (A_6.R_s.\gamma_o) + (A_7.R_s.\gamma_g) + (A_8.\gamma_o^2) - (A_9.\gamma_g^2) - (A_{10}.R_s) \tag{2.6}$$

Where:

A1= 1.08199630980282

A2= 0.0080557753122104

A3= 0.294012771678018

A4= 0.0000994971241197314

A5= 0.00402938539792537

A6= 9.11296341546098E – 06

A7= 0.000207478721653173

A18= 0.000133609923028163

A9= 0.111668392867066

A10= 5.24239118889319E – 08

Bubble point pressure: the first correlation was developed by Standing [19] for California's crude oil, he used 105 data set to get the graphical correlation, in 1981 he then express it in mathematical form with average error of 4.8% as function of API, gas solubility, gas gravity and temperature.

By 1958 Laster developed new correlation with 3.8% average error, his 137 data set was provided from Canada, South America and Western and Mid Continental USA, his proposed graphical correlation had a factor as function of solution gas mole fraction as long with GOR and Stock Tank properties.

In 1980 Glaso [17] used Standing model to develop new Bubble point pressure correlation for North Sea with 45 crude sample report, same parameters were utilized, and the results presented 1.28% average error.

$$P_b = 10^{1.7669+1.7447\log A-0.30218(\log A)^2}$$

$$A = \left(\frac{R_{sb}}{\gamma_g}\right)^{0.816} \times \frac{T^{0.172}}{API^{0.989}} \tag{2.7}$$

In 1993 Petrosky and Farshad [11] proposed similar model compared to Standing's model, they used 81 data set from Gulf of Maxico crude sample reports, and their modification was to get each parameter the ability to have exponent or multiplier equal to 1, they got more accurate result as 3.28% average absolute error.

$$P_b = 112.727\left(\frac{R_{sb}^{\ 0.5774}}{\gamma_g^{\ 0.8439} \times 10^x} - 12.34\right)$$

$$x = 7.916\times10^{-4} API^{1.541} - 4.561\times10^{-5} T^{1.3911} \tag{2.8}$$

2.2 Group Method of Data Handling Approach

This modeling method had been used proudly in many areas such as weather modeling, mechanical diagnostics, marketing and environment systems as metioned by Osman and Abdel-Aal [5]. Partial models is the main function in GMDH, its algorithm is applied to several component divisions.

By utilizing the least square method the coefficients of these models are estimated. Also another self-organization process for the models through the algorithm in GMDH is to be run in order to build a model with optimum configuration. It progressively change and increase the number of the components of the partial model, then the minimum value would consider as an indication of the optimum complexity of the external criterion.

During past 30 years, GMDH is developing as a method of inductive modelling and forecasting of complex systems according to Godefroy et al [6]. Hence, GMDH modelling approach has been proposed as an alternative modeling tool to forecast and proposing of the new correlations where it can avoid the restriction and limitation of the existing correlations.

As discussed by Ward Systems Group Inc [7], this method works by building and linking successive layers, it uses linear and non-linear regression of the data to create the layers in polynomial forms. The consecutive layers started by the first layer in which the best set preferred by the regression analysis of the input variables, the second layer created through another regression examination of the data in the first layer, this computing process continues till the results are getting better.

To predict an output Y, this method require to find out an approximation F function for available input variables as x = (X1, X2, X3 … Xn) where these variables has significant contribute in estimating Y value. As assumption:

Y = F ((X1, X2, X3 … Xn)

Semenov et al (2010) suggested that the multilayer algorithm function F is used as polynomial reference as follow:

$$F(x) = a_0 + \sum_{i=1}^{d} a_i x_i + \sum_{i=1}^{d}\sum_{j=1}^{d} a_{ij}\, x_i\, x_j + \ \dots$$

This function provide window to describe or simulate the outputs and inputs, GMDH also can fix any problem or trouble by simply adding two terms in the previous multilayer polynomial function.

The GMDH network structure according to Semenov et al [8] is as follow:

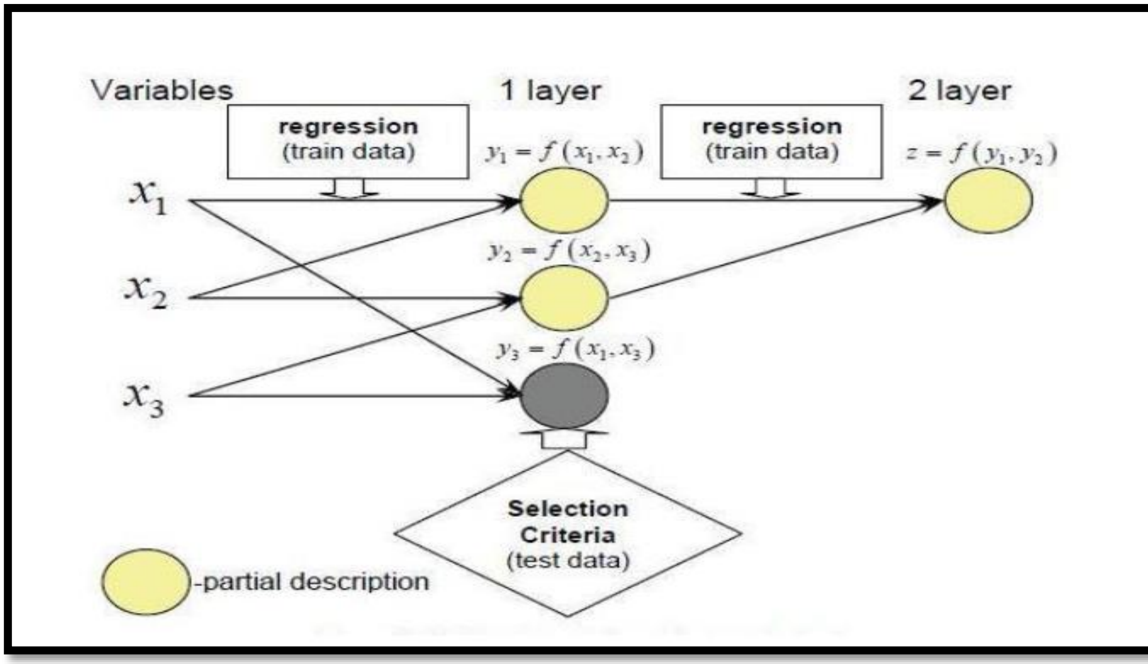


Figure 2.1: GMDH network structure

Through this self-orgnizing technique, GMDH will be able to neglect the effect of small, less accurate and noisy data, therefore higher accuracy of the model is achived. GMDH will also be able provide simple structure of the new developed model.

Group Method of Data Handling approach is not widely practiced in oil and gas industry. The litrature search founded that only few studies had been accomplished using GMDH modeling, it has been used in the prediction of permeability using well logs, Lim et al [9], in forecast of PVT properties, by Osman & Abdel-Aal [5], for the improvement of porosity prediction Semenov et al [8] and also used as a prediction of estimating tool life in drilling engineering Lee et al[21].

Chapter 3

Methodology

3.1 Research methodology:

The research methodology for the project is illustrated on figure3.1.

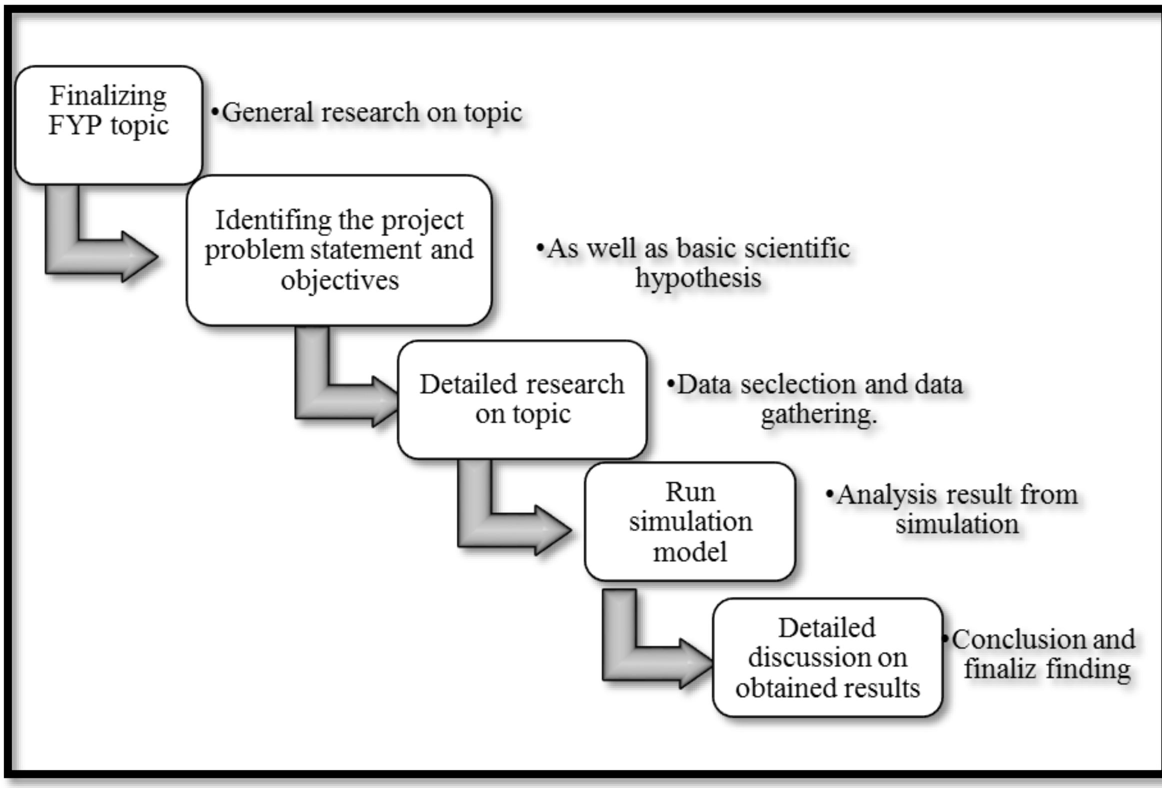


Figure 3.1: Research methodology

3.2 Project workflow:

The flow of this project is structured as illustrated in figure 2

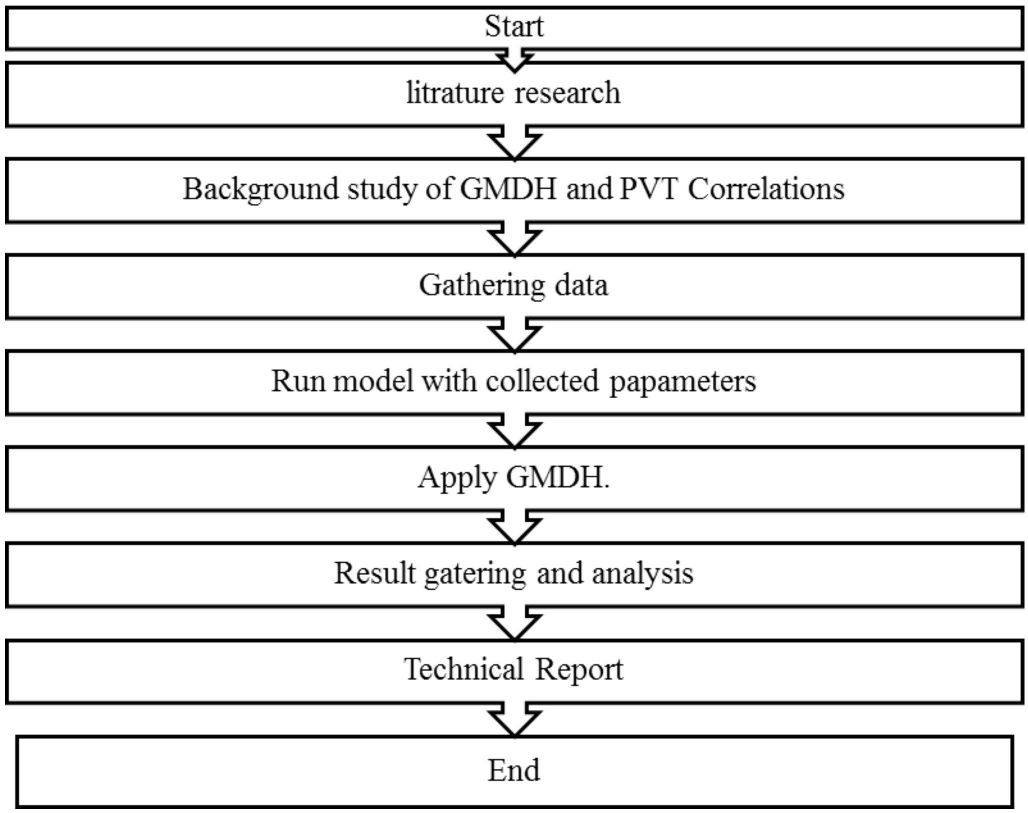


Figure 3.2: Project workflow

3.3 Key milestone:

- Submission of extended proposal (week 6)
- Proposal defense, oral presentation (week 8 and 9)
- Submission of interim draft report (week 13)
- Submission of interim report (week 14)

3.4 Gantt chart:

Table 3.3: Gantt chart for FYP

No	Detail/ week	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1	Selection of project title																												
2	Preliminary research work																												
3	Submission of extended proposal defense																												
4	Proposal defense																												
5	Project work continue																												
6	Submission of intermit draft report																												
7	Submission of intermit report																												
8	Project work commences																												
9	Submission of Progress Report																												
10	Poster Exhibition																												
11	Submission of Final Draft																												
12	Viva																												

3.5 Tools:

All of the project required tools are software base only listed in table

Table 3.4: Tools

Tool	Function
Matlab	Develop GMDH modelling
Microsoft office word	For reporting
Microsoft office excel	To prepare data sheet collection
Microsoft office power point	To prepare presentation

3.6 Software utilized:

Microsoft excel sheets were used to save, modify and arrange the collected data, it was also used to apply the published model on the data for comparison and analysis purposes. For this Project, MATLAB (version R2009b) environment was used due to flexible code programming and also graphs visualization. MATLAB provided simple monitoring of the performance of all three data sets (training, validation and testing data) simultaneously which aid the optimization process as well as sensitivity analysis. The MATLAB code was built and input parameters were revised in order to ensure that there parameters are well optimized. The code is in appendix 4.

Developing polynomial Neural Networks was brought by Group Method of Data Handling (GMDH), in which the algorithm build its network in form of layer by layer arrangement by using the training data set. It consists of an evaluation criterion that control the number of layers and the connectivity in between them. The corrected Akaike's Information Criterion or Minimum Description Length options given either to use assessing performance with validation data clearly taking network's complexity into account. The other parameters such as, Max Number Of Inputs for individual neurons, the Degree Of Polynomials in the neurons, it is also applicable to allow the neurons to take inputs from the preceding layer or it can use inputs from the original input variables, Number Of Neurons in a layer to decrease the number of neurons of the following layers.

3.7 Developing a model of GMDH

GMDH approach sets a several algorithms for different solutions of a problems. This inductive method based on building and arranging of gradually complicated models, then it select the best solution by the defined criterion characteristic. For basic models GMDH built non-linear, polynomials and probabilistic functions. Polynomial technique offer a representation of input regimes to the outputs via an application of Regularity Criterion, which is Average Absolute Percentage Error. Its objective is to reduce the error between the measured and calculated value targeted in the layers. The method also apply threshold level before adding each layer.

3.8 Trend analysis

This trend analysis proposed for GMDH models is carried out to check if its models are providing physically correct trends. Synthetic sets are built, each set have one input parameter to be given range from its minimum to its max values, meanwhile the other parameters are fixed value of choose set from the dataset. In the trend analysis parameters effects of different inputs are tested such as; temperature, gas gravity, API gravity and solution gas oil ratio are all studied.

3.9 Statistical Error Analysis

Statistical Error Analysis was used to check the accuracy of the developed models and also the other studied models. These statistical parameters used in this project are average absolute percent relative error, average percent relative error, maximum absolute percent error and minimum absolute percent error.

3.10 Graphical Error Analysis

Graphic tools assist in visualization the trend curves, accuracy and the performance of the developed model. Plots used were:

3.10.1 Cross-plots

Cross plots were used to comparison of the outputs of all the investigated models. It is a 45° straight line between the predicted Oil Formation Volume Factor versus measured Oil Formation Volume Factor plot which represent the perfect correlation accuracy line. As the values go closer to the 45° line, it indicate as closer results of the measured as to the estimated values.

3.10.2 Error Distribution

Error distribution shows the error sharing histograms for the new GMDH model of all the three datasets: training, validation and testing. Normal distribution curves had been fitted to each one of them. The normal distribution is used to describe any variable that tends to cluster around the mean. In this case it was used to describe the error tendency around the estimated mean, it is also known as Gaussian distribution. Refer to appendix2.

Chapter 4

Results and discussion

These data were used in the available PVT models which were mentioned in the literature, some results are obtained for Bubble point pressure, Pb (psia), Dead oil viscosity, µod (cp), Saturated oil viscosity, µob (cp), Under-saturated oil viscosity, µo (cp), Solution gas oil ratio, Rs (scf/STB) and Oil formation volume factor, B_o, to be compared later with the output result of the generated models from GMDH.

4.1 Data Gathering

A data set was collected to represent measured data consisting of formation volume factor, reservoir temperature, API, gas gravity, solution gas-oil ratio to generate new model of the oil formation volume factor.

The total number data collected was 220 data points from different published. Each of the papers were from different regions. To avoid replication of the data each data groups were screened for duplicates and randomized. Rang of all PVT data used to build B_o model is described as in the following table:

Table 4.1: Description of the Data Used For B_o

	B_o	° API	GG	Tr (°F)	Rs(scf/STB)
Min	1.0165	10.0000	0.0906	85.0100	2.9
Max	1.3300	22.2000	1.517	253.814	429.16

- Data selection and data gathering is considered one of the difficulties in this project due to:
  - Publications in heavy oils correlations did not have significant attention as lighter oils.
  - Mostly the released available data are out of this project scope which refer to heavy oil as in range of 10 < API < 22.3.
  - Most of published papers provide only range of the data set used.
  - Confidentiality of the data to the oil and gas companies.

4.2 GMDH model: Oil Formation Volume Factor

Set of The equations has been developed by GMDH approach to estimate oil formation volume factor, the new model has come up with the result of 2 layers of the network of all possible inputs to achieve the desired output;. The best correlation for estimating oil formation volume factor by GMDH has involved selected inputs was:

- API ° as X1.
- γ_g as X2.
- T_r (°F) as X3.
- R_s (scf/STB) as X4.

The model equations are as follow:

Number of layers: 3

Number of used input variables: 4

Execution time: 2.52 seconds

Layer #1

Number of neurons: 1

x5=0.798657345309226+0.00133160720995408*x4+0.00169956363066512*x3+0.27385250500804*x2-1.44287093334376e-06*x3*x4-9.13564419605414e-05*x2*x4

-0.0019171472010573*x2*x3-1.64243542639056e-06*x4*x4+

2.34324112686453e-06*x3*x3 -0.0419930241820582*x2*x2

Layer #2

Number of neurons: 1

x9=-0.970751785028963+3.42173773602385*x5+1.83210680628967*x2

0.14571899825903*x1-1.69158713713385*x2*x5+0.137626254974143*x1*x5

+0.00525168829206646*x1*x2-1.41916834844809*x5*x5

0.0369540001188934*x2*x2 -0.000326058875175248*x1*x1

Layer #3

Number of neurons: 1

y=5.34533921204889-10.6692712819784*x9+0.00529794350033261*x4+ 0.0120118206627359*x3-0.00557241595723045*x4*x9-0.0135637845020399*x3*x9+ 2.91817138610996e-06*x3*x4 + 6.40317017736758*x9*x9+ 2.03064344807061e-06*x4*x4+7.94794993577465e-06*x3*x3

4.3 Bo Statistical Error Analysis

Oil formation volume factor B_o: Vasquez-Begs model also gave good estimation for B_o with only 4.59% average error, the new built correlation gave 1.2675 %average absolute percentage relative error.

The statistical error analysis parameters used are; minimum and maximum absolute percent error, average absolute percent relative error and the correlation coefficient R^2 as in table 4.2.

Table 4.2: Statistical Error Analysis

Statistical Error Analysis	This project model	Standing correlation	Vazquez & Beggs correlation	Glaso correlation	Al-Marhoun correlatio
Minimum Absolute Percent Relative Error, Emin	0.0022	0.0509	0.022	0.015	0.074
Maximum Absolute Percent Relative Error, Emax	4.9936	16.950	23.273	19.145	16.35
Average Absolute Percent Relative Error, AAPRE	1.2675	4.751	4.590	5.28	5.368
Square of Correlation of Coefficient, R^2	0.99124	0.658	0.717	0.6365	0.686

4.4 Graphical Error Analysis

4.4.1 Cross-plots

Following is the distribution of the data estimated by GMDH and other models, all models results are shown in the following Cross plot of Measured vs. Predicted B_o:

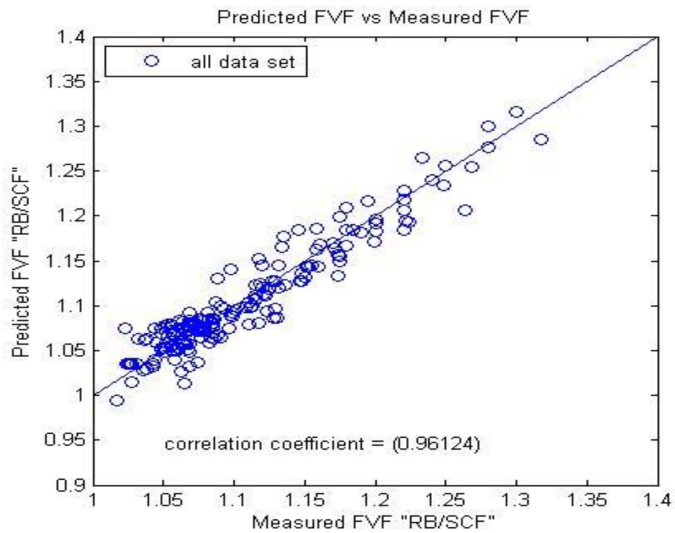


Figure 4.3 Cross plot of Measured vs. Predicted B_o by GMDH Model

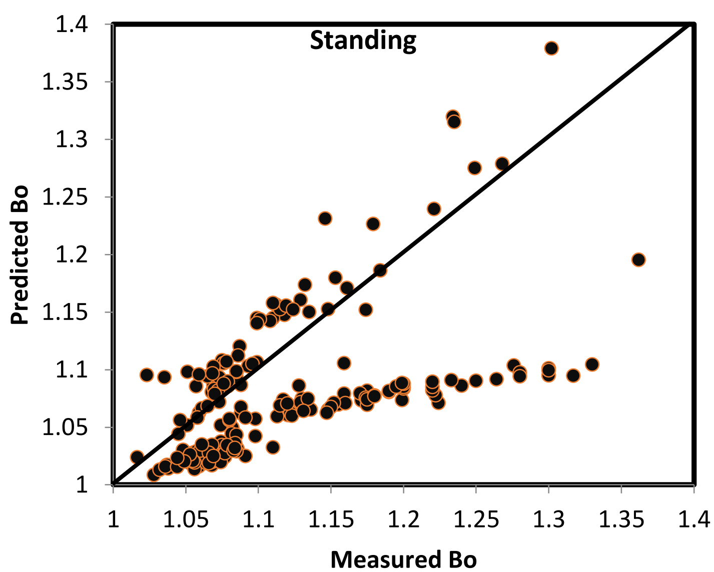


Figure 4.4 Cross plot of Measured vs. Predicted B_o by Standing model

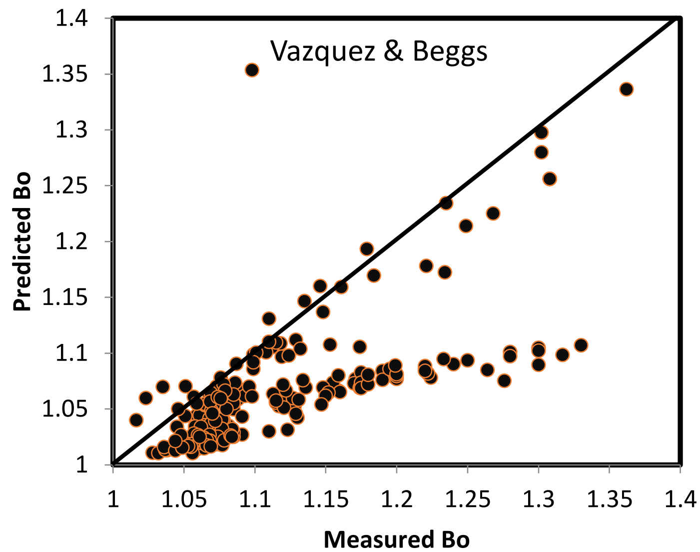


Figure 4.5 Cross plot of Measured vs. Predicted Bo by Vazquez & Beggs Model

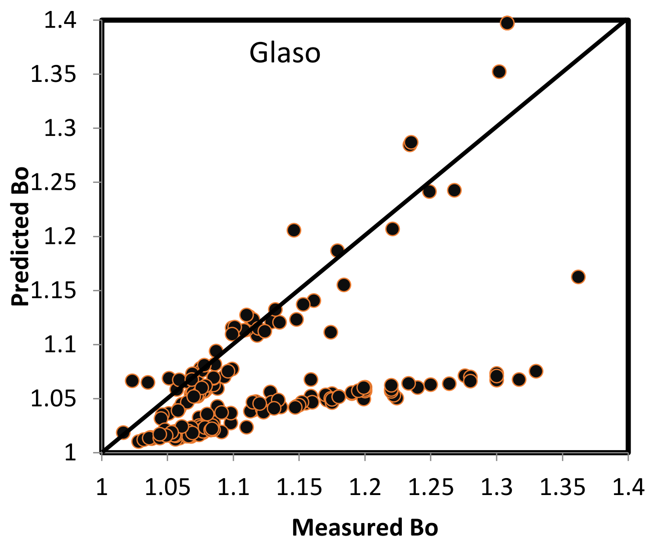


Figure 4.6 Cross plot of Measured vs. Predicted B_o by Glaso Model

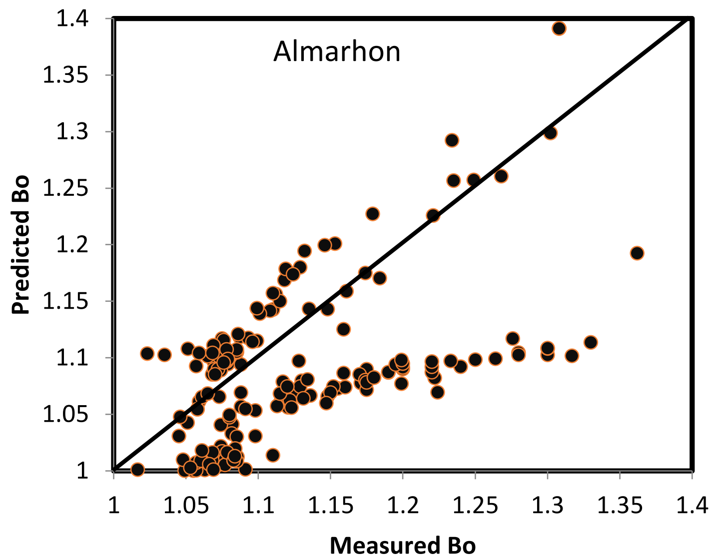


Figure 4.7 Cross plot of Measured vs. Predicted B_o by Almarhon Model

4.4.2 Simulation trend analysis:

Simulation trend analysis were carried out to check if its models are providing physically correct trends. While changing one parameter fixing all the other parameters the result shows good results as in following figures:

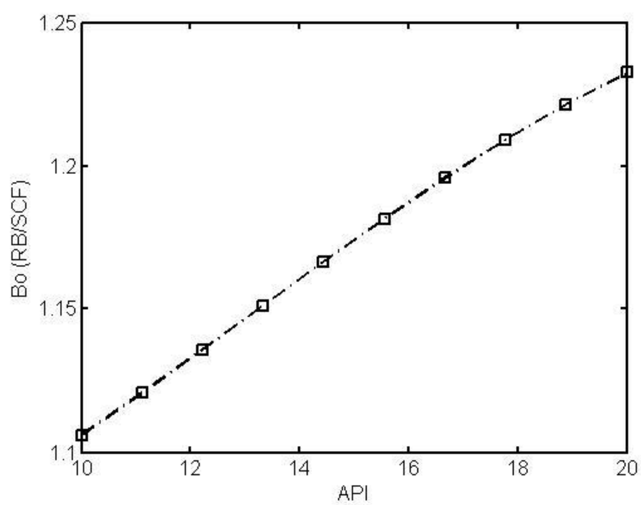


Figure 4.8 Variation of B_o with API

- API gravity describe in gravity terms how heavy or light the oil is, for heavy oil we get less amount of volume of oil compared to lighter oil. Therefore as long as the API gravity increase the Bo increase as well, the model provided truth physical trend of API variation to Bo.

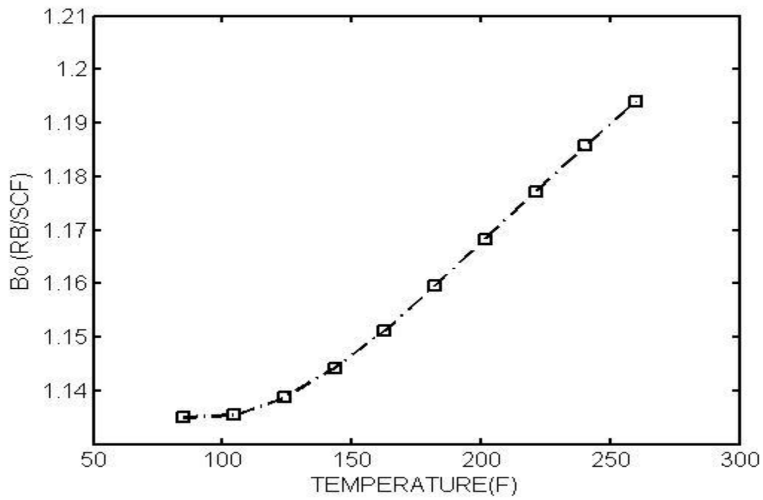


Figure 4.9 Variation of B_o with Temperature

- Temperature trend is also physically correct, as heated the oil is under the reservoir pressure, it give higher B_o.

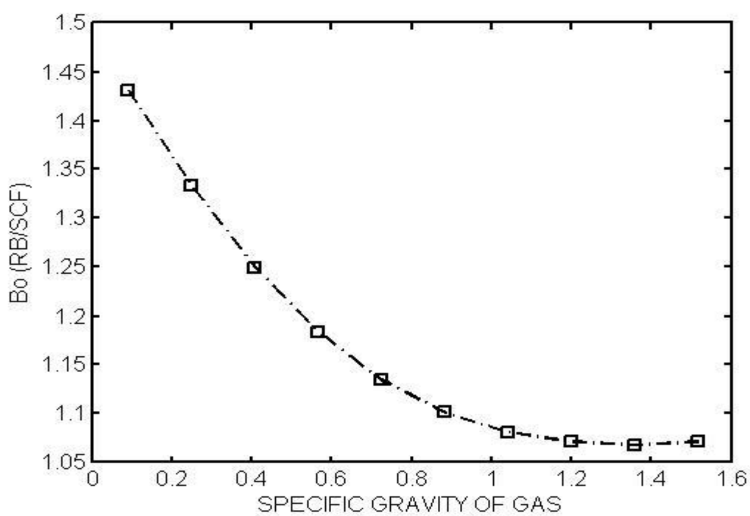


Figure 4.10 Variation of B_o with GG

- The trend was not at best for specific gravity of gas. It gives same trend as resulted by Vazquez & Beggs correlation. This due to the difference of reservoir properties, some data set has low B_o while its gas gravity is high, such data cause uncertainty in calculation.

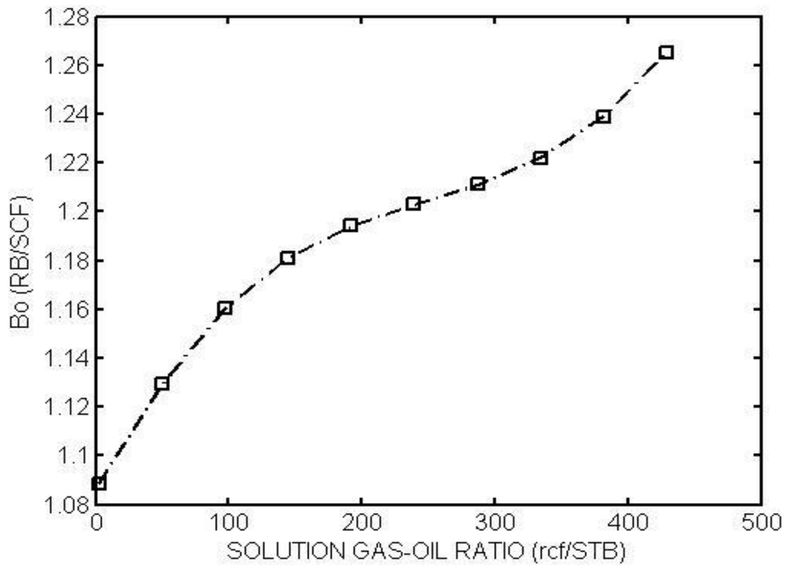


Figure 4.11 Variation of B_o with Rs

- The solution gas oil ratio related to how much gas is soluble in the oil and not free, the oil will get expand as it arrives to surface cause this gas was pressurized in the reservoir, that increase in volume represented by B_o, as Rs increases the oil formation volume factor increases.

4.4.3 Error Distribution

Figures 4.12, 4.13 and 4.14 show below are the error distribution histograms for GMDH Model data sets.

Analysis of error distribution is important, it provides clear indication about the new model presentation of the datasets. From the estimated results, all sets have acceptable distribution without any clear moving towards positive or negative ends, therefore pointing out a good estimation.

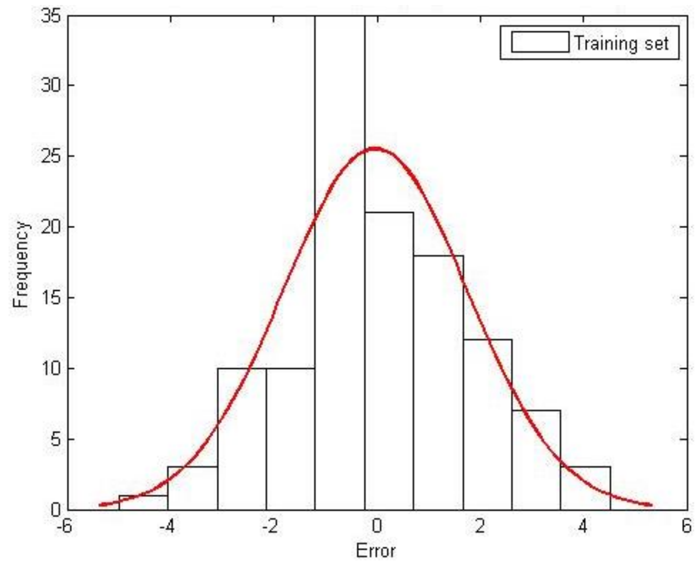


Figure 4.12 Error Distribution for Training set

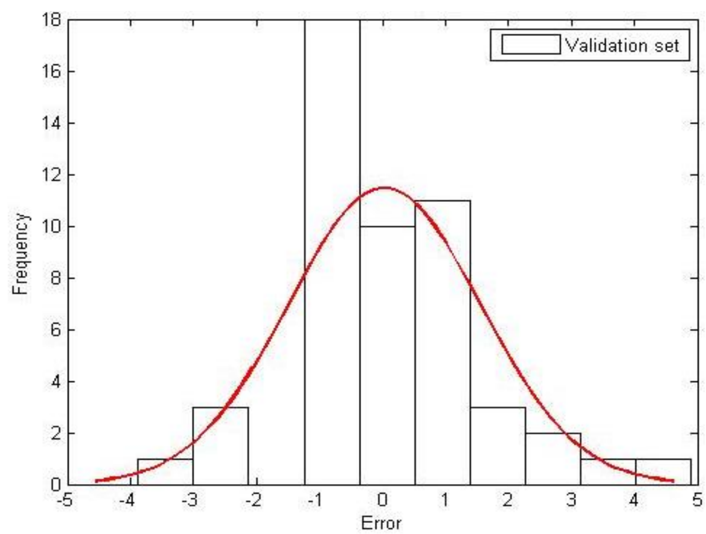


Figure 4.13 Error Distribution for validation set.

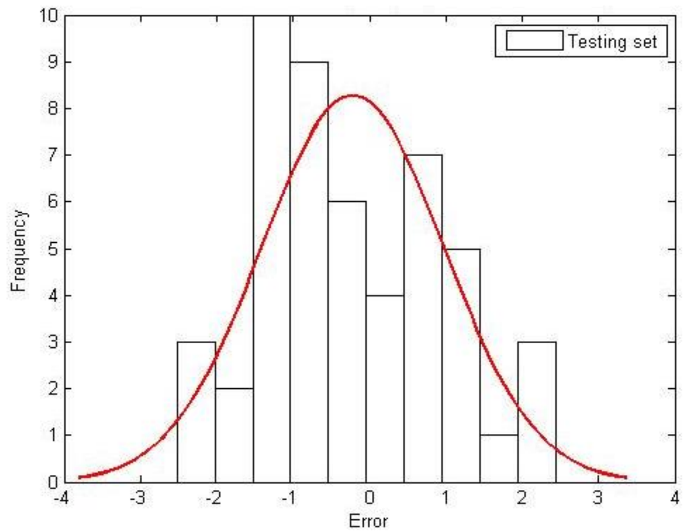


Figure 4.14 Error Distribution for testing set.

4.5 Other current models

4.5.1 GMDH model: Bubble point pressure

Set of equations has been developed by GMDH approach for predicting the Bubble point pressure, The model came up with the model with 1 layer of the network of all possible inputs to achieve the desired output;. The best correlation for estimating Bubble point pressure by GMDH has involved selected inputs was:

- API ° as X1.
- GG as X2.
- Tr (°F) as X3.
- Rs (scf/STB) as X4.

- Number of used input variables: 2 (Rs and GG)
- Execution time: 0.17 seconds
- Number of layers: 1
- Layer #1
- Number of neurons: 1

$y = -12464.083375 + 87.5012165745*x4 + 20663.5629038*x2 - 104.065584294*x2*x4 - 0.0831240503581*x4*x4 - 9205.09410135*x2*x2 + 0.0497301529539*x2*x4*x4 + 32.3464237869*x2*x2*x4 + 2.62933696006e-05*x4*x4*x4 + 626.60257929*x2*x2*x2$

- **P_b Statistical Error Analysis**

Bubble point pressure P_b:

The statistical error analysis parameters used are; average absolute percent relative error, minimum and maximum absolute percent error and the correlation coefficient as in table 4.3.

Table 4.3: Statistical Error Analysis

Statistical Error Analysis	This Study	standing	Kartoatmodjo and Schmidt	Glaso	De gattoh
Minimum Absolute Percent Relative Error, Emin	0.0518	1.2398	1.127	0.789	1.23
Maximum Absolute Percent Relative Error, Emax	79.7868	48.78211	83.95	103.04	61.168
Average Absolute Percent Relative Error, AAPRE	14.7547	17.47773	25.2058	36.04	28.08
Square of Correlation of Coefficient, R^2	0.97	0.93	0.918	0.916	0.93

The best estimation for this P_b was from Standing’s correlation with average error 17.4%, however most of the other correlations could not predict or over estimated it. Kartoatmodjo model provide 25.0% absolute average error for representative samples.

- P_b **Simulation trend analysis**

The generated model only utilized 2 input variables (R_s and GG), the uncertainty and the small number of data sets used (only 53 datasets) caused the new model not to give all trend for all variables the P_b depends on it. Only two trends were obtained as following in the following figures:

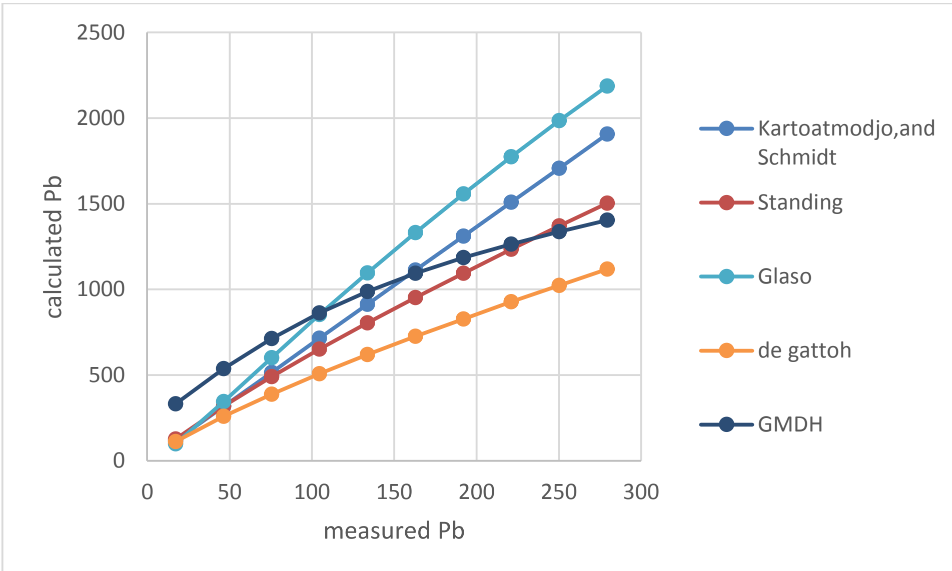


Figure 4.15 Variation of P_b with R_s

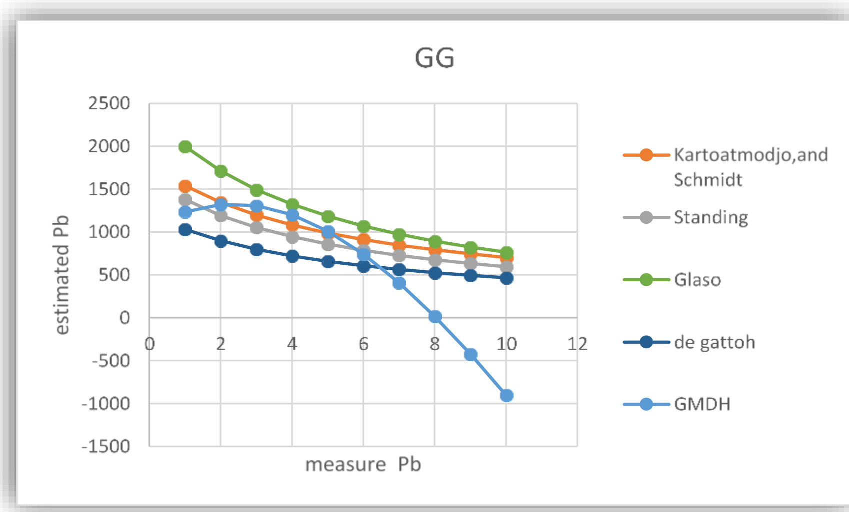


Figure 4.16 Variation of P_b with γ_g

The model provides wrong estimation of P_b with negative values, if more data were added the performance and the model could be changed and gave the correct trends.

There for this model is not valid for now.

4.5.2 Dead oil viscosity

This was mostly unpredicted with most of the models, only Egbogah – Jack resulted in 41.89% average error, for less than 10 representatives samples Kartoatmodjo model result in high average error of 34.4 %. The justification for the high error related to that this property model only depend on the reservoir temperature and the API. This property measurement is also difficult to obtain from the labs. Only 48 data sets were gathered and doesn’t provide acceptable models.

4.5.3 Saturated oil viscosity

The best estimations was from Kartoatmodjo correlation wiith 16.1% average error. Chew-Connally model also gave good prediction of 17.8% average error and 15.7 SD. Only 64 data sets were gathered and doesn’t provide acceptable models.

4.5.4 Under-saturated oil viscosity

Kartoatmodjo model delivered its best estimations with average error of 10.1%. Only 43 data sets were gathered and doesn’t provide acceptable models.

Chapter 5

Conclusion

5.1 Conclusion

Most of the existing correlations do not give accurate estimations, due to some assumption in their derivation to simplify the correlation and also each correlation was derived for specific field. Furthermore only few modification were made to some correlation to be applied on heavy oil. There is not enough attention for building models of PVT for heavy oil in practice as reviewed by literature.

GMDH as a suitable predictive tools should be used when experimentally measured values are not obtainable, this is because of some inappropriate sampling technique or in most cases the expensive PVT test. (GMDH) is introduced in this project as an alternative modeling approach that helps to overcome the limitations. GMDH had been used in previous similar studies to develop Bo.

New model for oil formation volume factor is generated from GMDH approach, it has provided good results in estimating oil formation volume factor of heavy oil. The new model had square of correlation of coefficient, R^2 of 0.99 with average absolute percent relative error, AAPRE of 1.27%.

Another set of models were established for Bubble point pressure, Pb (psia), Dead oil viscosity, µob (cp), saturated oil viscosity, µod (cp), Under-saturated oil viscosity, µo (cp) and Solution gas oil ratio, Rs (scf/STB), but due to the small number of datasets gathered for each property the method was not generate correct model with the available data.

References:

[1] G. De Ghetto and M. Villa, "Reliability Analysis on PVT Correlations," presented at the *European Petroleum Conference*, Londen, U.K., 1994.

[2] M. S. Hossain, C. Sarica, H. Q. Zhang, L. Rhyne, and K. L. Greenhill, "Assessment and Development of Heavy Oil Viscosity Correlations," 2005.

[3] J. T. H. Ng and E. O. Egbogah, "An Improved Temperature-Viscosity Correlation For Crude Oil Systems," vol. 5, pp. 197-200, 1983/1/1/ 1983.

[4] B. Dindoruk and P. G. Christman, "PVT Properties and Viscosity Correlations for Gulf of Mexico Oils," 2001.

[5] Osman, E., & R. E. Abdel-Aal, "Abductive Networks: A New Modeling Tool for the Oil and Gas Industry," presented at the *SPE Asia Pacific Oil and Gas Conference and Exhibition*, Melbourne, Australia., 2002.

[6] S. N. Godefroy, S. H. Khor, and D. Emms, "Comparison and Validation of Theoretical and Empirical Correlations for Black Oil Reservoir Fluid Properties," presented at the *Offshore Technology Conference*, Houston, Texas, 2012.

[7] Ward Systems Group Inc. (2008). *Network Architecture - GMDH*. Retrieved April 9, 2013, from NeuroShell2 Help: http://www.wardsystems.com/manuals/neuroshell2/

[8] Semenov, A., Oshmarin, R., Driller, A., & Butakova, A. Application of Group Method of Data Handling for Geological Modeling of Vankor Field. SPE 128517 at SPE North Africa Technical Conference and Exhibition, February 14-17. Cairo, Egypt. 2010.

[9] Lim, J.-S., Park, H.-J., & Kim, J. (2006). A New Neural Network Approach to Reservoir Permeability Estimation from Well Logs. *SPE 100989 presented at the SPE Asia Pacific Oil & Gas Conference and Exhibition.* Adelaide, Australia, September 11-13. 2006: Society of Petroleum Engineers.

[10] AI-Marhoun M.A.: "PVT Correlations for Middle East Crude Oils," JPT (May 1988), pp 650-66.

[11] Petrosky, G.E. and Farshad, F.F, 1993. Pressure-Volume-Temperature Correlations for Gulf of Mexico Crude Oils. SPE 26644 presented at the 68th Annual Technical Conference and Exhibi- tion of the Society of Petroleum Engineers, Houston, Texas, 3–6 October

[12] Lohrenz, J. and Bray, B.G. “Calculating Viscosities of Reservoir Fluids from Their Compositions,” Journal of Petroleum Technology (1964), 1171

[13] Beal, C. “The Viscosity of Air, Water, Natural Gas, Crude Oils and Its Associated Gases at Oil Field Temperature and Pressures,” SPE Reprint Series No. 3, Oil and Gas Property Evaluation and Reserve Estimates, Society of Petroleum Engineers of AIME, Dallas, Texas (1977), 114.

[14] Chew, J. and Connally, C.A. Jr. “A Viscosity Correlation for Gas-Saturated Crude Oils,” Transactions AIME (1959)216, 23.

[15] Beggs, H.D. and Robinson, J.R. “Estimating the Viscosity of Crude Oil Systems,” Journal of Petroleum Technology, (1975), 1140.

[16] Vazquez, M.E. and Beggs, H.D. “Correlation for Fluid Physical Property Prediction,” Journal of Petroleum Technology (1980), 968.

[17] Glaso, O. “Generalized Pressure-Volume-Temperature Correlations” Journal of Petroleum Technology (1980), 785.

[18] Kartoatmodjo, R.S.T. and Schmidt, Z. “Large Databank Improves Crude Physical Property Correlations,” Oil and Gas Journal (1994), 51. 14.

[19] Standing, M.B. “A Pressure Volume-Temperature Correlation for Mixtures of California Oils and Gases,” Drilling and Production Practices, API (1947), 275.

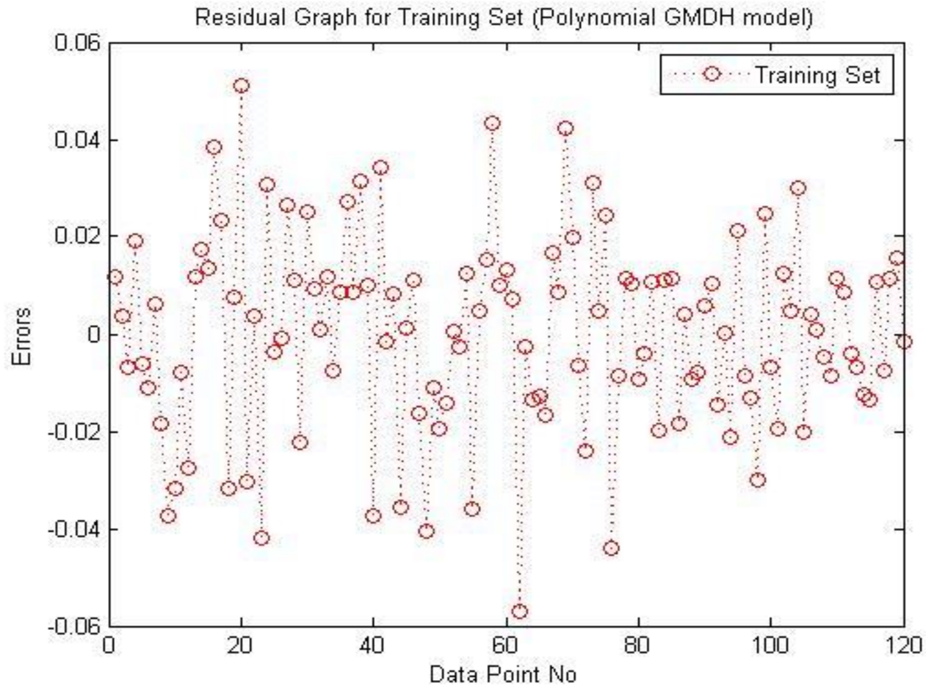
[20] Labedi, R. “Improved Correlations for Predicting the Viscosity of Light Crudes,” Journal of Petroleum Science and Engineering (1992)8, 221.

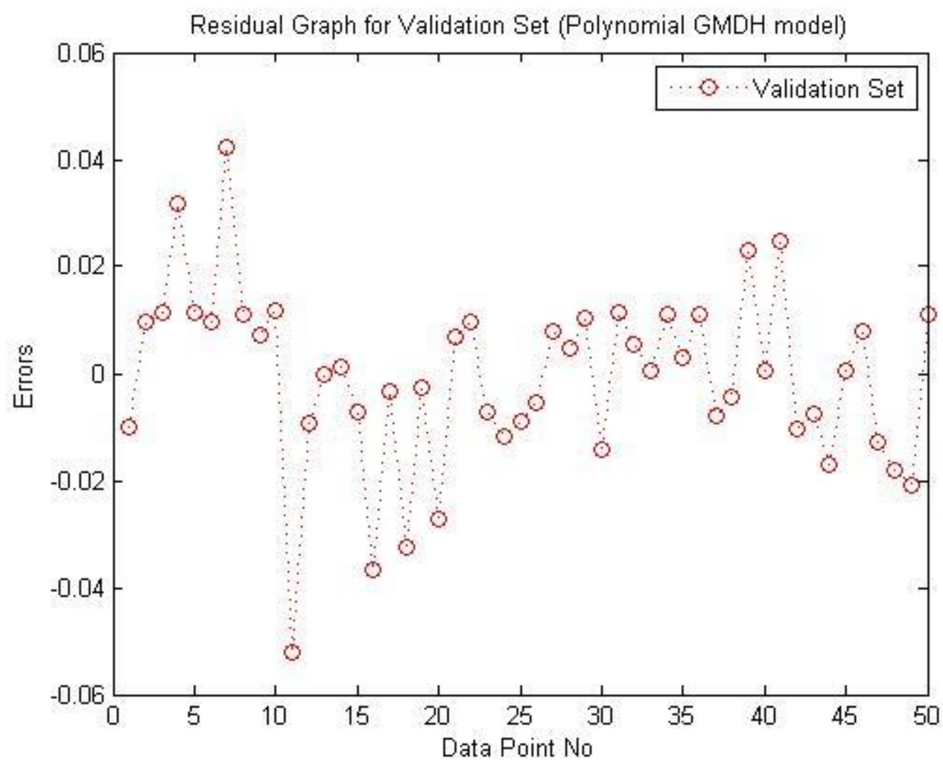
[21] Lee, B., Liu, H., & Tarng, Y. (1995). An Abductive Network for Predicting Tool Life in Drilling. *IEEE Transactions on Industry Application.*

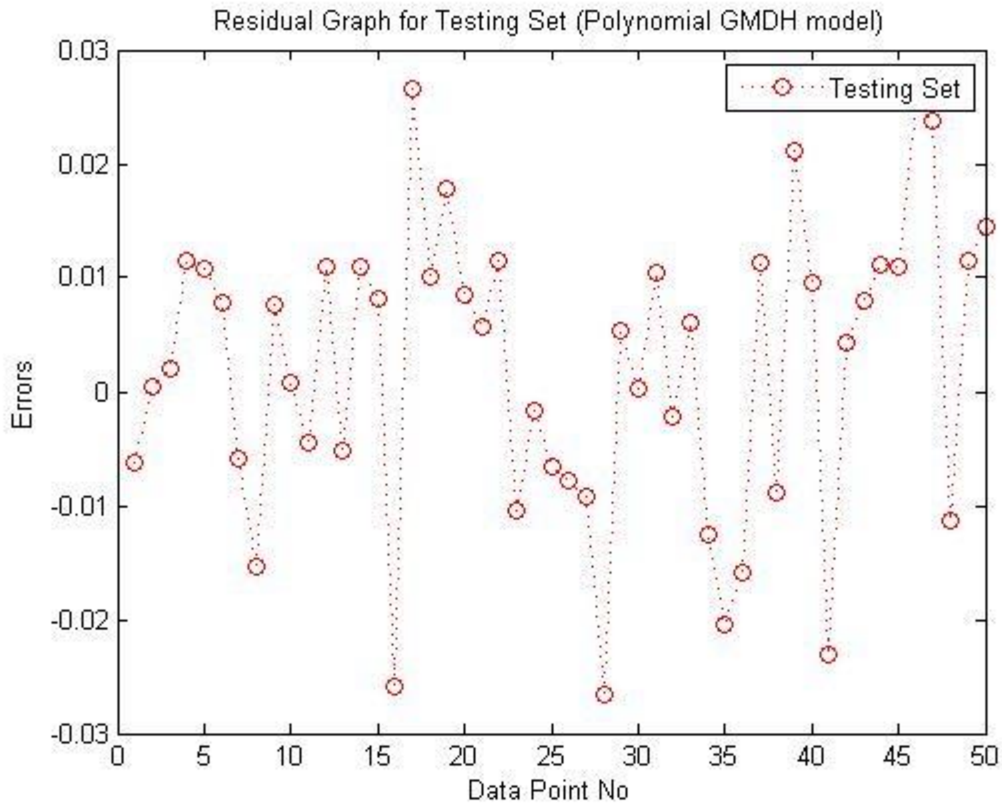
[22] Sulimon, A.A., Ramli, N. and Adeyemi, B. J. " New correlation for Oil Formation Volume Factor" presented at the S*PE Nigeria Annual International Conference and Exhibition*, Lagos, Nigeria, 2014.

.

Appendix 1: Error Distribution for Bo model






Residual Graph for Testing Set (Polynomial GMDH model)
Testing Set
Errors
Data Point No
0.03
0.02
0.01
0
-0.01
-0.02
-0.03
0
5
10
15
20
25
30
35
40
45
50

Appendix 2: Bo data used

	x1	x2	x3	x4
OFVF	° API	GG	T (°F)	R_s(scf/STB)
1.0237	20.0000	0.7400	140.0000	5.0000
1.0930	15.2000	1.0640	214.0000	54.1300
1.0640	13.2000	0.6600	105.9800	17.9000
1.2800	21.7000	0.7500	170.0600	110.8300
1.1100	19.7000	1.3360	170.6000	186.4500
1.1219	21.8000	0.7500	177.8000	16.5800
1.0694	15.7000	1.1664	187.8000	42.6407
1.0580	20.2000	1.4660	157.3000	24.5747
1.0740	17.3000	0.7410	89.9960	28.2300
1.1100	15.0000	0.8500	111.2000	31.1700
1.1205	14.1000	0.7700	149.0000	52.0000
1.2220	21.7000	0.7500	170.0600	63.7500
1.0238	20.0000	0.7400	140.0000	5.0000
1.0870	19.8000	1.2560	150.8000	147.9600
1.0450	19.4000	1.2510	120.0000	39.0000
1.1460	16.8000	1.5170	140.0000	320.3400
1.0690	12.8000	1.3230	215.6000	17.2100
1.3170	21.7000	0.7500	170.0600	104.2900
1.1150	15.0000	0.7800	161.6000	54.0000
1.0231	18.1000	1.2560	187.8000	63.3369
1.2240	21.5000	0.7490	143.0060	85.5000
1.1590	14.4000	0.8000	253.8140	13.3600
1.1280	20.0000	1.2200	210.0200	22.6200
1.0320	9.5000	0.6300	95.0000	2.9000
1.1150	18.2000	0.7500	177.9800	13.4600
1.0860	16.5000	1.1880	188.1000	97.3200
1.0512	21.6000	1.1850	187.8000	69.6347
1.0245	20.0000	0.7400	140.0000	5.0000
1.0165	21.9000	0.5610	100.0000	29.0000
1.1200	21.1000	0.7500	188.6000	21.0600
1.0760	22.2000	1.1720	186.0000	85.6877
1.0760	19.2000	1.1430	187.8000	51.3941
1.0239	20.0000	0.7400	140.0000	5.0000
1.0570	14.5000	0.6500	122.0000	8.3700
1.0271	20.0000	0.7400	140.0000	5.0000
1.1590	21.1000	0.7500	188.6000	41.4500

1.0690	17.4000	1.2540	187.8000	76.8007
1.0440	10.8000	0.7500	96.0800	8.2700
1.0254	20.0000	0.7400	140.0000	5.0000
1.2210	19.5000	1.1690	178.7000	332.6100
1.1180	19.9000	1.0050	231.8000	121.4600
1.1290	19.6000	1.0920	231.8000	140.5200
1.0550	13.9000	0.8200	100.0400	15.0000
1.0620	19.3000	0.7170	114.0080	19.4100
1.0740	13.0000	0.7300	135.0140	12.5000
1.0570	21.4895	1.1850	183.0000	51.2772
1.0850	14.0000	0.7000	111.2000	25.0000
1.1740	19.5000	1.0590	240.8000	115.9800
1.1310	18.2000	0.7500	177.9800	19.6880
1.0838	14.0000	0.8000	125.6000	11.0000
1.1120	19.4000	1.4110	172.4000	177.8300
1.1270	17.9000	0.7500	179.9600	23.0880
1.0510	13.8000	0.7100	111.2000	8.0150
1.0682	9.5000	0.6600	134.0600	7.0000
1.1170	19.7000	1.0150	197.9960	13.7100
1.1190	18.8000	1.2060	244.4000	111.7600
1.0560	9.5000	0.6700	95.0000	4.2000
1.0980	21.1000	0.0906	145.9940	27.2500
1.0650	15.4000	1.2570	187.0000	62.4413
1.0530	10.7000	0.6500	100.0400	12.7400
1.0520	12.0000	0.6800	105.9800	10.1500
1.2640	17.9000	0.7500	179.9600	84.0930
1.2200	21.7000	0.7500	170.0600	73.8000
1.2680	17.6000	0.9340	194.0000	429.1600
1.0769	15.0000	0.6700	111.2000	27.6000
1.2000	17.9000	0.7500	179.9600	64.8700
1.3000	21.7000	0.7500	170.0600	119.3000
1.0680	19.2000	1.1870	187.8000	61.0487
1.1350	12.4000	0.7140	152.6000	269.9000
1.0490	11.0000	0.7000	116.0600	5.9000
1.0840	14.4000	0.7800	122.0000	26.7000
1.0829	14.4000	0.7400	120.2000	13.0000
1.1340	21.1000	0.7500	188.6000	31.1800
1.0307	20.0000	0.7400	140.0000	5.0000
1.1750	21.1000	0.7500	188.6000	47.9300

1.1300	15.0000	1.0150	198.0140	15.4550
1.0591	11.0000	0.6300	100.0400	3.9000
1.0630	11.8000	0.7700	110.0120	13.3600
1.0252	20.0000	0.7400	140.0000	5.0000
1.0850	14.5000	1.2920	187.8000	68.2028
1.0640	14.0000	0.7200	122.0000	11.2210
1.0248	20.0000	0.7400	140.0000	5.0000
1.1750	18.2000	0.7500	177.9800	44.4400
1.0244	20.0000	0.7400	140.0000	5.0000
1.0240	20.0000	0.7400	140.0000	5.0000
1.0580	20.2000	1.4660	157.3000	24.5747
1.0680	14.0000	1.2950	183.2000	40.9700
1.0980	21.7000	0.7500	170.0600	13.3500
1.0610	19.4000	1.2510	160.0000	39.0000
1.0780	15.1000	1.3440	207.7000	25.2100
1.0816	11.8000	0.6600	143.6000	20.0000
1.2490	21.2000	1.0620	183.2000	404.0100
1.1840	10.9000	0.8100	154.2000	331.3400
1.1480	21.7000	0.7500	170.0600	32.9200
1.1950	21.1000	0.7500	188.6000	56.0000
1.0380	14.2000	0.7100	95.0000	5.8800
1.2200	18.2000	0.7500	177.9800	82.0830
1.1230	9.9000	0.7200	181.4000	4.8000
1.0370	10.3000	0.6500	102.2000	3.9200
1.1710	21.8000	0.7500	177.8000	42.3000
1.0682	14.7000	0.7600	100.0400	15.0000
1.1320	19.8000	1.3470	244.0000	135.4700
1.0307	20.0000	0.7400	140.0000	5.0000
1.1790	16.0000	0.7840	211.3000	338.0000
1.1470	13.2000	0.6700	158.0000	47.0000
1.0480	13.2000	0.7200	110.0120	6.7700
1.0740	21.0000	1.0970	184.3000	44.8790
1.0440	14.3000	0.6900	114.0800	3.8500
1.0660	13.7000	0.7900	100.0400	11.6000
1.0242	20.0000	0.7400	140.0000	5.0000
1.0270	20.0000	0.7400	140.0000	5.0000
1.0850	10.9000	0.7700	150.8000	5.6000
1.0650	21.8000	1.1230	160.0000	45.0000
1.1480	10.5000	0.8150	152.6000	260.0000

1.0280	14.4000	0.7000	85.0100	3.6000
1.0248	20.0000	0.7400	140.0000	5.0000
1.1150	19.0000	1.2920	163.4000	188.8200
1.0241	20.0000	0.7400	140.0000	5.0000
1.0560	10.7000	0.6500	100.0400	17.8000
1.0590	19.5000	1.1050	167.0000	25.7300
1.1550	21.8000	0.7500	177.8000	33.3700
1.0260	20.0000	0.7400	140.0000	5.0000
1.0241	20.0000	0.7400	140.0000	5.0000
1.2330	21.7000	0.7500	170.0600	94.8400
1.0240	20.0000	0.7400	140.0000	5.0000
1.0258	20.0000	0.7400	140.0000	5.0000
1.0880	21.1000	0.7500	188.6000	13.2200
1.0244	20.0000	0.7400	140.0000	5.0000
1.0770	14.9000	1.3070	207.9000	25.0400
1.0236	20.0000	0.7400	140.0000	5.0000
1.0650	21.9000	0.7670	112.0100	21.7300
1.1530	17.0000	1.2320	250.7000	146.4000
1.0730	14.4000	0.8900	96.9800	19.7700
1.0730	16.7000	1.2630	187.8000	59.8322
1.0785	12.9000	0.9000	131.0000	9.1000
1.0690	19.6000	0.7120	114.0080	24.5800
1.2800	21.8000	0.7500	177.8000	91.9620
1.1290	12.7500	0.8400	183.2000	30.0000
1.0750	13.2000	0.7400	128.0120	14.6000
1.1990	21.7000	0.7500	170.0600	53.6750
1.1610	11.4000	0.7760	153.1000	305.8000
1.0570	12.0000	0.7200	112.0100	10.7000
1.0990	19.2000	1.4020	165.2000	166.3300
1.0990	19.5000	1.4170	177.8000	145.1800
1.0850	18.2000	1.2440	187.8000	69.5853
1.1010	19.3000	1.4060	154.4000	175.4400
1.0910	14.8000	0.7700	174.2000	11.0000
1.0307	20.0000	0.7400	140.0000	5.0000
1.0251	20.0000	0.7400	140.0000	5.0000
1.1130	21.7000	0.7500	170.0600	18.8800
1.0242	20.0000	0.7400	140.0000	5.0000
1.0680	12.0000	0.6800	86.0000	28.0000
1.0850	14.6000	1.1780	205.9000	41.9200

1.0245	20.0000	0.7400	140.0000	5.0000
1.0820	11.3000	0.7800	158.0000	12.0000
1.0243	20.0000	0.7400	140.0000	5.0000
1.2000	18.2000	0.7500	177.9800	71.0000
1.2000	17.9000	0.7500	179.9600	74.7500
1.0610	10.0000	0.6900	127.4000	18.0000
1.1200	17.9000	0.7500	179.9600	17.8300
1.0380	11.0000	0.7000	100.0400	4.8100
1.1080	19.8000	1.3330	163.4000	167.8900
1.0360	13.9000	0.6800	100.0400	3.6000
1.1740	17.9000	0.7500	179.9600	45.1000
1.2400	21.7000	0.7500	170.0600	83.9400
1.0684	15.3000	1.2480	187.8000	66.5481
1.1800	18.2000	0.7500	177.9800	52.9150
1.1500	18.2000	0.7500	177.9800	28.1500
1.0960	20.5000	1.2880	186.0000	81.2956
1.0243	20.0000	0.7400	140.0000	5.0000
1.0780	19.2000	1.4120	158.0000	109.9300
1.0550	14.0000	0.7500	120.0200	7.5000
1.0510	21.8000	1.1230	130.0000	45.0000
1.0590	18.2000	1.4310	187.8000	60.4344
1.0248	20.0000	0.7400	140.0000	5.0000
1.0560	11.0000	0.6400	100.0400	14.0700
1.1240	19.0000	1.1720	238.3000	113.7000
1.0670	13.0000	0.7000	104.0000	10.0000
1.2200	21.8000	0.7500	177.8000	70.2040
1.0750	18.6000	1.3530	187.8000	83.3488
1.0790	21.8000	1.1230	190.0000	45.0000
1.0246	20.0000	0.7400	140.0000	5.0000
1.0770	13.8000	0.8100	102.2000	22.0000
1.0246	20.0000	0.7400	140.0000	5.0000
1.0770	10.0000	0.8080	116.0600	12.2000
1.1750	21.7000	0.7500	170.0600	43.5060
1.0530	9.5000	0.7100	118.0400	7.1250
1.0254	20.0000	0.7400	140.0000	5.0000
1.0642	11.0000	0.8000	100.0400	11.2000
1.0271	20.0000	0.7400	140.0000	5.0000
1.0660	12.0000	0.7300	116.0600	13.5400
1.0241	20.0000	0.7400	140.0000	5.0000

1.0430	14.6000	0.7200	100.0400	7.1300
1.0730	13.1000	0.9200	118.4000	8.9000
1.0600	14.7000	0.7200	123.0080	9.0800
1.0430	13.7000	0.6800	100.0400	6.9500
1.1900	18.2000	0.7500	177.9800	62.6440
1.0910	14.9000	0.7400	100.0400	28.5000
1.0480	15.3000	0.7710	122.0000	11.2200
1.1700	17.9000	0.7500	179.9600	54.9900
1.0251	20.0000	0.7400	140.0000	5.0000
1.0790	21.7000	1.1340	187.0000	56.3485
1.2500	21.8000	0.7500	177.8000	82.7130
1.1360	21.8000	0.7500	177.8000	23.0700
1.0880	20.7000	1.2260	184.0000	51.4706
1.1600	18.2000	0.7500	177.9800	36.6800
1.0242	20.0000	0.7400	140.0000	5.0000
1.1520	17.9000	0.7500	179.9600	34.7100
1.0550	10.0000	0.6900	105.0800	10.5000
1.0260	20.0000	0.7400	140.0000	5.0000
1.0855	14.5000	0.7500	122.0000	14.9500
1.1800	21.8000	0.7500	177.8000	51.2500
1.0720	15.4000	1.2760	203.0000	21.4900
1.0244	20.0000	0.7400	140.0000	5.0000
1.0246	20.0000	0.7400	140.0000	5.0000
1.0490	10.9000	0.7500	105.0800	8.3700
1.0550	10.0000	0.7200	113.0000	9.4400
1.0680	14.8000	0.7400	120.2000	13.1800
1.0240	20.0000	0.7400	140.0000	5.0000
1.0690	10.5000	0.8200	113.0000	10.1000

Appendix 3: GMDH code

```
clc;
% Final Year Project
% Awab
% 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('OFVF.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=1:120
    atr(i,:)=ndata(i ,:);
end
for i=121:170
    aval(i-120,:)=ndata(i,:);
end

for i=171:length(ndata)
    atest(i-170,:)=ndata(i,:);
end
for i=1:length(ndata)
    all(i,:)=ndata(i,:);
end
Ytr=atr(:,1);
Xtr=atr(:,2:5);
Xtst=atest(:,2:5);
Ytst=atest(:,1);
Yv=aval(:,1);
Xv=aval(:,2:5);

Yall=all(:,1);
Xall=all(:,2:5);
[model, time] = gmdhbuild(Xtr, Ytr, 3, 1, 4, 0, 2, 2, 1, Xv, Yv,1);

% [model, time] = gmdhbuild(Xtr, Ytr, 3, 1, 4, 0, 2, 2, 1, Xv, Yv,1);

gmdheq(model, 15);
[Yqtst] = gmdhpredict(model, Xtst);
[Yqval] = gmdhpredict(model, Xv);
[Yqtr] = gmdhpredict(model, Xtr);
[Yqall] = gmdhpredict(model, Xall);
```

```
[MSE, RMSE, RRMSE, R2] = gmdhtest(model, Xtst, Ytst);

Eall1=(Yall-Yqall)./Yall*100;
[m,n] = size(Eall1);

 AAPEall1 = sum(abs(Eall1))/m
 MinErrall1 = min(abs(Eall1))
 MaxErrall1 = max(abs(Eall1))
 Rall1=corrcoef(Yall,Yqall)
 Rall11=min(Rall1(:,1))


% Evaluating Relative Error for all set:
%============================================
Eall1=(Yall-Yqall)./Yall*100;
[m,n] = size(Eall1);;
figure
plot(Yall,Yqall,'o')
grid off
set(gcf, 'color', 'white')
axis square

title('Predicted FVF vs Measured FVF');
xlabel('Measured FVF "RB/SCF"');
ylabel('Predicted FVF "RB/SCF"')
legend('all data set', 'location', 'Northwest')
% Addding Reference Line with 45 degree slope
line([1 ; 1.4],[1 ; 1.4])
%HINT: Select the y-value based on your data limits
hold
% Evaluating the correlation coefficient for All set:
% =======================================================
Rall1=corrcoef(Yqall,Yall);
Rall11=min(Rall1(:,1));
gtext(['correlation coefficient = (' num2str(Rall11) ')']);
hold

% Addding Reference Line with 45 degree slope
line([1 ; 1.4],[1 ; 1.4])
%HINT: Select the y-value based on your data limits

% 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 FVF vs Measured FVF');
xlabel('Measured FVF "RB/SCF"');
ylabel('Predicted FVF "RB/SCF"')
legend('Training set', 'location', 'Northwest')
```

```
% Addding Reference Line with 45 degree slope
line([1 ; 1.4],[1 ; 1.4])
%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 ; 1.4],[1 ; 1.4])
%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 FVF vs Measured FVF');
xlabel('Measured FVF "RB/SCF"');
ylabel('Predicted FVF "RB/SCF"')
legend('Validation set', 'location', 'Northwest')
% Addding Reference Line with 45 degree slope
line([1 ; 1.4],[1 ; 1.4])
%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 FVF vs Measured FVF');
xlabel('Measured FVF "RB/SCF"');
ylabel('Predicted FVF "RB/SCF"')
```

```
legend('Testing set', 'location', 'Northwest')
% Addding Reference Line with 45 degree slope
line([1 ; 1.4],[1 ; 1.4])
%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('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(Errortt1);

% Evaluating Average Absolute Percent Relative Error
% =====================================================
AAPETT1 = sum(abs(Ett1))/m;
 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 API while fixing the other parameters
% % ------------API-------------
ps1=[linspace(10,20,10); %API  [min=10    max=22.2   mean=16.95569756]
linspace(0.67,0.67,10);%SPECIFIC GRAVITY OF GAS [min=0.0906   max=1.517
mean=0.902843]
linspace(158,158,10);%RESERVOIR TEMPERATURE[min=85.01    max=253.8140
mean=160.08456]
linspace(79.086,79.086,10)]';%SOLUTION GAS-OIL RATIO[min=0    max=429
mean=68.34259219]
%
% Now simulate
[Yq_API] = gmdhpredict(model, ps1);
% Plot Figures for API
figure
px1=plot(ps1(:,1),Yq_API(:,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('API','FontSize',12)
ylabel('Bo (RB/SCF)', 'fontsize',12)

%  %Simulation: Variation SPECIFIC GRAVITY OF GAS while fixing the
other parameters
% ------------SPECIFIC GRAVITY OF GAS ------------------

ps2=[linspace(13.2,13.2,10); %API  [min=10    max=22.2
mean=16.95569756]
linspace(0.0906 , 1.517 ,10);%SPECIFIC GRAVITY OF GAS [min=0.0906
max=1.517   mean=0.902843]
linspace(158,158,10);%RESERVOIR TEMPERATURE[min=85.01    max=253.8140
mean=160.08456]
linspace(79.086,79.086,10)]';%SOLUTION GAS-OIL RATIO[min=0    max=429
mean=68.34259219]

% Now simulate
[Yq_GG] = gmdhpredict(model, ps2);
% Plot Figures for SPECIFIC GRAVITY OF GAS Variation
figure
px2=plot(ps2(:,2),Yq_GG(:,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('SPECIFIC GRAVITY OF GAS','FontSize',12)
ylabel('Bo (RB/SCF)', 'fontsize',12)
```

```matlab
% Simulation: Variation of TEMPERATURE while fixing the other
parameters
% % ------------ TEMPERATURE--------------------

 ps3=[linspace(13.2,13.2,10);%API  [min=10    max=22.2
mean=16.95569756]
linspace(0.67,0.67,10);%SPECIFIC GRAVITY OF GAS [min=0.0906   max=1.517
mean=0.902843]
linspace(85.01,260 ,10);%RESERVOIR TEMPERATURE[min=85.01
max=253.8140  mean=160.08456]
linspace(79.086,79.086,10)]';%SOLUTION GAS-OIL RATIO[min=0    max=429
mean=68.34259219]


% Now simulate
[Yq_T] = gmdhpredict(model, ps3);
% Plot Figures for TEMPERATURE
figure
px3=plot(ps3(:,3),Yq_T(:,1),'-rs');
set(gca,'YGrid','off','XGrid','off')
set(gca,'FontSize',12,'LineWidth',2);
set(px3,'LineStyle','-.','LineWidth',1.5,'Color','k','MarkerSize',6)
xlabel('TEMPERATURE(F)','FontSize',12)
ylabel('Bo (RB/SCF)', 'fontsize',12)

% Simulation: Variation of SOLUTION GAS-OIL RATIO while fixing the
other parameters
% % ------------ SOLUTION GAS-OIL RATIO--------------------------

 ps4=[linspace(13.2,13.2,10);% API  [min=10    max=22.2
mean=16.95569756]
linspace(0.67,0.67,10);%SPECIFIC GRAVITY OF GAS [min=0.0906   max=1.517
mean=0.902843]
linspace(158,158,10);%RESERVOIR TEMPERATURE[min=85.01    max=253.8140
mean=160.08456]
linspace(2.9 , 429 ,10)]';%SOLUTION GAS-OIL RATIO[min=2.9    max=429
mean=68.34259219]


% Now simulate
[Yq_Rs] = gmdhpredict(model, ps4);
% Plot Figures for Yq_SOLUTION_GAS_OIL_RATIO
figure
px4=plot(ps4(:,4),Yq_Rs(:,1),'-rs');
set(gca,'YGrid','off','XGrid','off')
set(gca,'FontSize',12,'LineWidth',2);
set(px4,'LineStyle','-.','LineWidth',1.5,'Color','k','MarkerSize',6)
xlabel('SOLUTION GAS-OIL RATIO (rcf/STB)','FontSize',12)
ylabel('Bo (RB/SCF)', 'fontsize',12)
```