FINAL YEAR PROJECT REPORT

Prediction of Bubble-point Pressure (P_b) and Formation Volume Factor (B_o) Using Group Method of Data Handling (GMDH) approach and the effect of reducing correlating parameters; a comparative study

BY:

MOHD HAZIQ FARHAN BIN MOHD ZAKI

12682

PETROLEUM ENGINEERING DEPARTMENT

1ST AUGUST 2013

SUPERVISE BY:

DR. ABDALLA AYOUB

UNIVERSITI TEKNOLOGI PETRONAS
BANDAR SERI ISKANDAR
31750 TRONOH
PERAK DARUL RIDZUAN

CERTIFICATION OF APPROVAL

CERTIFICATION OF APPROVAL

Prediction of Bubble-point Pressure (P_b) and Formation Volume Factor (B_o) Using Group Method of Data Handling (GMDH) approach and the effect of reducing correlating parameters; a comparative study

By:

Mohd Haziq Farhan Bin Mohd Zaki (12682)

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. ABDALLA AYOUB)

UNIVERSITI TEKNOLOGI PETRONAS

TRONOH, PERAK

May 2013

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.

____

(MOHDHAZIQ FARHAN BIN MOHD ZAKI)

ABSTRACT

Final year Petroleum Engineering undergraduate students at Universiti Teknologi PETRONAS (UTP) are required to complete a Final Year Project (FYP), as part of the graduating requirement. In this course, students are given opportunity to use their knowledge as well as problem-solving tools and methods that they have acquired throughout their study to independently carry out research or design work. The project supervision and evaluations are mainly done by lecturers although practicing engineers from the industries are invited to participate.

The author's FYP title is "The Prediction of Bubble Point Pressure and Oil Formation Volume Factor using Group Method of Data Handling (GMDH) and the effect of reducing correlating parameters; a comparative study". A precise description of the reservoir fluid properties, hold importance essence in finding solution and solving petroleum reservoir engineering related problems. It is important for engineers to get ahold of the physical reservoir fluid properties, because they help in designing the best approach and strategies for the development of any oilfields.

A broad literature review was done to assist the author to apprehend and establish the parameter, boundary, limitations of the existing correlation as well as the method he's going to apply for his study.

This paper will be evaluating published correlations aimed to predict bubble point pressure and oil formation volume factor, the author's propose correlation which is his ambition to improve the current correlation's accuracy used by the industry in predicting the reservoir's bubble point pressure and oil formation volume factor through the method known as GMDH. Statistical analysis will be conducted to evaluate the model's performance.

This paper aims to generate two correlation models; bubble point pressure P_b, and Oil formation volume factor Bo, with high accuracy and less number of parameters. On the other hand, this paper will also discuss the effect of reducing parameters used in previously published correlation that took into account while developing them.

CONTENTS

CHAPTER 1

INTRODUCTION

Petroleum or also known as the crude oil is a complex naturally occurring liquid consisting of hydrocarbons of various molecular weights and other liquid organic compounds that are found in geologic formations beneath the Earth's surface. The physical reservoir properties of crude oil vary considerably in different reservoirs depending on concentration of the various types of hydrocarbons.

Reservoir fluids' properties are principally based on pressure-volume-temperature (PVT) analysis. This PVT analysis is usually determined from a detailed laboratory procedure intended to provide the key values of the reservoir fluids' properties. Reservoir fluid properties essentially used to:

- Classify reservoirs
- Classify the naturally occurring hydrocarbon systems
- Describe the phase behavior of the reservoir fluid

In general, reservoirs are commonly categorized on the foundation of the point representing the initial reservoir pressure and temperature with respect to the PVT diagram of the reservoir fluid. There are several key terms in PVT diagram; they are the cricondentherm, cricondenbar, critical point, two-phase region, the bubble point curve, the dew point line curve, the bubble point pressure. However in this paper, the author wishes not to stress on all of the term but rather to bubble point pressure and the oil formation volume factor.

Bubble point pressure (P_b) defines the highest pressure at which the first bubble of gas is liberated from crude oil thus shifting the crude from single phase into 2 phases. Identifying the bubble point pressure is critical because to gain maximum production rate, it is vital for reservoir pressure to be above than the bubble point pressure so that our reservoir fluid could be maintained in 100% liquid phase. Figure 1 shows the Pressure-Temperature diagram presenting the oil's phase at downhole.

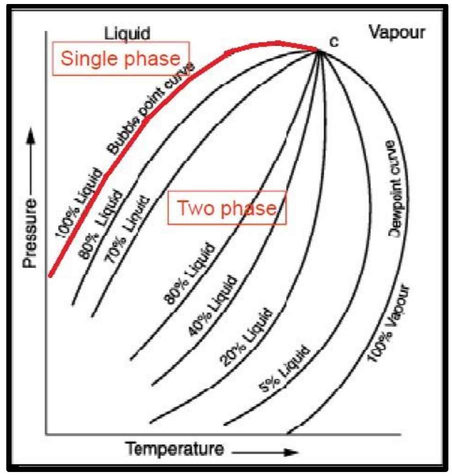

Figure 1 Oil Phase Diagram

Meanwhile, Oil Formation Volume Factor (B_o) is the ratio of the volume of oil at reservoir (in-situ) conditions to that at stock tank (surface) conditions. It is a measurement of reduction in the volume of crude oil as it is produced. Through many conducted research before, ideally, oil formation volume factor of any reservoir is the highest at the bubble point pressure. The impact of knowing oil formation volume factor is that it allows the industry to estimate the initial volume of oil in a reservoir, so that they could estimate the production they'll gain from producing crude oil of the respected reserves. Thus, allow them to figure the best approach and strategies to maximize the production of hydrocarbon economically.

The paper's objective is to create improved correlation of P_b and B_o through a method called as the Group method of data handling (GMDH). GMDH is a set of several algorithms for different problems solution. It consists of parametric, cauterization, analogues complexion, and probability algorithms. This inductive approach is based on sorting-out of gradually complicated models and selection of the best solution by minimum of external criterion characteristic. According to E. A. Osman in his paper [1] GMDH approach seems to be virtually unknown to the oil and gas industry. Along time, the author will be using this approach with this paper as a reference.

Most existed correlations that are used to predict P_b and B_o, include these parameters; bubble point solution GOR (R_{sb}), reservoir temperature (T_r), gas specific gravity (Υ_g), and stock tank oil specific gravity (Υ_{ost}).

P_b and B_o are two important reservoir fluid properties and were always used in many of existing oil property correlation. Even though, this paper stress on P_b and B_o correlation, due to the author have limit his scope of study to only these two properties, other reservoir fluid properties are also important because these properties help engineers to estimate the oil volume original in place, design the surface facilities during production stage, etc. Therefore, it is very important that these fluid properties are estimated as accurately as possible.

1.1 PROBLEM STATEMENTS

Ideally, the fluid properties for hydrocarbon reservoirs are obtained through laboratory analysis on a preserved or recombined sample of the reservoir fluid. The analysis involves series of laboratory procedures designed to provide the values of the reservoir fluid properties required in material balanced calculations, well test analysis, etc. However, there are many times in the industry;

1. the laboratory analysis privileges are absent,
2. economic issues to acquire a downhole fluid sample,
3. insufficient volume of sample to conduct a full analysis,
4. Poor sample quality for lab to conduct test due to human error while sampling processes or field transfers, and the possibility of lab errors.

Thus, correlations are widely used to estimate reservoir fluid properties. However, correlations are known to be more accurate when used to estimate reservoir fluids properties of regional crude oil. This is because due to difference oil physical and chemical properties of the reservoir itself for each reservoir geographically. [2] [3]

1.2 OBJECTIVES

The objectives of the conducted study are to;

1. Evaluate the existing correlation on bubble point pressure (P_b) and the oil formation volume factor (B_o)

2. Improving the existing correlation by gathering published data using group method of data handling (GMDH) approach.

3. Evaluate performance through statistical analysis

4. Checking the effects of reducing number of variables on the correlation accuracy.

1.3 SCOPE OF STUDIES

This paper will cover;

1. Data used to correlate will be gathered from published journal

2. Parameters related to correlate Bubble point pressure and oil formation volume factor

3. Effect of reducing parameters while developing the correlation to be used to predict P_b and B_o

CHAPTER 2

LITERATURE REVIEW

2.1 BUBBLE POINT PRESSURE, P_b

The bubble point pressure, Pb of a hydrocarbon system is defined as the pressure at which a bubble of gas is first liberated from the crude oil. This property can be measured experimentally for a crude oil system by conducting a constant composition expansion test [4].

With continued production, the reservoir pressure would decline further, producing appreciable quantities of gas that may dominate the multiphase flow of liquids in the reservoir. Once enough gas is produced, a high gas-oil ratio, HGOR at producing wells is expected.

Major decisions in reservoir engineering require knowledge of the bubble point pressure. Early pressure maintenance operation may be necessary to maintain reservoir pressure above the bubble point and avoid gas evolution and its eventual dominance in oil production. If the initial reservoir pressure is below the bubble point pressure with a gas cap present, reinjection of the produced gas may be necessary to maintain reservoir pressure at an optimum level.

In the absence of the experimental measured bubble point pressure, it is necessary for the engineer to make an estimation of the crude oil property from the readily available measured producing parameters. A large number of correlations for estimation of bubble point pressures of the reservoir oils have been offered in the oil and gas literature over the last few decades since the 1940's when petroleum engineers started to realize the importance of predicting an estimation of crude oil property.

These correlations mostly are based on the assumptions of bubble point pressure is a strong function of solution gas-oil ratio, Rs, gas gravity, ϒg, oil gravity, ϒo, and reservoir temperature, Tr

Pb = f { Rs, ϒg, , ϒo, Tr }

2.2 OIL FORMATION VOLUME FACTOR, BO

The oil formation volume factor, Bo is a measure of the shrinkage or reduction in the volume of crude oil as it is produced. Accurate evaluation of the oil formation is importance because it is related directly to the calculation of the oil initial in place under stock tank conditions. It is the ratio of reservoir oil barrels under the reservoir pressure and temperature over stock tank barrels of oil at the surface [5].

The oil formation volume factor can be expressed mathematically as;

Bo = (Vo)p,t / (Vo)sc

A typical oil formation factor curve, as a function of pressure for an under saturated crude oil (Pb>Pb) is shown below.

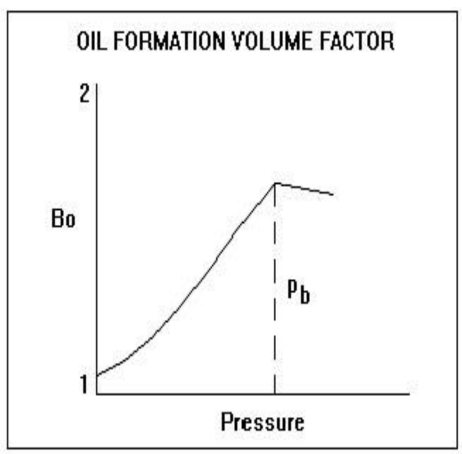

Figure 2 Oil Formation Volume Factor Curve

The pressure is reduced below the initial reservoir pressure, Pi, the oil volume increases due to the oil expansion. This behavior results in an increased in the oil formation volume factor and will continue until the bubble point pressure is reached. At Pb, the oil reaches it maximum expansion and consequently attains a maximum value of Bob for the oil formation volume factor.

As the pressure is reduced below Pb, volume of the oil and Bo, are decreased as the solution gas is released from the crude oil. When the pressure is reduced to atmospheric to surface pressure and temperature, the value of Bo is equal to 1.

Most of the published empirical correlations for Bo assume Bo to be a strong function of;

$$Bo = f\{Rs, \Upsilon g, \Upsilon o, Rt\}$$

2.3 EXISTING CORRELATIONS FOR P_b AND B_O

Correlation for reservoir fluid properties has been an interest of petroleum engineers as early as 73 years back; the 1940's. During that time, studies have been carried out to assist field engineers on site. Correlation by Kartz [6], Standing [7], and Vasquez and Beggs [8] are among the commonly used correlation used in the industry. As time passed, more improved correlations have been published such as Al- Shammasi [9], El-Mabrouk [10] and the recent one improved correlation by Parag Bandyopadhyay [11].

From the literature review, it has been observed that correlation model is suggested for a specific geographical region. [9] It is also has been an assumption that Pb and Bo are functions of four field parameters; (1) solution GOR at bubble point, Rsb (2) reservoir temperature, Tr (3) gas specific gravity; ϒg and (4) stock tank oil specific gravity ϒost [10] . On top of that, this study also requires a large size of data set so that the correlation models from this paper would be reliable.

According to Al-Shammasi (1999) the relationships between variable in various forms were explored through plots. From these plots then, he used linear regression method to test his correlation performance. His study then leads him into proposing his Pb and Bo model;

$$P_b = \Upsilon_o^{5.527215} * e^{-1.841408*|\Upsilon o*\Upsilon r|} * [R_s * (460+T) * \Upsilon_g]^{0.783716}$$

Equation 1 Al-Shammasi P_b Model

Al-Shammasi [9] also was able to produce 2 correlations for Bo in his paper with one of them used only three parameters instead of common four that is usually used. He excluded the gas gravity parameter. His new correlation's statistical performance for Pb model, gave a 0.9987 correlation coefficient, 17.85% average absolute error, while for Bo with reduced parameter, his model's test data average absolute error is 19.86%. Below is the new correlation with four parameters followed by correlation with three parameters.

$$B_o = 1+5.53*10^{-7} (R_s*(T-60)) +0.000181*(R_s/\Upsilon_o) +0.000449*(T-60)/\Upsilon_o) +0.000206*(R_s*\Upsilon_g/\Upsilon_o)$$

Equation 2 Al-Shammasi Bo four variables

$$B_o= 1+0.000412*(R_s/\Upsilon_o) +0.000650*((T-60)/\Upsilon_o)$$

Equation 3 Al-Shammasi Bo three variables

So far, the author has only found Al-Shammasi paper that studies developing regression analyses model with reduce parameters in correlation. [9] This paper could be made as reference for future study under this paper scope to study the impact towards the accuracy of correlations with less parameter.

In 2010, Elmabrouk [10] published his paper with his objective, to overcome the limitations faced by the previous correlations by developing multiple regression models using directly measured field parameters as input to predict Pb and Bo. He believed that the previous correlation input variables required the knowledge of parameter like solution GOR and specific gravity gas; it is difficult to apply them in the absence of

PVT analysis. Instead of the conventional he develop his correlation under four readily available field parameters (1) solution GOR at separator, Rsp (2) Pressure at separator, Psp (3) Stock tank oil specific gravity, ϒost and (4) reservoir temperature, Tr. The correlation he made, after calibrated, has an average percent relative error for PB model, 2.83% and for his Bo model, average percent relative error, 0.038%. The models he proposed;

$$P_b = R_{sp}^{\ 0.683}\, P_{sp}^{\ 0.18}\, \Upsilon_{ost}^{\ 4.98}\, T_r^{\ 0.658}$$

Equation 4 El- Mabrouk Pb model

From his paper, he found that oil formation volume factor at bubble point pressure, Bob is equal to flash separation oil formation volume factor, Bofb. His proposed model to estimate Oil formation volume factor at bubble point, Bob;

$$B_{ob} = 1.6624+0.000512R_{sp}+0.00015P_{sp}-0.802\Upsilon_{ost}+0.000501T_r$$

Equation 5 El- Mabrouk Bob model

In 2011, Parag Bandyopadhyay proposed his own model to predict only Pb resulting from a complex regression analysis. [11]. He compared his model with two equations of state (EOS) model; Soave- Redlich- Kwong (SRK) [12] and Peng-Robinson (PR) [13]. In his study, he relates Pb with molar concentrations, molecular weight of heavy fraction and temperature of reservoir fluid. However, according to his findings, he stated that his proposed model's input parameter are molar percentage of crude oil components and temperature at which bubble point is to be estimated. He named his model as Exponential Interaction (EI)

$$Pb = \sum_{i=0}^{7} (Ai + Bi\, T^n)\, Ci$$

Equation 6 Porag's Pb Exponential interaction

The equation he proposed is calibrated to see its accuracy agaist SRK and PR EOS model. His model then is proved to be accurate than these two models by 3%. In his paper, he stated that bubble point pressure is a (1) combination of linear and non-linear functions of percentage molar concentrations of the individual components in crude oil (2) bubble point pressure is a non-linear power function of temperature and (3) sensitivity of bubble point pressure varies with temperature as well as the composition of the crude oil (4) the model can be used to improve Bo and Rs predictions to populate Black Oil PVT tables for application in reservoir engineering.

M. I. Omar published his paper in 1993 on developing a modified a black oil model for Malaysian crude oil using Standing correlation [14]. His paper mentioned that the obstacle of using Standing's correlation in Malaysia is that, the correlation is difficult to determine its accuracy since the correlating factor of Standing is develop using different sample (California's crude) than the sample obtain in Malaysian's crude in terms of physical and chemical properties. To overcome this, he introduce 'bias factor' as other published paper has done previously when made comparison with Standing's correlation. His model uses the same variables as Standing (Rs, ϒg, Tf, ϒapi, and Bob) but with different coefficient obtained through non-linear regression analysis on each of the Standing's correlation coefficient. His model shows a better accuracy in predicting Pb (Absolute Average Relative Percentage Error; AARPE of 7.17%) and Bo (AARPE of 1.44%) of Malaysian crude when compared with Standing's and other known correlations such as Vasquez [8] and Marhoun's [15] PVT correlation.

For GMDH approach by E.A Osman [1] conducted in 2002, he developed two models for predicting Pb and Bo. He used Neural Network for his comparison with his model. His approach using GMDH as the tool to develop new correlation when statistically assessed shows that his model to be more accurate than Neural Network for Pb shows absolute percentage error of 5.62% and for Bo model, the absolute percentage error was 0.86%. Nevertheless, his paper objectives are to show the usefulness and the power of GMDH approach in developing a correlation for reservoir engineering. His model work its way by formalized the paradigm for iterated polynomial regression of wide range of data, capable of producing high degree polynomial model in effective predictors. The algorithm will select the polynomial relationships and input combinations that minimize the prediction error in every phase. Iteration is stopped automatically at a point of balance between models of data and fit them to the training data and creates a model that could be generalized with new data.

From the literature review, it is observed that most correlations existed previously, would be accurate if the correlations are made for regional crude using the region's crude oil as the sample. The normal parameters used for Pb and Bo correlations are;

1. Solution GOR at bubble, R_{sb}
2. Reservoir Temperature, T_r
3. Specific Gravity of Gas, Υ_g
4. Specific Gravity of Stock Tank Oil, Υ_{ost}

Using existing correlation; new correlation made to predict Pb and Bo could be developed through regression analysis and the author can safely assume the new correlation from this paper should have an Absolute Average Relative Percentage Error of less than 20% to be acceptable.

2.4 GROUP METHOD OF DATA HANDLING (GMDH)

The Group Method of Data Handling (GMDH) was developed for complex systems analysis, forecasting, knowledge discovery and data mining, diagnostics and decision support. It is based on sorting-out of gradually complicated models and evaluation of them by external criterion on separate part of data sample.

In GMDH as input variables can be used any parameters, which can influence on the process. Computer is found structure of model and laws which acts in the system by external criterion value itself.

This inductive method is different from deductive techniques used commonly for modeling on principle. Only by this self-organizing method for inaccurate, noisy or small data can be found optimal non-physical model, accuracy of which is higher and structure is simpler than structure of usual full physical model

The method was originated in 1968 by Prof. Alexey G. Ivakhnenko in the Institute of Cybernetics in Kiev, Ukraine. This approach from the very beginning was a computer-based method so, a set of computer programs and algorithms were the primary practical results achieved at the base of the new theoretical principles.

Later on different GMDH variants were published by Japanese and Polish scientists. The scientists concluded that the GMDH approach is the best method for solving AI problems- identification, short term and long term forecasting of random processes and pattern recognition in a complex system.

Major involvements of GMDH were recorded in the Ukrainian journal "Automatica" [16] and as follow; -

1968-1971

This period is distinguished by the application of one regularity criterion to solve problems of identification, pattern recognition and short-term forecasting. As reference functions polynomials, logical nets, fuzzy Zadeh sets and Bayes probability formulas were used. Authors were stimulated by very high accuracy of forecasting with the new approach. Noise-immunity was not investigated.

1972-1975

The problem of modeling of noised data and with incomplete information basis was solved. Multi-criteria selection and utilization of additional priory information for noise-immunity increasing were proposed. Best experiments showed that with extended definition of the optimal model by additional criterion noise level can be ten times more than signal. Then it was improved using Shennon's theorem of General Communication theory.

1976-1979

The convergence of multilayered GMDH algorithms was investigated. It was shown that some multilayered algorithms have "multi-layered error" - analogical to static error of control systems. In 1977 solution of objective systems analysis problems by multilayered GMDH algorithms was proposed. It turned out that sorting-out by criteria ensemble allows to choose the only optimal system of equations and therefore to show complex object elements, their main input and output variables.

1980-1988

Many important theoretical results were received. It became clear that full physical models cannot be used for long-term forecasting. It was proved, that non-physical models of GMDH are more accurate for approximation and forecast than physical models of regression analysis. Two-level algorithms which use two different time scales for modeling were developed.

1989- Current

New algorithms for non-parametric modeling of indistinct objects and simplified learning programming for expert systems were developed and investigated. Present stage of work is devoted to development and implementation, mainly into economical systems, of twice-multilayered neuronets, which open a new solution to the problem of self-organization of artificial neuronets - models of human brains.

GMDH approach [17] can be useful because:

- The optimal complexity of model structure is found, adequate to level of noise in data sample. For real problems solution with noised or short data, simplified forecasting models are more accurate.
- The number of layers and neurons in hidden layers, model structure and other optimal NN parameters are determined automatically.
- It guarantees that the most accurate or unbiased models will be found - method doesn't miss the best solution during sorting of all variants (in given class of functions).
- As input variables are used any non-linear functions or features, which can influence the output variable.
- It automatically finds interpretable relationships in data and selects effective input variables.
- GMDH sorting algorithms are rather simple for programming.
- TMNN neural nets are used to increase the accuracy of other modeling algorithms.
- Method uses information directly from data sample and minimizes influence of previous author assumptions about results of modeling.
- Approach gives possibility to find unbiased physical model of object.

2.4.1 APPLICATIONS OF GMDH IN OIL AND GAS INDUSTRY

GMDH approach have been used for modeling and sorting inputs in many fields including weather forecasting, medical diagnostics, marketing, and many more. However, the technique's applications seem to have not received the same spotlight as the technique had in other fields.

A search has shown that only a small fragment of research papers on GMDH modeling approach in the oil and gas industry have been published up until today. Among papers on GMDH has been published is by Lee, Y. B., Liu, H. S., and Tarng, Y. S have conducted a research to predict the drill life under varying cutting conditions. They used the GMDH approach to define the period of drilling time of an average flank wear land. The research paper prove that GMDH approach can be effectively used to predict drill life under varying conditions, with prediction error of less than 9 % [18].

Besides that, in 2002, E.A Osman also has published a paper together R.E Abdel –Aal that discussed the applications and the high potential of GMDH approach in intelligent modeling when applied in the oil and gas industry. The authors then prove their claimed by building two GMDH modeling approach to develop correlation for bubble point pressure and oil formation volume factor. Their models outperformed the rest of empirical correlations [1].

In 2007, M.A Abdalla published his research on evaluation the below bubble point viscosity correlations using GMDH modeling approach. His paper evaluates two below bubble point viscosity correlations, and he tried to develop a new correlation using GMDH to estimate this property. The result was the new model outperformed the previous two models. His new model was able to predict the estimation of below bubble point viscosity with outstanding correlation coefficient of 99.3% [19].

Furthermore, in 2010, A.A. Semenov, R.A. Oshmarin, A.V. Driller, and A.V. Butakova, Vankoroil co-wrote a paper on GMDH modeling technique to model the geography of Vankor field in Russia. The paper claimed by using multilayered algorithm of GMDH, based on polynomial reference function allowed maximizing amount of information being used from different types of well logs in reference wells for target relation. Data from neutron, density, resistivity and PS logs were the most significant for the final model of Dolgan reservoir. So using of GMDH cybernetics algorithms may significantly increase precision of rock properties forecast for further geomodeling purposes [20].

Two years later, Mohammad, N., Gholam- Abbas, B., and Haji, M. A conduct. A prediction of pipeline scour depth in clear-water and live-bed conditions. They suggest that the parameters which can affect the scour depth are sediment size, geometry of pipeline, and the approaching flow characteristics. They then compared their result with other methods of prediction including support vector machines (SVM) and commonly used empirical equations. The result is significance, where GMDH outperform the other methods. The prediction made using GMDH have lower error and higher accuracy [21].

CHAPTER 3

METHODOLOGY

The findings of this paper will mostly be based through a qualitative method. The author starts by reviewing published articles and journals of related topic and within the scope of this paper study. Using a wide set of data from published papers of related topic, the author would be able to analyze the information, patterns of existed correlation and will come out with a newly improved correlation for predicting oil bubble point pressure and oil formation volume factor.

The new correlation will be the results of in-depth data gathering using the available data point of published journal. Through the GMDH method, these data will be arranged according to its implicit pattern and will be calibrated with regression analysis method using MATLAB software.

Along the way, the author will study the effect of reducing the correlating parameters. Figure 3, below summarizes the methodology of this paper;

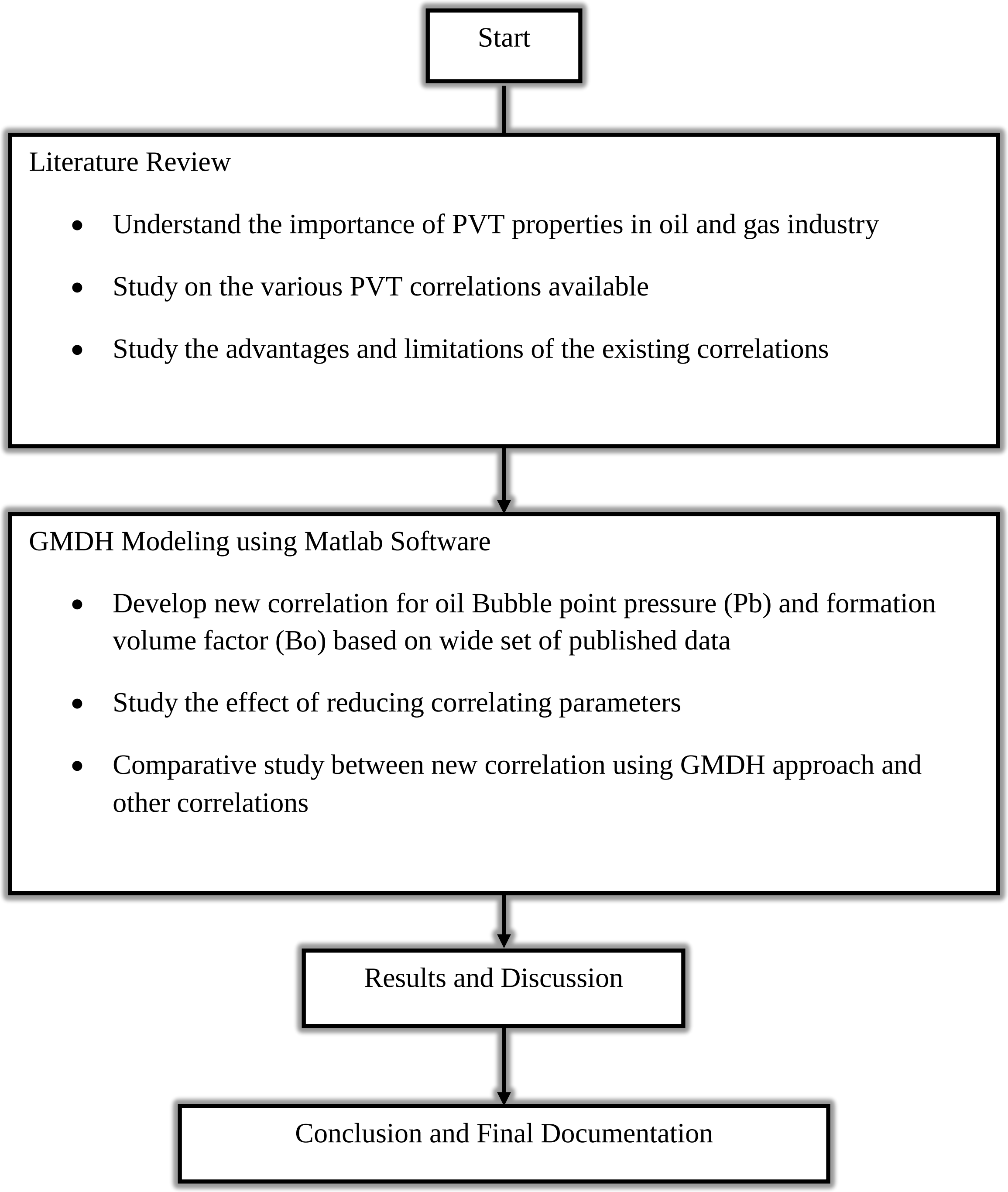

Figure 3 Summary of methodology

3.1 LITERATURE REVIEW

Literature review is done to obtain a wider view of this study's perspective. It helps the author to understand the importance of PVT properties in oil and gas industry, study on the various PVT correlations available and able to study the advantages and limitations of the existing correlations.

3.2 THE GMDH MODELING APPROACH

Step 1: Data Gathering

The first step in generating a successful model to correlate bubble oil pressure, PB and Oil formation volume factor, Bo is by gathering the suitable data. The sources of data will be collected either by extensive internet research from published paper from the industry, directly contacting the authors of previous published paper on related topic, or by applying for data from any oil and gas companies to be used for the study.

The attributes of the data as the inputs should be well known to be contributing to the desired outputs. On top of that, the data volume also should be big enough so that the data will be able to provide a firm model and able to improve of the previous correlations for P_b and B_o.

Step 2: Pre-Processing

The second step in generating the model is by carefully cleaning and consolidates the data gathered according to the objective of this study which is the correlation to estimate Pb and Bo. There will be two stages in this second step. They are database consolidate and data filtration.

In database consolidate; data that is gathered in the first step will be tabulated so that the data collected will be organized, and not scattered. By organizing data into tabulated form, the author could found any inconsistencies, irregularities, repetitive, or any missing entries of the data that is going to be use.

Data filtration defines the effort to remove the data outliers, any non-normal distributions and other glitches within the data. However, data filtration is not just about removing bad data or interpolating missing values, but about finding hidden correlations in the data, identifying sources of data that are the most accurate, and determining which columns are the most appropriate for use in analysis.

Step 3: Data Handling

The third step to generate the correlation is by appropriately divide the data gathered into three different sets; training set, validation set, and the test set. The impact of handling the data suitably, is it determines the success rates of the Pb and Bo correlations output.

A training set consists of the inputs from data and the actual output, and is used together with a supervised learning method (GMDH) to *train* a knowledge database. The training set should have sufficient numbers of data because the set is used to develop and adjust the weights in the model to be produced. The higher number of data allocate for training usually the better the model's performance would be.

Validation set is presented to the model during the training phase to ensure the optimum generalization of the developed model. In other words, data validation serves as the model check and balances before the model is tested with the test data sets. The validation set is used to compare their performances and decide which data to be applied, and finally, the test set is used to obtain the performance characteristics such as the model's accuracy, sensitivity, etc.

The test set is a set of data that is independent of the training data, but that follows the same probability distribution as the training data. These sets allow the author to examine the final performance of the model.

The ratio that has been decided for data that will be used in this study is 2:1:1. Half of the data set will be reserved for training. One quarter for validation set and the remaining quarter will be used as the test set.

Step 4: Developing Model

To develop the correlation model for Pb and Bo, the software that will be used in this study is the Matlab software. Matlab is chosen because the software has high range of flexibility related to programming and graphs visualization. The software also has a good way to observe the performance of the three sets of data at the same period. To achieve the prime objective of this study, a matlab code is developed to cater the function of the model's training, validation and test data sets.

Step 5: Model Performance

Graphical Error Analysis is used a graphical tool aids to visualize the performance and accuracy of correlation. The techniques used for this error analysis would be the cross plot technique.

Average Percent Relative Error (APE) – measures the relative deviation from the data.

$$Er = \frac{1}{n}\sum_{i=1}^{N} | Ei |$$

Equation 7 APE formula

Ei is the relative deviation of an estimated value from an experimental value.

$$Ei = \left| \frac{(u)\exp - (u)est}{(u)exp} \right| \; x\ 100, i = 1, 2, 3, ..n$$

Equation 8 Ei formula

Average Absolute Percent Relative Error (AAPE) – measures the relative absolute deviation from the experimental values.

$$Ea = \frac{1}{n}\sum_{i=1}^{n}|Ei|$$

Equation 9 AAPE formula

Minimum Absolute Percent Relative Error (Emin)

$$Emin = |Ei|$$

Equation 10 Min APE

Maximum Absolute Percent Relative Error (Emax)

$$Emax = \max|Ei|$$

Equation 11 Maximum APE

Root Mean Square Error (RMSE) – measure the data dispersion around zero deviation.

$$RMSE = \left[\frac{1}{n}\sum_{i=1}^{n}Ei^2\right]^2$$

Equation 12 RMSE formula

Standard Deviation (SD)

$$S = \sqrt{\left[\left(\frac{1}{m-n-1\sum_{i=1}^{m}\left\{\frac{(Uexp - Uest)}{uexp}\right\} 100}\right)\right]^2}$$

Equation 13 Standard deviation formula

(m-n-1) are the degrees of freedom in multiple-regression. A lower value indicates a smaller degree of scatter.

Correlation Coefficient (R^2) – represents the degree of success in reducing the standard deviation by regression analysis.

$$R = \sqrt{\frac{1 - \sum_{i=1}^{n}[Uexp - U\,est]_i^n}{\sum_{i=1}^{n} Uexp - \overline{\Delta U}}}$$

Where; $\overline{\Delta U = } \sum_{i=1}^{n}[(\Delta U)exp]i$

R values range between 0 and 1. The closer to 1 represents a flawless correlation.

3.3 PROJECT GANTT CHART AND MILESTONES

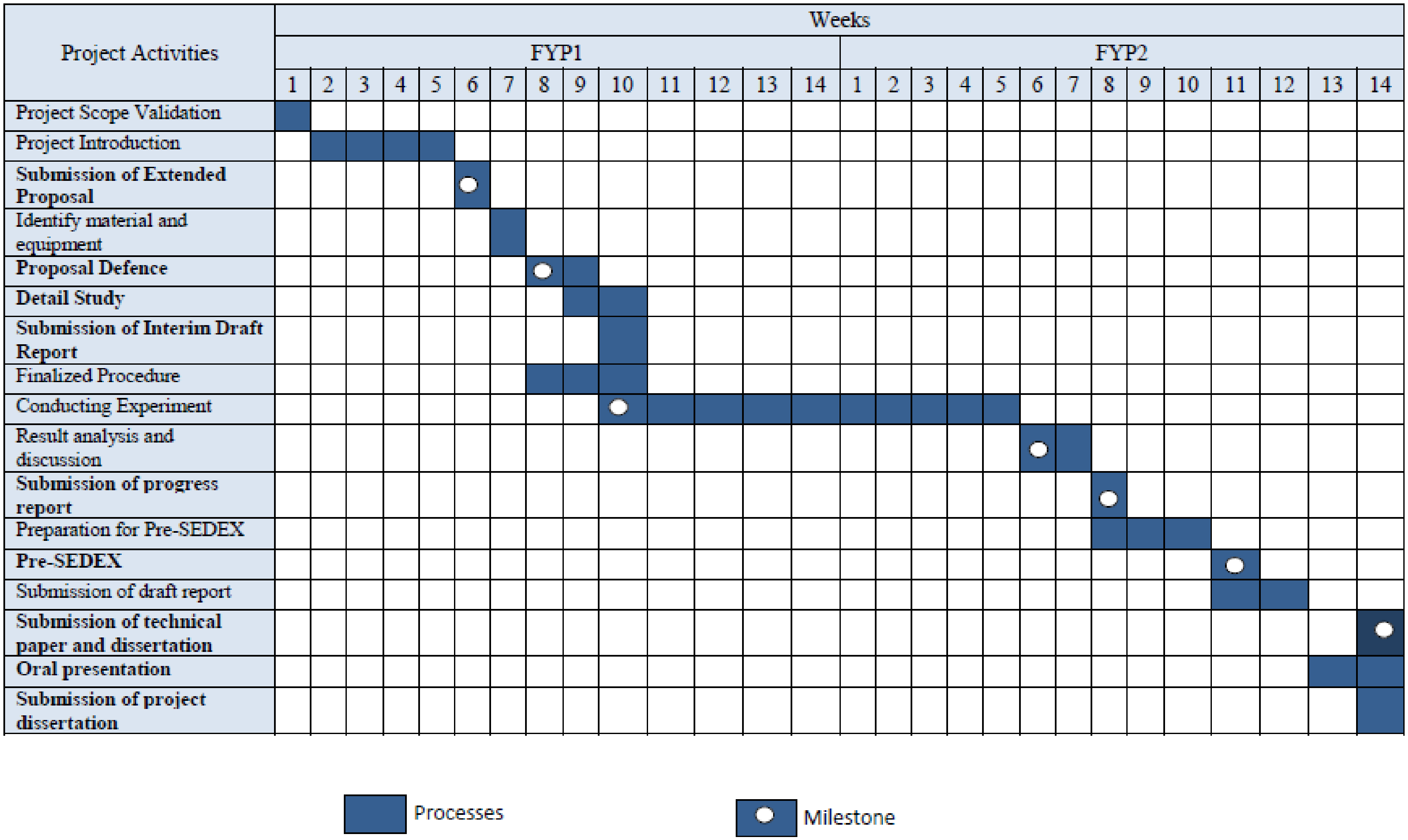

Figure 4 Project Gantt chart and Milestones

3.4 TOOLS AND EQUIPMENTS USED

Table 1 Tools and Equipments used

Tool	Function
Matlab	To develop GMDH modeling approach for new PVT correlation
Gints Jekabson	
Microsoft Office Word	To write reports, data etc
Microsoft Office Excel	To prepare data sheets and calculations
Microsoft Office Power Point	To prepare presentations

CHAPTER 4

RESULTS AND DISCUSSIONS

4.1 THE GATHERING AND SORTING OF PVT DATAPOINTS

A data set is given to represent raw data consisting of reservoir temperature, bubble point pressure, formation volume factor, gas gravity, solution gas-oil ratio, and oil density.

Data used for this study is available in the previous literatures and consists of series of PVT data points from three different regions.. The PVT data points that were used in this study are Oil API, Oil specific gravity, Gas specific gravity, Solution Gas-Oil Ratio (Rs), Reservoir Temperature, and Oil density for Bubble point pressure (Pb), Oil formation volume factor (Bo). The selections of PVT data points are based on the input necessities by majority of published correlations.

A total of 268 data points from 3 different published papers were collected. Each of the papers were from different regions; UAE, Middle East, Malaysia. Each of the data groups were screened for duplicates and crosschecked with other data groups to avoid repentance of data.

Table below shows the group of data and the range of every PVT data used in this study.

Table 2 Data group and ranges

	Malaysia		Middle East		UAE	
	Min	Max	Min	Max	Min	Max
Pb	790	3851	508	4640	590	4640
Bob	1.085	1.954	1.096	2.493	1.216	2.493
Rs	142	1440	127	2266	181	2266
Gas SG	0.612	1.315	0.752	1.29	0.798	1.29
API	26.6	53.2	21.9	44.6	28.2	40.3
Temp	125	280	74	275	190	275
Oil SG	0.766	0.922	0.804	9.22	0.8236	0.886
Data	92		125		51	
Total	268					

The datasets are sort together randomly in Excel Spreadsheet before divided into ratios of 2:1:1 for data training, data validation, and data test purposes. For this study, 136data are used for data training, 68 data are used for validation and the remaining 68 data points are used as the test data for this study's correlation.

4.2 GMDH CORRELATION MODEL: OIL BUBBLE POINT PRESSURE, P_b

The GMDH approach has develop number of equations for predicting the bubble point pressure. The best correlation for estimating oil bubble point pressure through GMDH approach has included only *four out of six* gathered reservoir and PVT properties as the correlation's input. The selected input was:

- Oil API
- Gas Specific Gravity, Y_g
- Solution Gas-Oil Ratio, R_s
- Reservoir Temperature, T_r

The model was the result of 2 hidden layers of the network of all possible inputs to achieve the desired output; the Bubble point pressure. Diagram below shows the layers of input with the heaviest weightage to produce the wanted output.

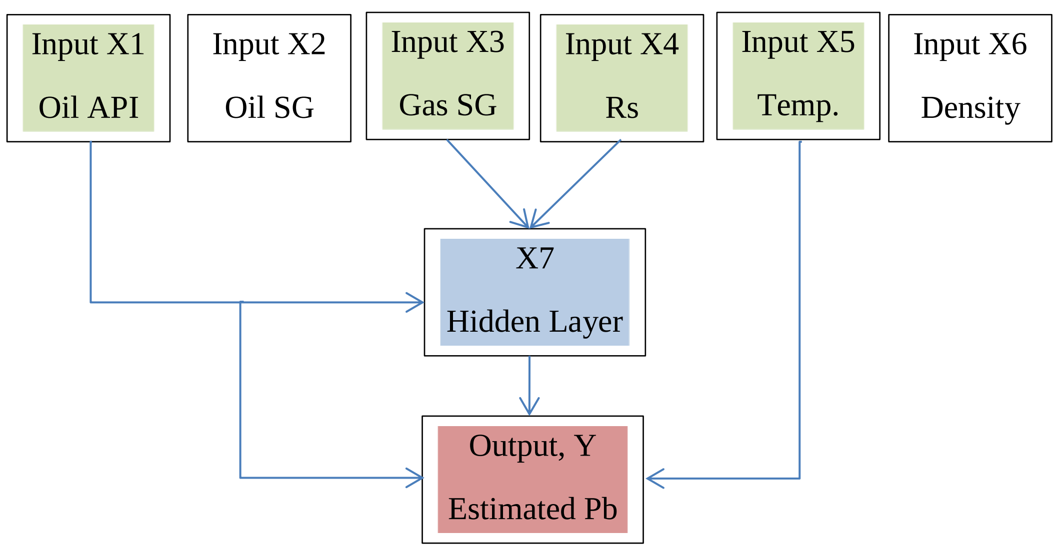

Figure 6 Layers of Inputs for Pb GMDH model 4 parameters input

Textbox below shows the model's equation to predict Pb as generated by MATLAB.

Model =

numLayers: 2 d: 6 maxNumInputs: 3 inputsMore: 1

maxNumNeurons: 6 p: 2 critNum: 2

Layer: [1x2 struct]

Time = 0.2406

Number of layers: 2

Layer #1

Number of neurons: 1

x7 = 9366.58879439684 +9.58414465571444*x4 -12914.5300942503*x3 -212.076724585407*x1 -3.92221694657826*x3*x4 -0.0792054058762421*x1*x4 +167.123826774092*x1*x3 -0.00062140630692062*x4*x4 +3915.39234380248*x3*x3 +1.17108748167652*x1*x1

Layer #2

Number of neurons: 1

y = -576.026418095619 +1.16180913428605*x7 +2.19383739988093*x5 +11.577828005077*x1 -0.000938537869568702*x5*x7 +0.00418047284863272*x1*x7 +0.377062362478756*x1*x5 -2.74297766705744e-005*x7*x7 -0.0330880310957772*x5*x5 -1.2764327206288*x1*x1

Figure 7 GMDH Pb correlation coefficient

The correlation has 19 coefficients for 2 layers of GMDH network with 4 key parameters as input to predict the bubble point pressure.

4.2.1 P_b STATISTICAL AND GRAPHICAL ERROR ANALYSIS

The correlation was then was evaluate with statistical error analysis to measure the reliability of the GMDH model predictions of P_b against the measured P_b value of the PVT data points set. The statistical error analysis parameters used are; average percent relative error, average absolute percent relative error, minimum and maximum absolute percent error, root means square and the correlation coefficient. Equations of those parameters are given in Chapter 3 Methodology section.

Besides that, graphical error analysis also is used in this assessment to aids visualizing the performance and accuracy of the GMDH model correlation. The table below summarizes the statistical error analysis done for this study's P_b correlation.

Table 3 Statistical Error Analysis for Pb GMDH correlation

Statistical Error Analysis	This Study P_b
Minimum Absolute Percent Relative Error, Emin	0%
Maximum Absolute Percent Relative Error, Emax	99%
Average Absolute Percent Relative Error, AAPRE	10%
Standard Deviation, SD	12%
Root Mean Square Error, RMSE	16%
Square of Correlation of Coefficient, R^2	0.9377

The correlation has quite large range of errors with min error is 0% while maximum error went as high as 99%. However, average absolute relative error is small with 10% only. The correlation's standard deviation is also small with 12% shows that the error distribution of this study correlation isn't spread far too wide from the true value of bubble point pressure. The correlation also recorded a small root means square error of 16% which means this study's correlation for P_b has good measure of accuracy.

The cross plot shown below compares on a linear line of the true value of bubble point pressure and the estimated bubble point pressure from this study correlation. The squared value of correlation of coefficient, R^2 is 0.9377 which is nearly to one which means the GMDH model for P_b correlation has a good prediction range.

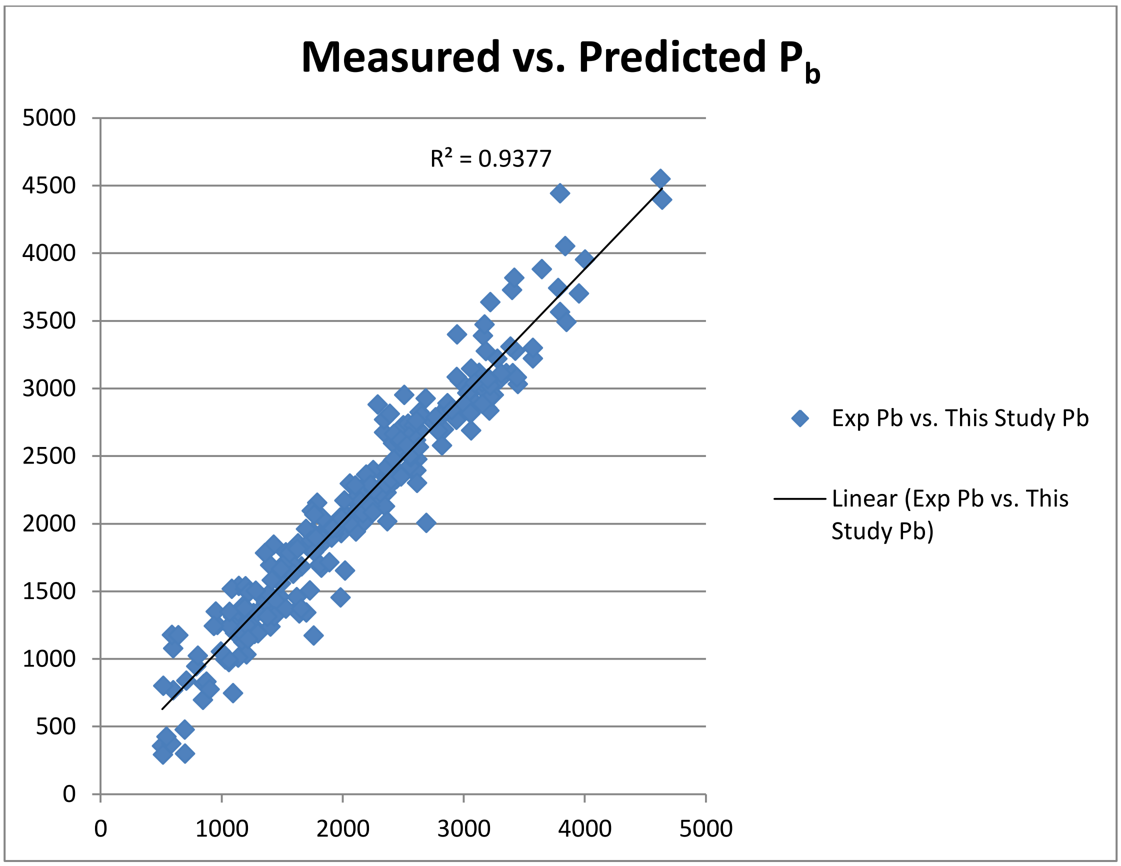

Figure 8 Cross plot of Measured vs. Predicted P_b for testing set

4.3 GMDH REDUCING PARAMETER EFFECTS on CORRELATION MODEL: OIL BUBBLE POINT PRESSURE, P_b

One of the objectives of this paper is to study the effect of reducing the GMDH model P_b correlation parameters. As discussed earlier, the correlation of this study has 4 PVT parameters as the key inputs to predict Bubble point pressure, P_b.

In this part, the key parameters; Solution Gas-Oil Ratio (R_s), Gas Specific Gravity (Y_g), crude oil API, and the reservoir's temperature will be erased one by one in order to know the impacts of each parameter to this study P_b correlation. The impacts of each parameter then are assessed using statistical error analysis with aids from graphical cross plot analysis.

The P_b correlation without the reservoir temperature as one of the parameter has only 11 coefficients for 2 layers of GMDH network with 4 key parameters as input to predict the bubble point pressure.

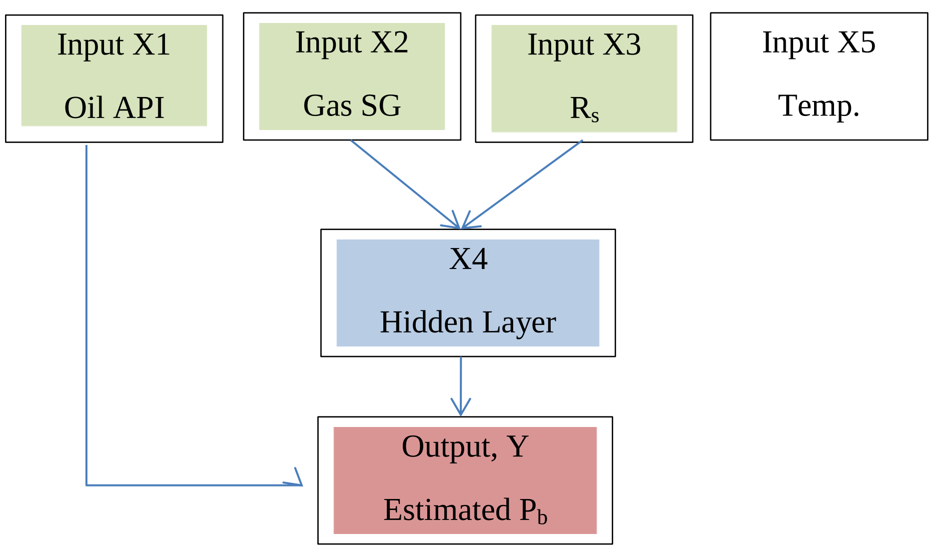

Figure 9 Layers of Inputs for Pb GMDH model 3 parameters input

Textbox below shows the model's equation to predict P_b without the reservoir temperature as generated by MATLAB.

Model =

numLayers: 2 d: 3 maxNumInputs: 2 inputsMore: 1

maxNumNeurons: 3 p: critNum: 2 layer: [1x2 struct]

Time =

0.1556

Number of layers: 2

Layer #1

Number of neurons: 1

x4 = 2646.8313853442 +5.95676886440463*x3 -4605.52338351011*x2 -3.39067984184884*x2*x3 -0.000674316659991972*x3*x3 +2363.33025938776*x2*x2

Layer #2

Number of neurons: 1

y = -1272.52172387022 +2.60279987306032*x4 +3.59913976057822*x1 -0.0497577378026607*x1*x4 +7.10283859621138e-005*x4*x4 +0.979862274593759*x1*x1

Figure 10 GMDH Pb correlation coefficient without Reservoir Temperature coefficient

This study's second correlation for P_b with 3 parameters is less complicated than the first P_b correlation with 4 parameters. On top of that the 3 parameters correlation reliability also was as good as the first correlation with 4 parameters as inputs. Figure 9 above shows the simplicity of the GMDH model P_b correlation with 3 key inputs.

4.3.1 REDUCING EFFECT FOR P_b STATISTICAL AND GRAPHICAL ERROR ANALYSIS

Table below shows the statistical error analysis of reducing P_b parameters.

Table 4 Statistical Error Analysis for Reducing Correlating Parameter

Statistical Error Analysis	**without R_s**	**without Y_g**	**without Oil API**	**without Temp, T_r**
Emin	0%	0%	0%	0%
Emax	156%	142%	54%	97%
AARE	28%	17%	14%	12%
Standard Deviation	27%	17%	11%	11%
RMSE	39%	24%	17%	16%
R^2	0.498	0.8067	0.8482	0.9084

The graphical cross plots below shows how by the effect of reducing each parameters as the key input to the P_b correlation.

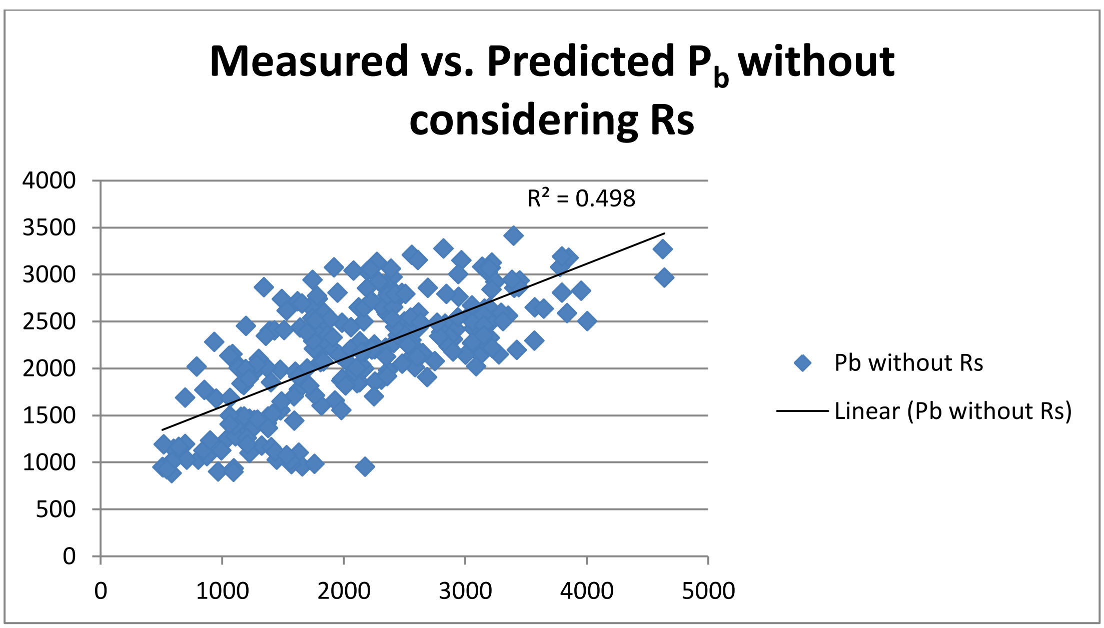

Figure 11 Measured vs. Predicted P_b without R_s considering

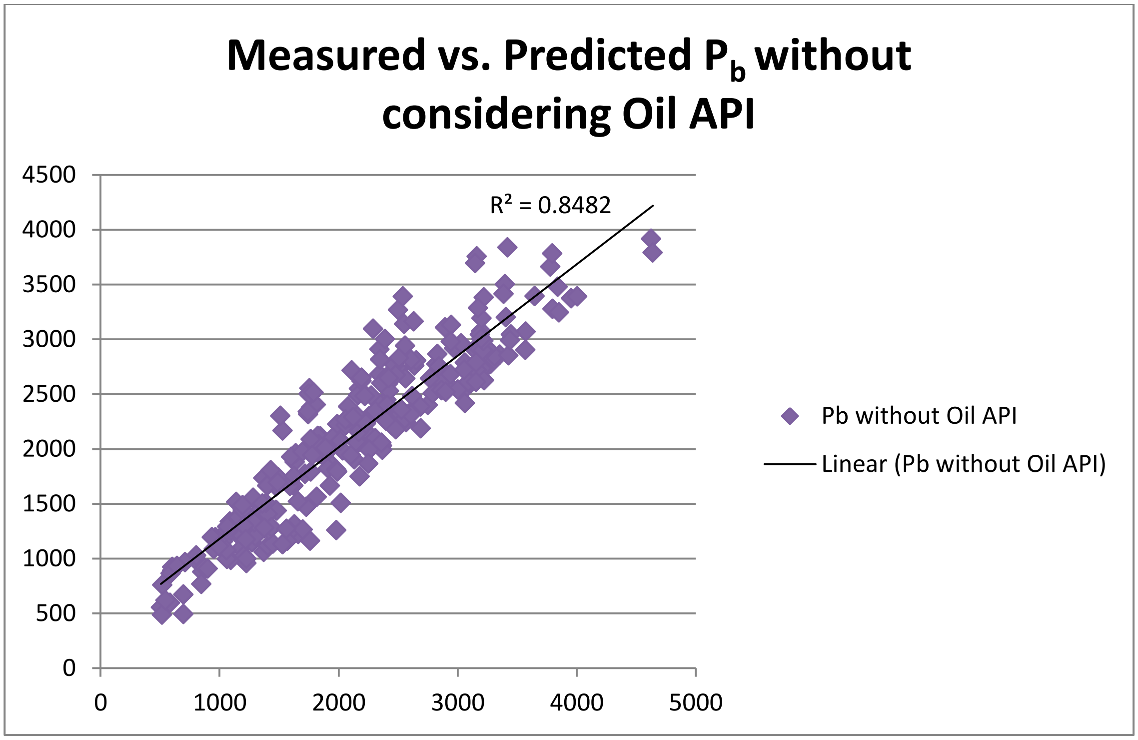

Figure 12 Measured vs. Predicted P_b without considering Oil API

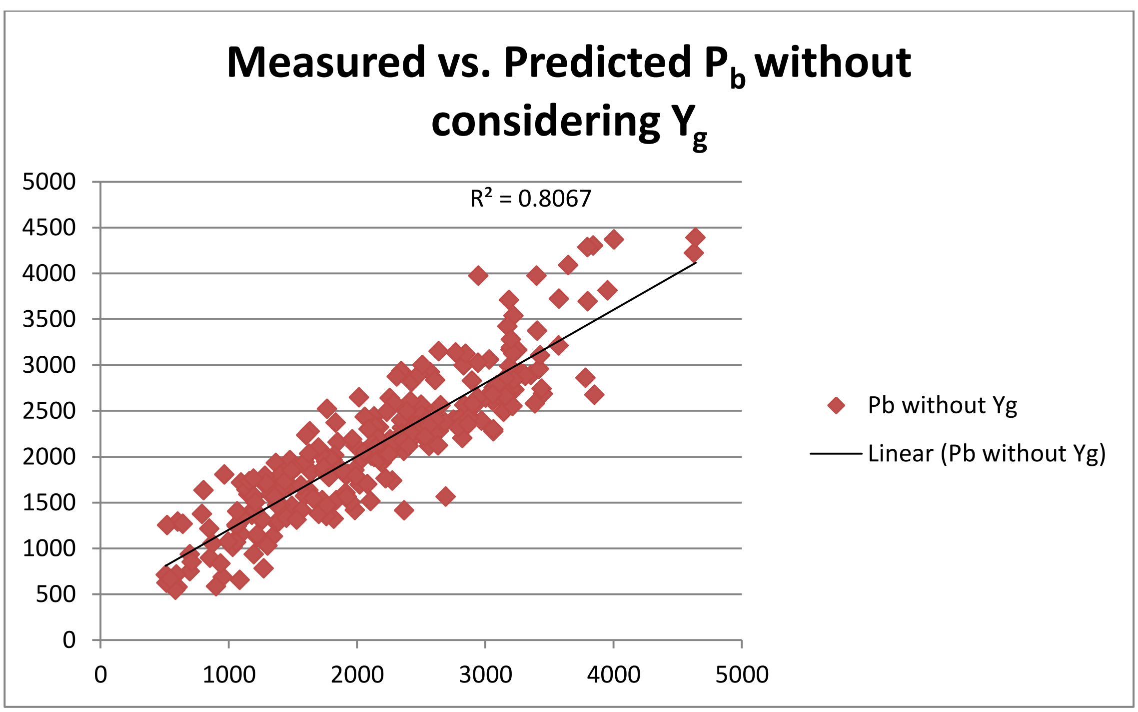

Figure 13 Measured vs. Predicted Pb without considering Yg

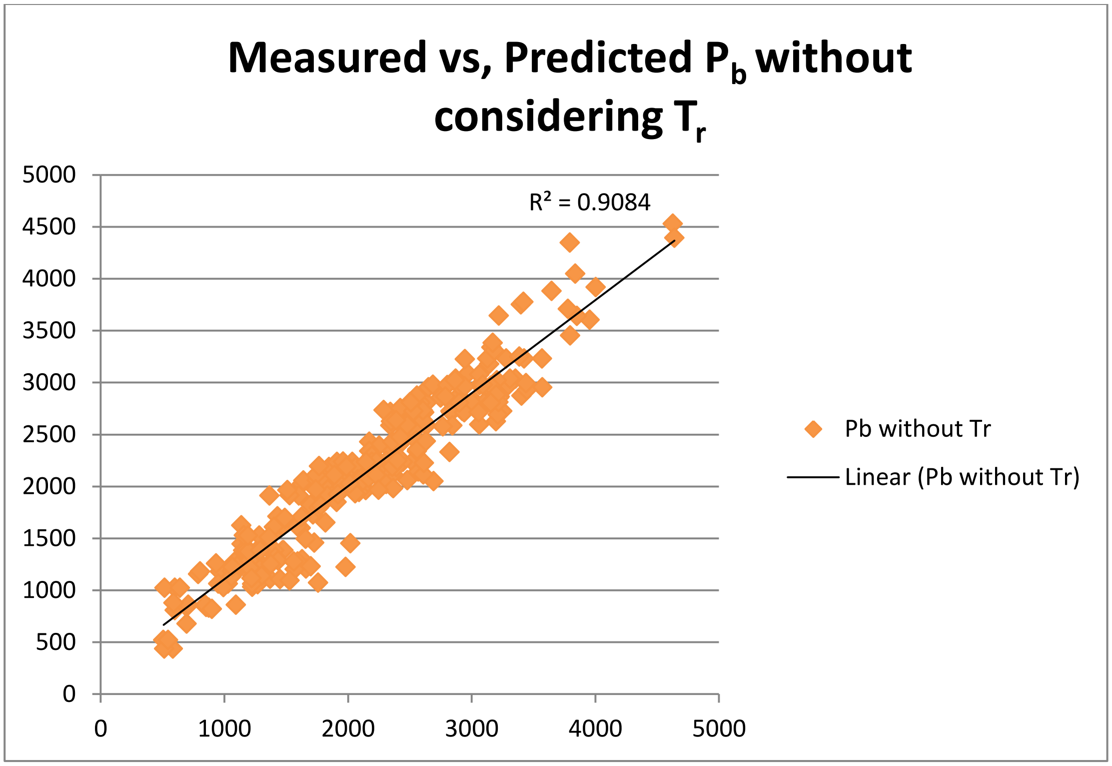

Figure 14 Measured vs, Predicted P_b without considering Tr

According to the statistical error analysis and the graphical analysis, the new P_b correlation without the reservoir temperature as one of the key inputs is nearly similar to the original P_b correlation. The withdrawal of Oil API also shows a good indication of P_b correlation with less number of key inputs. Nevertheless, the P_b correlation without reservoir temperature has higher value of squared correlation of coefficients which means the predictions is a lot closer to the measured P_b. Other reducing parameters however, do not yield the same capability as their RMSE, Emin, Emax, AARE, SD turn out to be bigger than this study's original correlation.

Table below compares this study 4 parameters P_b correlation against the Pb correlation without reservoir temperature as one of the correlating parameters.

Table 5 Comparison 4 parameters and 3 parameters GMDH Model Correlation

Statistical Error Analysis	4 parameters P_b correlation	P_b correlation without Reservoir Temperature
Emin	0%	0%
Emax	99%	97%
AARE	10%	12%
Standard Deviation	12%	11%
RMSE	16%	16%
R^2	0.9377	0.9084

4.4 NEW P_b GMDH MODEL PERFORMANCE EVALUATION AGAINST OTHER EXISTING CORRELATIONS

In this section, this study will compare the performance and accuracy of the new Pb correlation of 4 and 3 input parameters to other existence empirical correlations. For this matter, four correlations were selected to be use. The chosen correlations are:

- Standing's Correlation
- Al Marhoun's correlation
- Al Shammasi's correlation
- Dokla and Osman's correlation.

Equations describing those models may be reviews at the appendix at the end of this report.Figures below show the scatter diagrams of the predicted versus experimental Pb values. These cross plots indicates the degree of agreement between the experimental and the predicted values. If the degree is perfect, all the points should lay on the 45˚ line on the plot. The first two scatter diagrams show the tightest while the last scatter diagram shows the most scattered points.

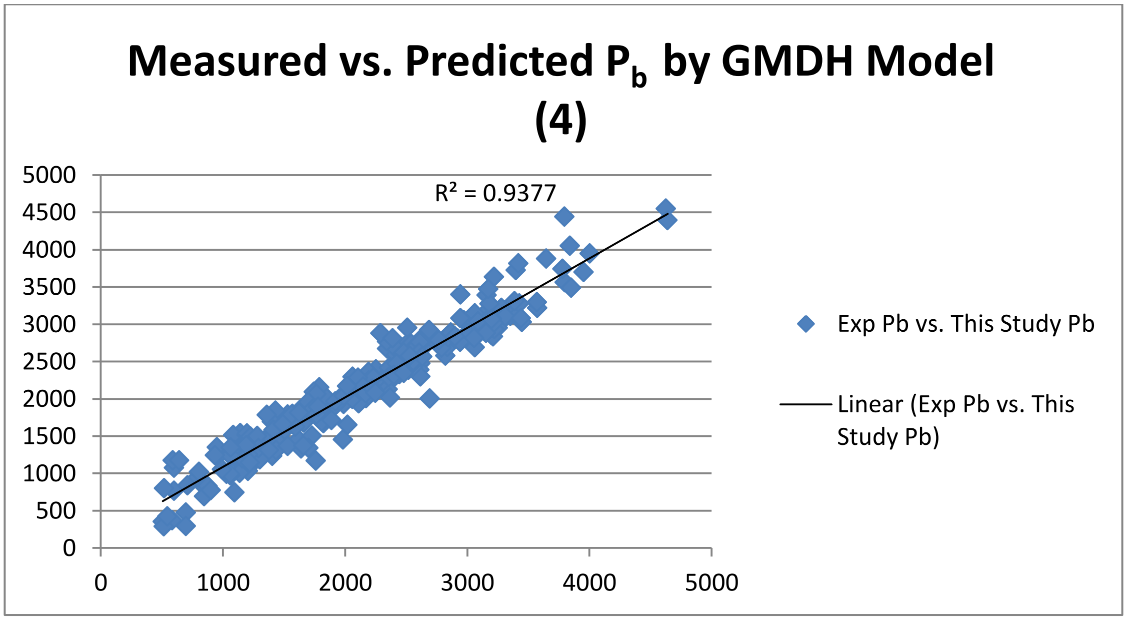

Figure 15 Measured vs. Predicted P_b by GMDH Model (4 parameters)

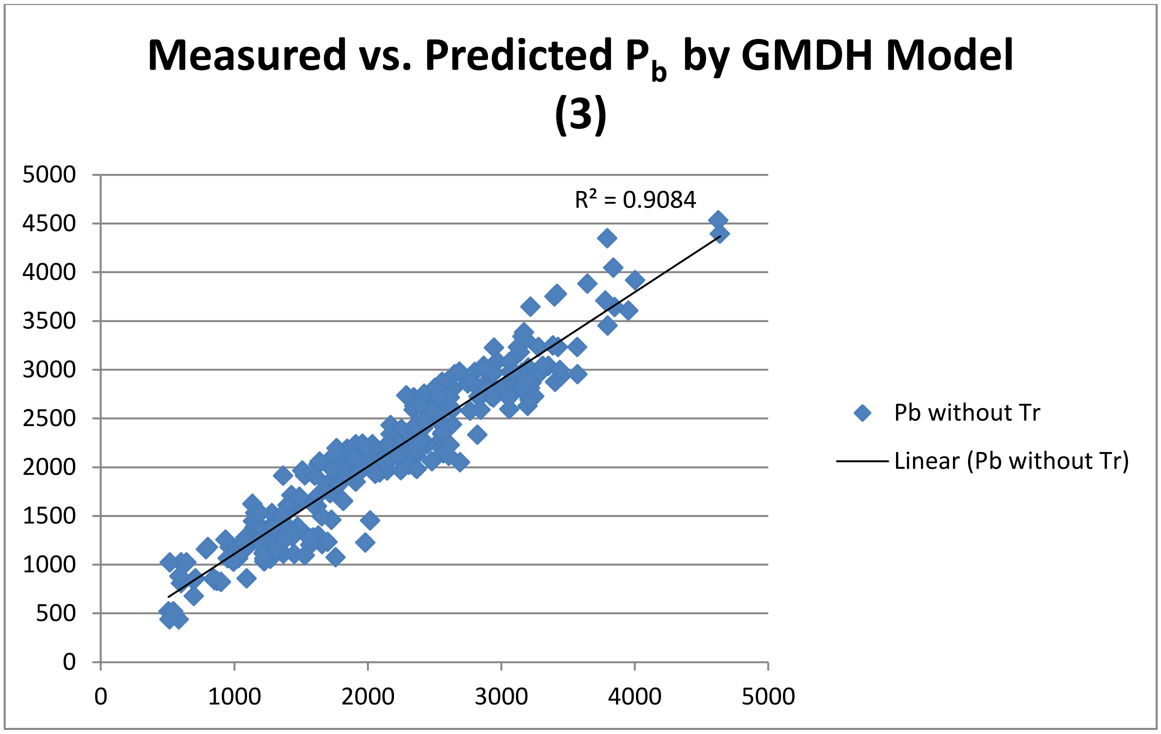

Figure 16 Measured vs. Predicted P_b by GMDH Model (3 parameters)

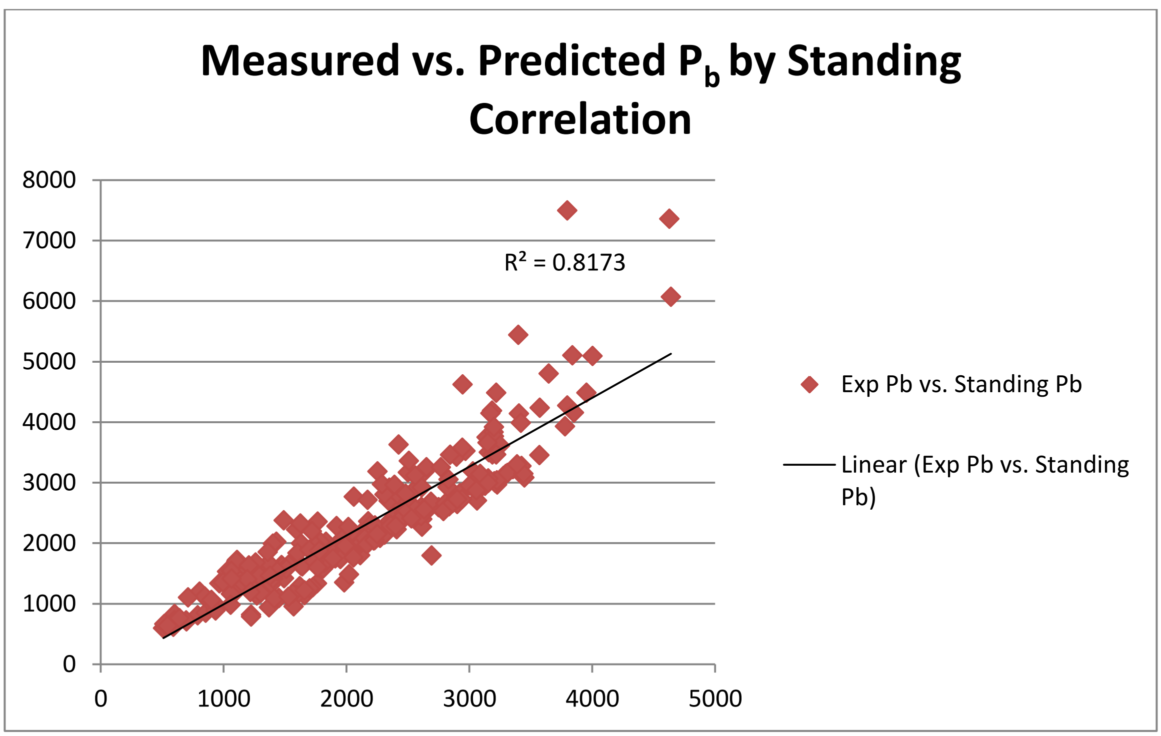

Figure 17 Current Data Regression Performance with Standing Correlation

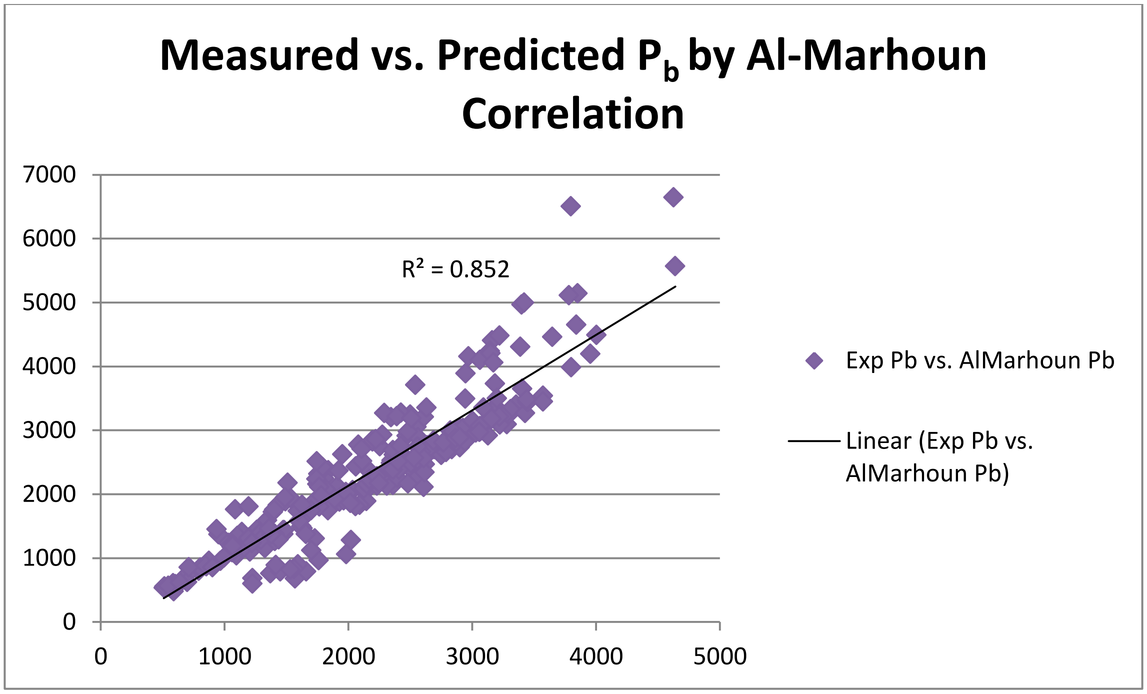

Figure 18 Current Data Regression Performance with Al- Marhoun Correlation

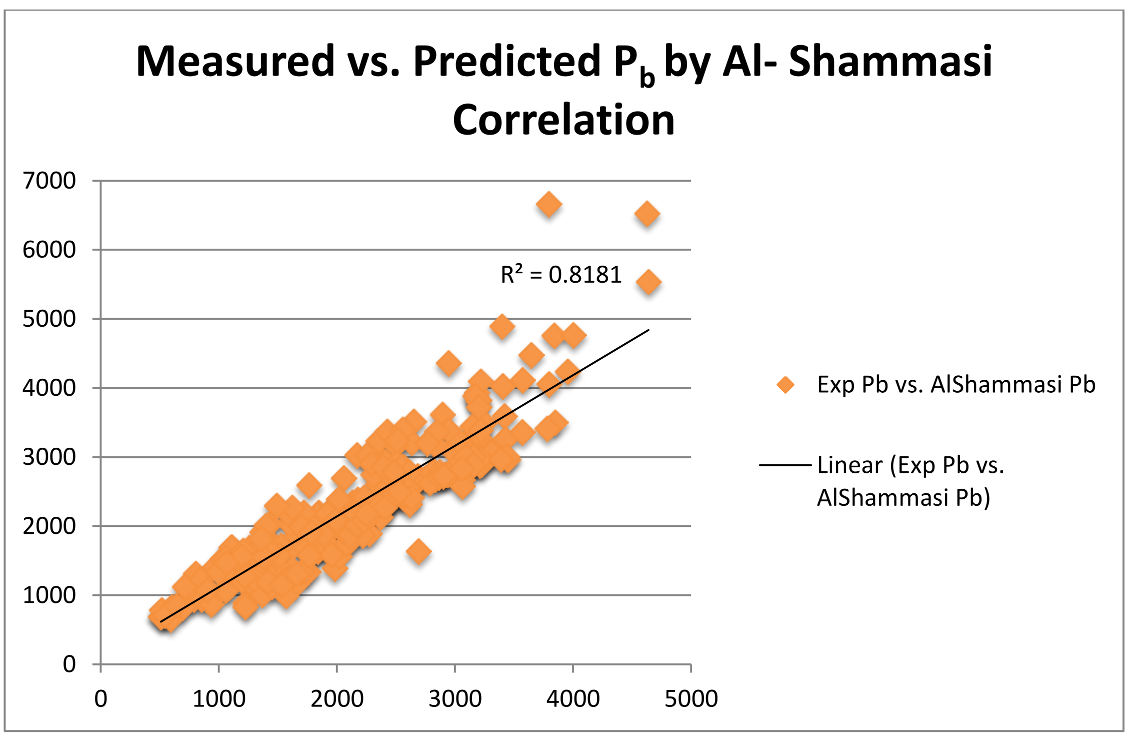

Figure 19 Current Data Regression Performance with Al-Shammasi Correlation

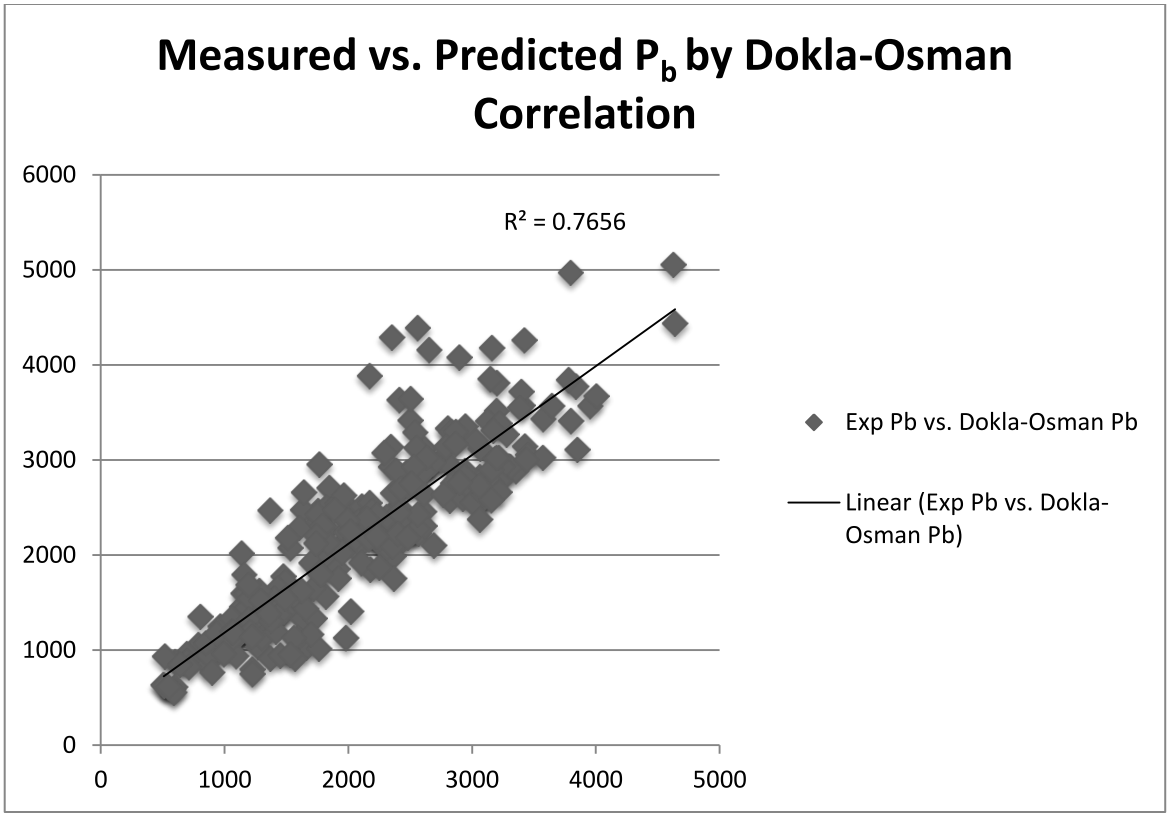

Figure 20 Current Data Regression Performance with Dokla-Osman Correlation

Below are the table of statistical error analysis and graphical analysis of the comparisons.

Table 6 Comparisons of Statistical Error Analysis with existing correlations

Statistical Error Analysis	**GMDH Model 4 parameters**	**GMDH Model 3 parameters**	**Standing.**	**AlMarhoun**	**AlShammasi**	**Dokla and Osman**
Emin	0%	0%	0%	0%	0%	0%
Emax	99%	97%	97%	71%	76%	83%
Eaverage	10%	12%	15%	14%	17%	17%
SD	12%	11%	14%	15%	14%	16%
RMSE	16%	16%	22%	20%	22%	24%
R^2	0.9377	0.9084	0.8173	0.852	0.8181	0.7656

According to the statistical error analysis, both of this study's models outperform all the empirical correlations. The proposed models achieved the lowest average absolute percent relative error (10% and 12%), the lowest standard deviation value (12% and 11%), and the lowest values for root means square error (both are 16%) and both of this study's correlation model are highest in squared correlation of coefficient (both are 0.9). However, both models also have the highest absolute percent relative error as high as 99% and 97%. Yet, other correlations also have quite a high error at this part.

Figures below show graphical comparisons of every statistical error analysis of both correlations against other empirical correlations.

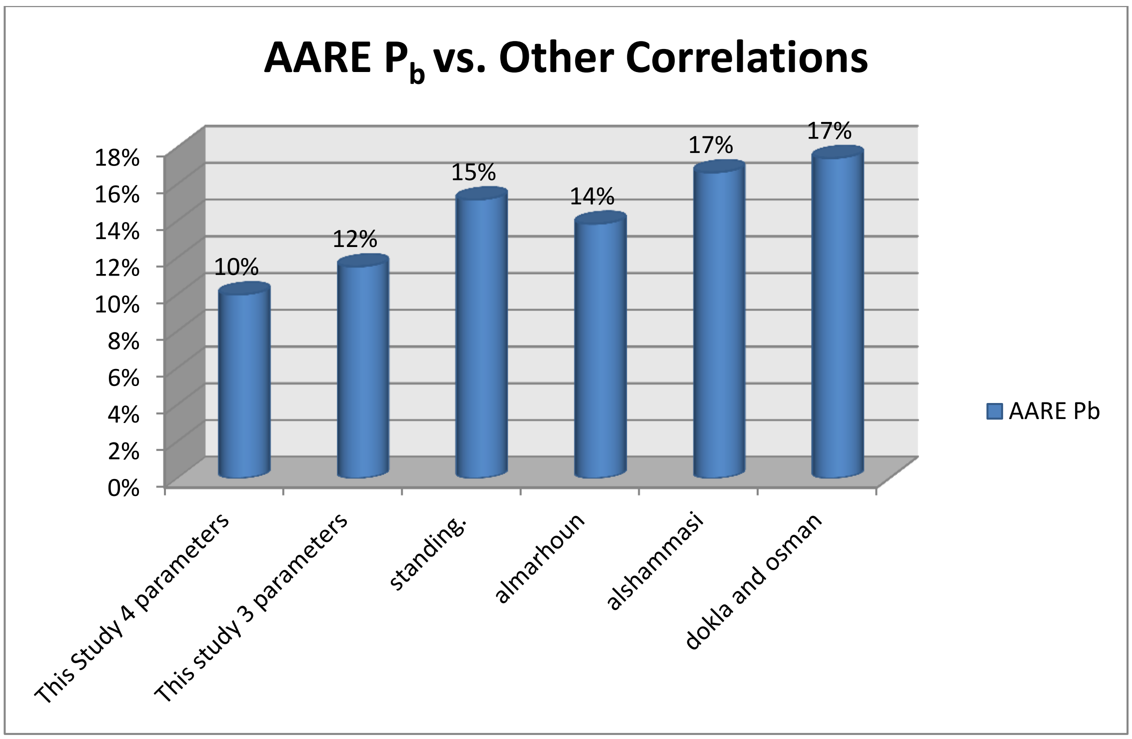

Figure 21 AARE P_b vs. Other Correlations

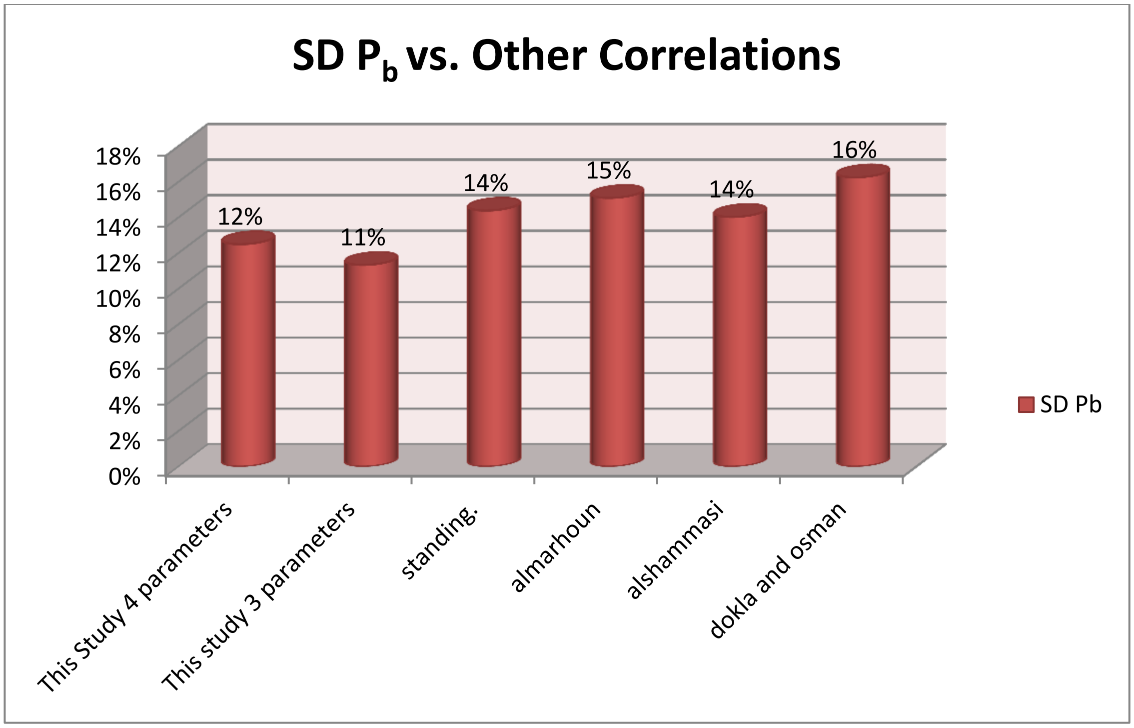

Figure 22 SD P_b vs. Other Correlations

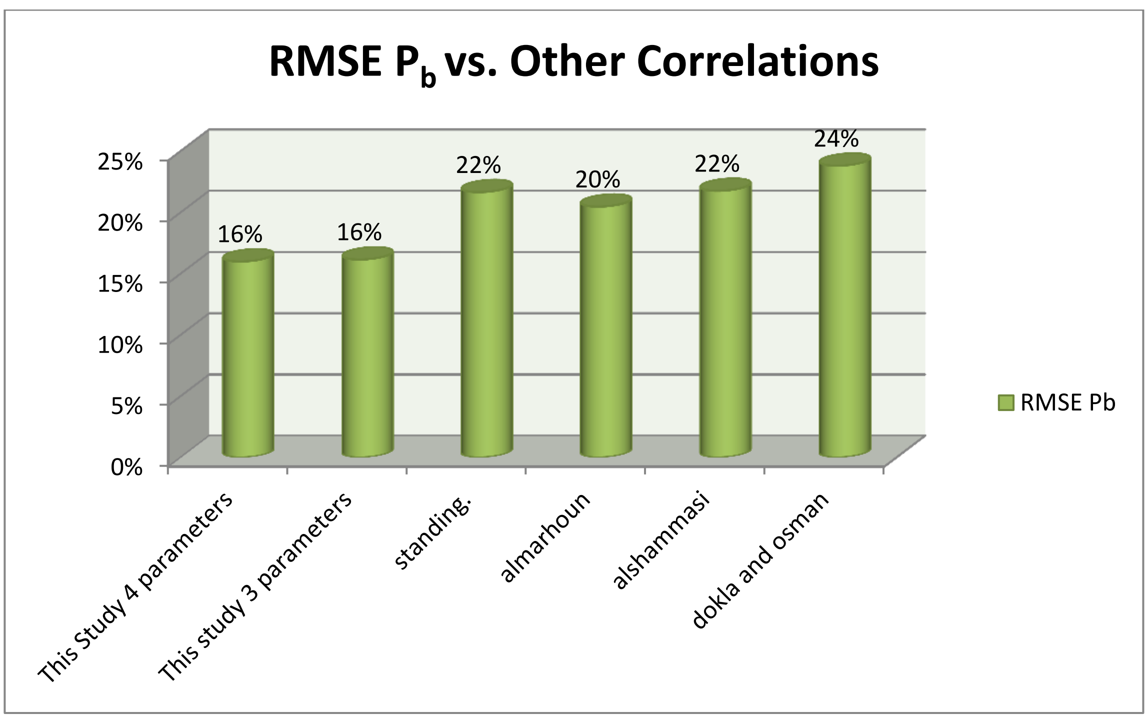

Figure 23 RMSE P_b vs. Other Correlations

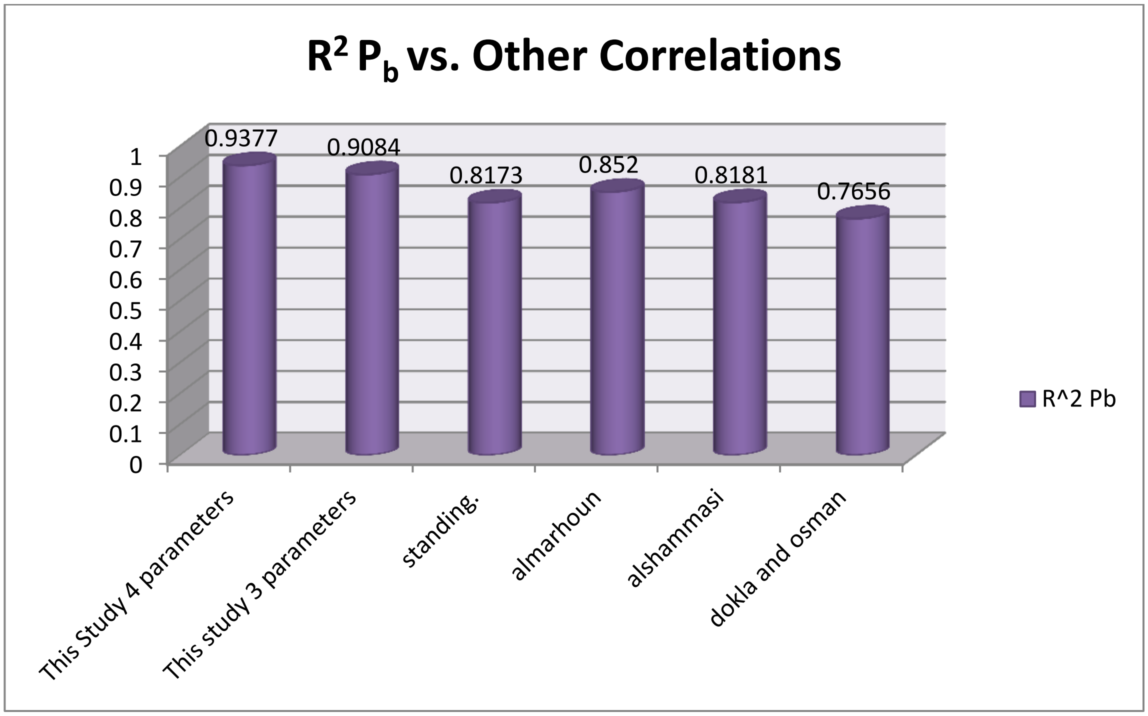

Figure 24 R^2 P_b vs. Other Correlations

The graphical analyses above are shown to indicate the superior performance of both GMDH model for P_b correlations against other empirical correlations.

4.5 GMDH CORRELATION MODEL: OIL FORMATION VOLUME FACTOR, B_o

The GMDH approach has develop numbers of equations for predicting the oil formation volume factor, B_o. The best correlation for estimating oil formation volume factor through GMDH approach has included only *two out of seven* gathered reservoir and PVT properties as the correlation's input. The selected input was:

- Oil Density
- Solution Gas-Oil Ratio, R_s

The model was the result of only 1 hidden layers of the network of all possible inputs to achieve the desired output; the Oil Formation Volume Factor. Diagram below shows the layers of input with the heaviest weightage to produce the wanted output.

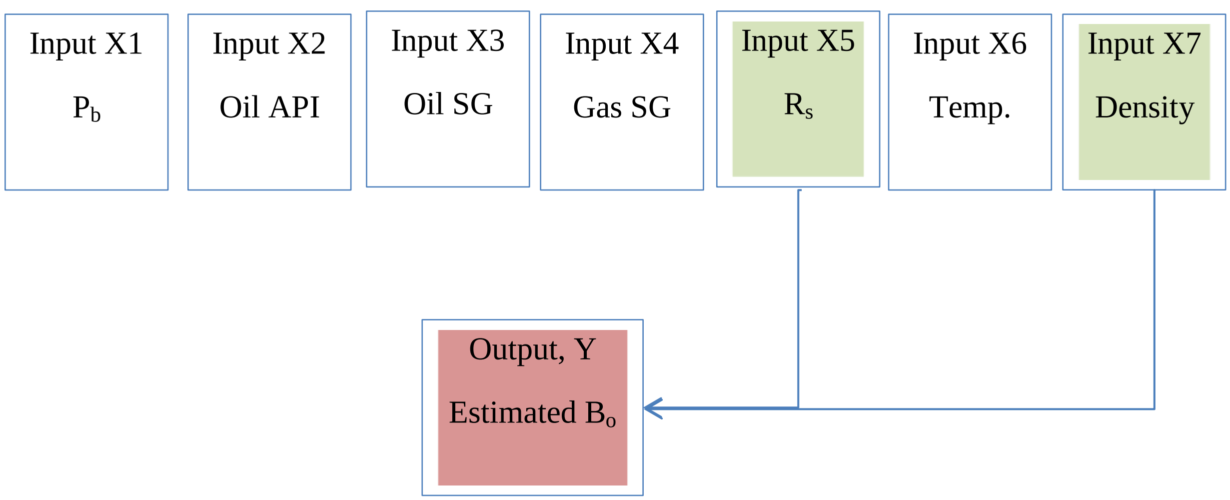

Figure 25 Layers of Inputs for B_o GMDH model correlation

Textbox below shows the model's equation to predict B_o as generated by MATLAB.

```
Model =

    numLayers: 1          d: 7       maxNumInputs: 2      p: 2

   inputsMore: 0   maxNumNeurons: 6       critNum: 2

Layer: [1x1 struct]

Time =



0.2989



Number of layers: 1

Layer #1

Number of neurons: 1

y = -0.000186487223862174 +0.0535558442790031*x7
+0.00178156763366563*x5 -3.13993421000316e-005*x5*x7 -
0.000623724463899113*x7*x7 -7.53941748412072e-008*x5*x5
```

Figure 26 GMDH Bo correlation coefficient

The correlation has only 6 coefficients for 1 layer of GMDH network with only 2 key parameters as input to predict the oil formation volume factor.

Since the inputs recognized by GMDH for Bo's correlation were only two, there will be no study on the effects of reducing correlating this study Bo parameters.

4.5.1 Bo STATISTICAL AND GRAPHICAL ERROR ANALYSIS

The table below summarizes the statistical error analysis for B_o GMDH model.

Table 7 B_o Statistical Error analysis

Statistical Error Analysis	This Study Bo correlation
Emin	0%
Emax	6%
Eaverage	0%
Standard Deviation	1%
RMSE	2%
R^2	0.9793

The error analysis shows small range of absolute relative errors up to 6% only while the average absolute error is maintain to be 0%. Standard deviation of the the proposed correlation for B_o is around 1% and the value of the proposed correlation RMSE is 2% which indicates a good measure of accuracy. Diagram below shows the scatter plot of measured B_o against the estimated B_o using this study's correlation indicating excellent agreement of experimental and estimated values of oil formation volume factor.

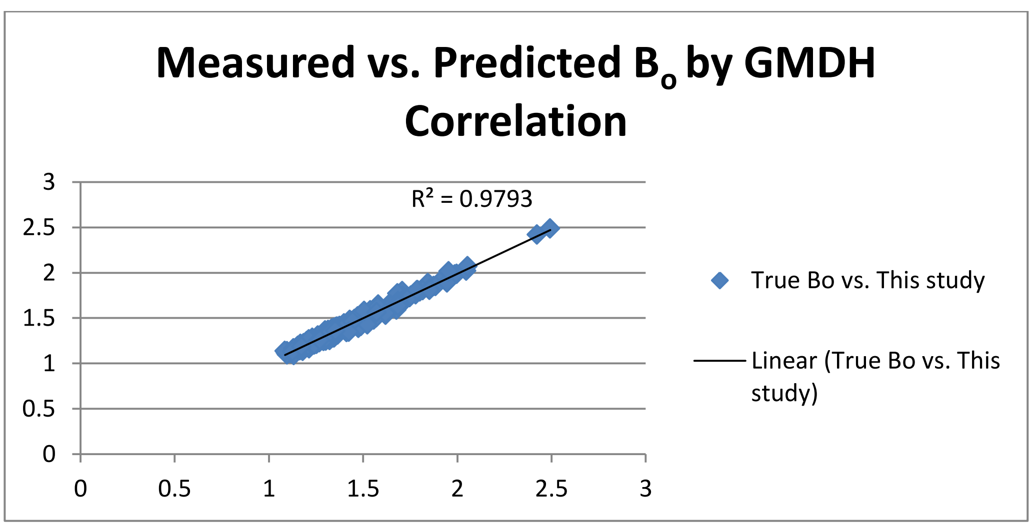

Figure 27 Measured vs. Predicted B_o by GMDH Correlation

4.6 NEW Bo GMDH MODEL PERFORMANCE EVALUATION AGAINST OTHER EXISTING CORRELATIONS

In this section, this study will compare the performance and accuracy of the GMDH Model for B_o to other existence empirical correlations. For this matter, four correlations were selected to be use. The chosen correlations are:

- Standing's Correlation
- Al Marhoun's correlation
- Al Shammasi's 3 parameters correlation
- Al Shammasi's 4 parameters correlation.

Equations describing those models may be reviews at the appendix at the end of this report.

Figures below show the scatter diagrams of the measured vs. predicted P_b values. These cross plots indicates the degree of agreement between the experimental and the predicted values. If the degree is perfect, all the points should lay on the 45° line on the plot. The first scatter diagrams (this study) show the tightest while the Standing's scatter plot shows the most spread points.

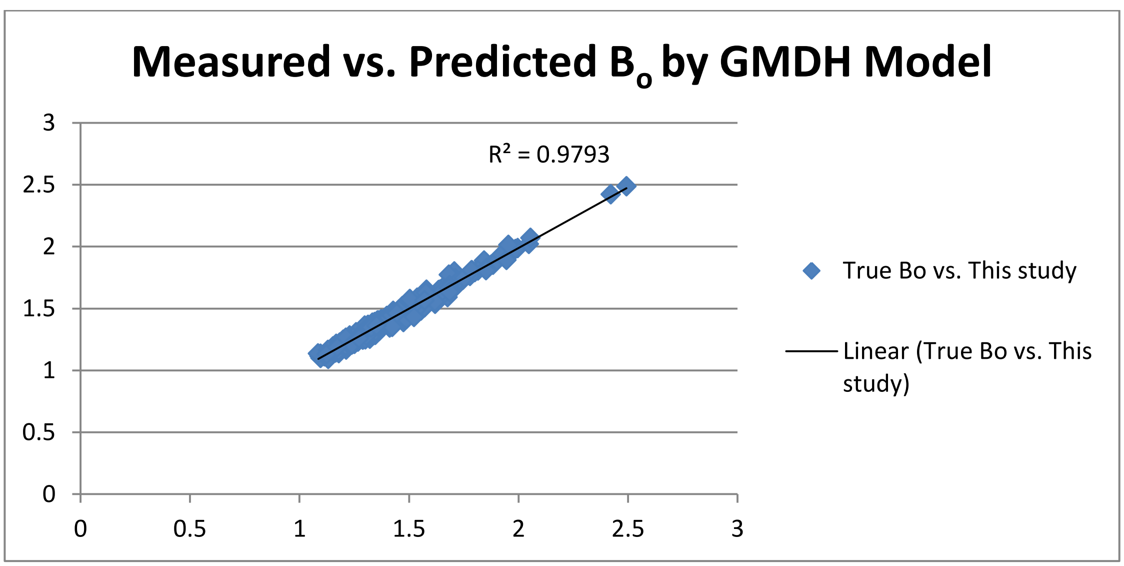

Figure 28 Cross plot of Measured vs. Predicted Bo by GMDH Model

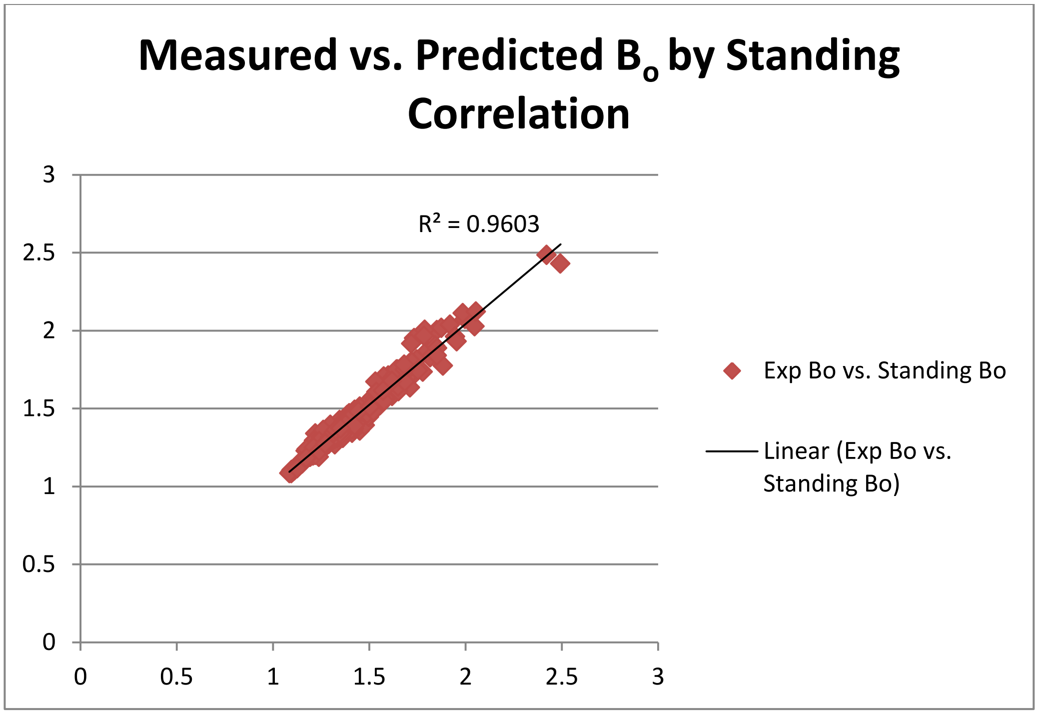

Figure 29 Current data regression performance with Standing Correlation

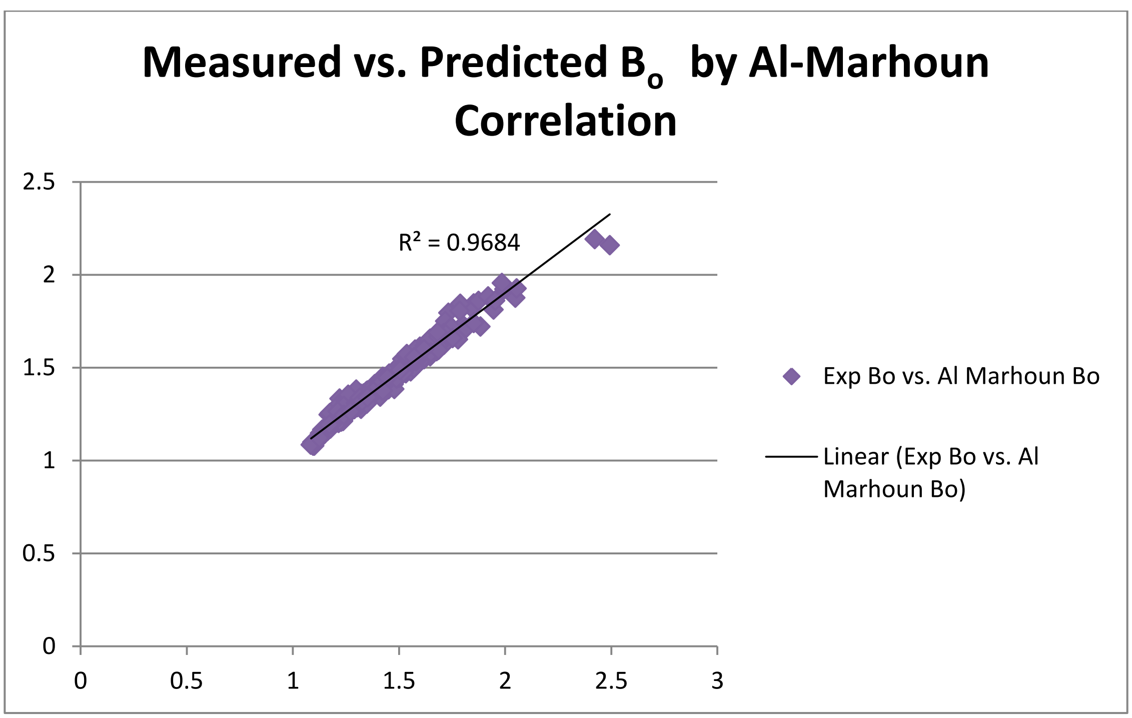

Figure 30 Current data regression performance with Al-Marhoun Correlation

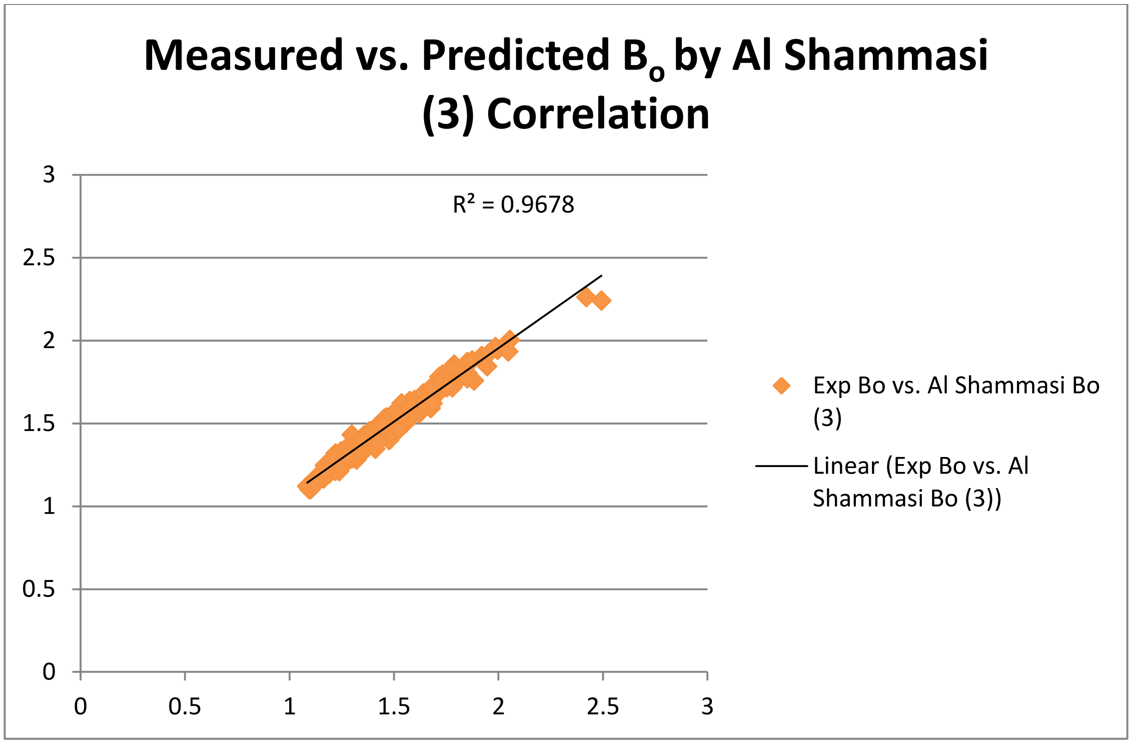

Figure 31 Current data regression performance with Al Shammasi Correlation 3 parameters

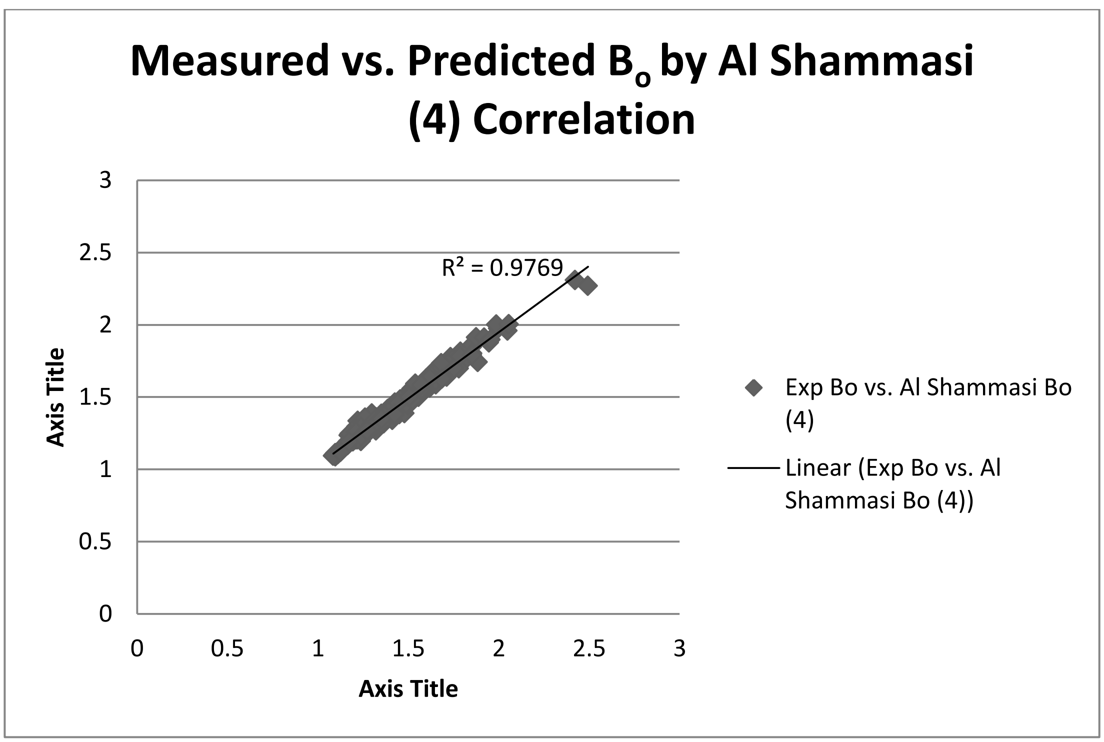

Figure 32 Current data regression performance with Al Shammasi Correlation 4 parameters

The table below summarizes this study B_o's correlation against other empirical correlations.

Table 8 Statistical Error Analysis for This study's B_o against other correlations

Statistical Error analysis	Bo GMDH Model	Standing	Al Marhoun	Al Shammasi 3 parameters	Al Shammasi 4 parameters
Emin	0%	0%	0%	0%	0%
Emax	6%	13%	13%	10%	9%
Eaverage	0%	3%	2%	3%	2%
Standard Deviation	1%	2%	2%	2%	2%
RMSE	2%	3%	3%	3%	2%
R^2	0.9793	0.9603	0.9684	0.9678	0.9769

Referring to the table 8 previously, this study's correlation for B_o also outperforms every other empirical correlation. The proposed model has a range of error up to only 6% while upholding low numbers for the correlation's standard deviation error (1%), RMSE (2%), and the correlation of coefficient of 0.973 which is extremely fit for both estimated and experimental value for Bo.

Next figures are the graphical error analysis for this study's GMDH modeling for Bo correlation against other correlations;

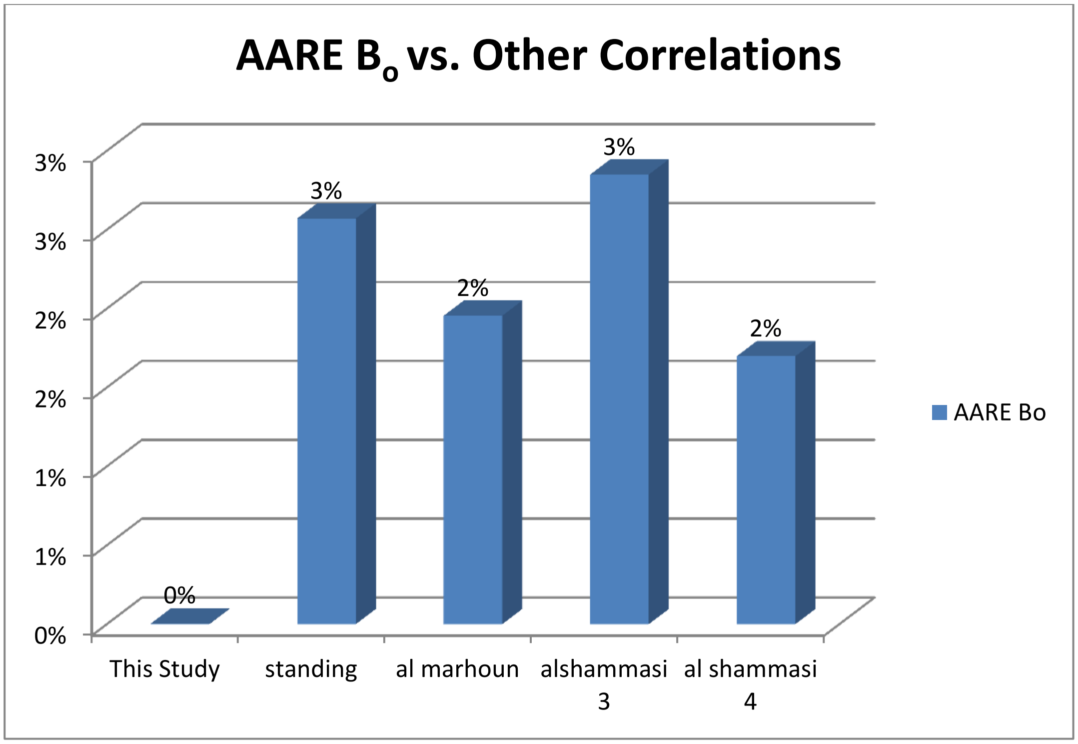

Figure 33 Comparison of AARE for this study against other correlations

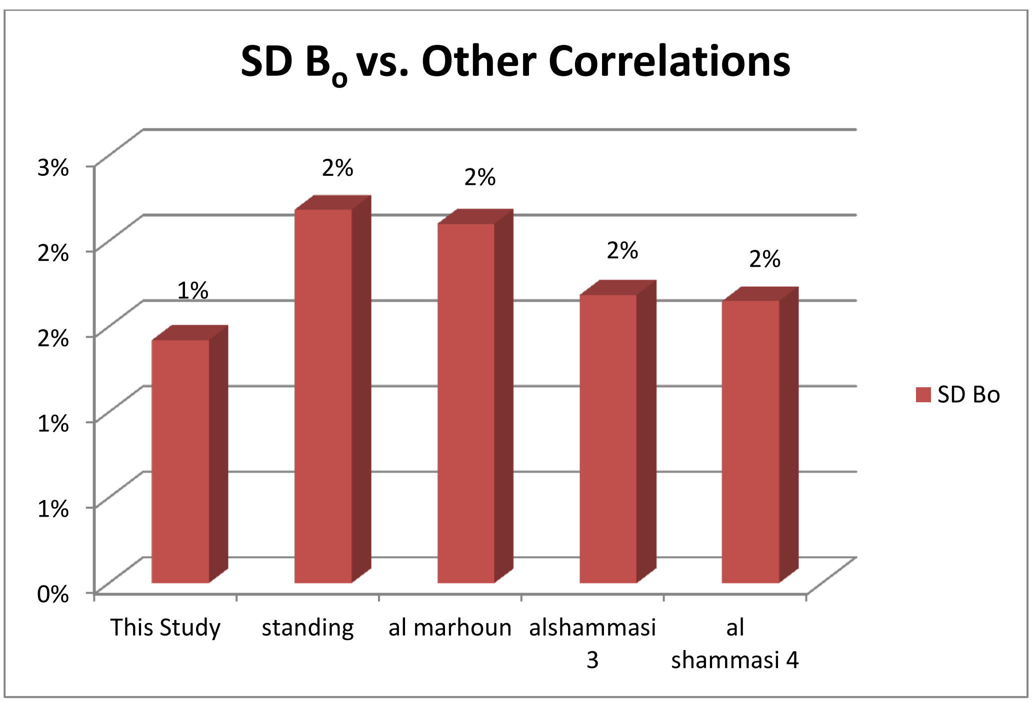

Figure 34 Comparison of SD for this study against other correlations

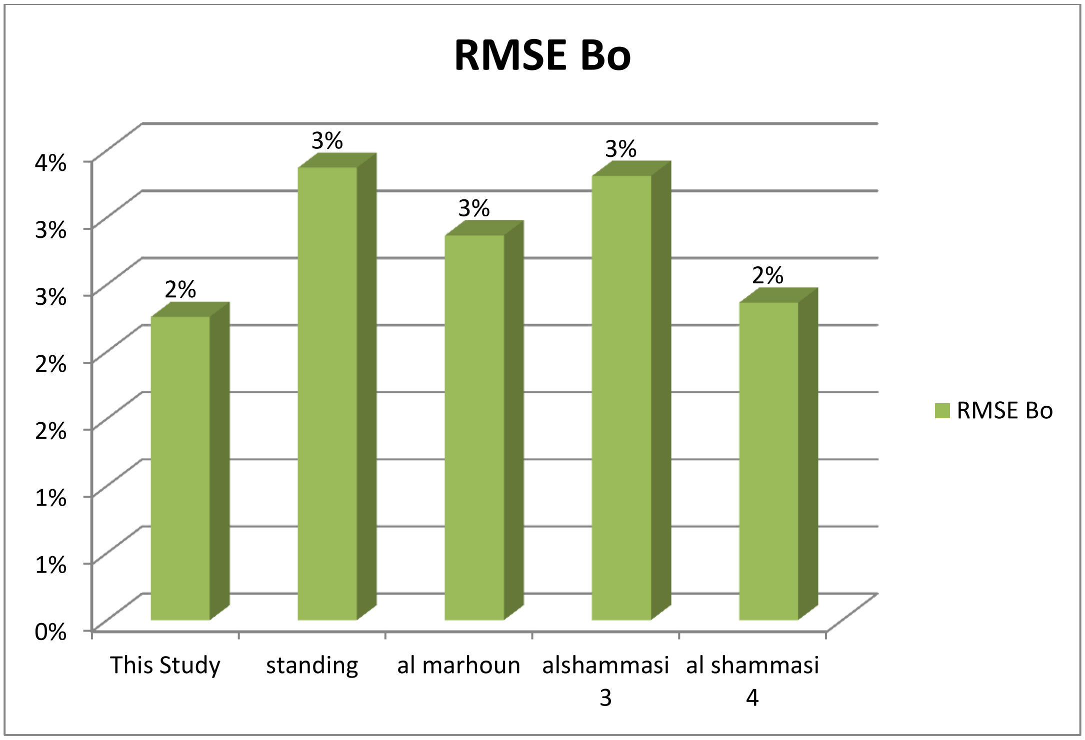

Figure 35 Comparison of RMSE for this study against other correlations

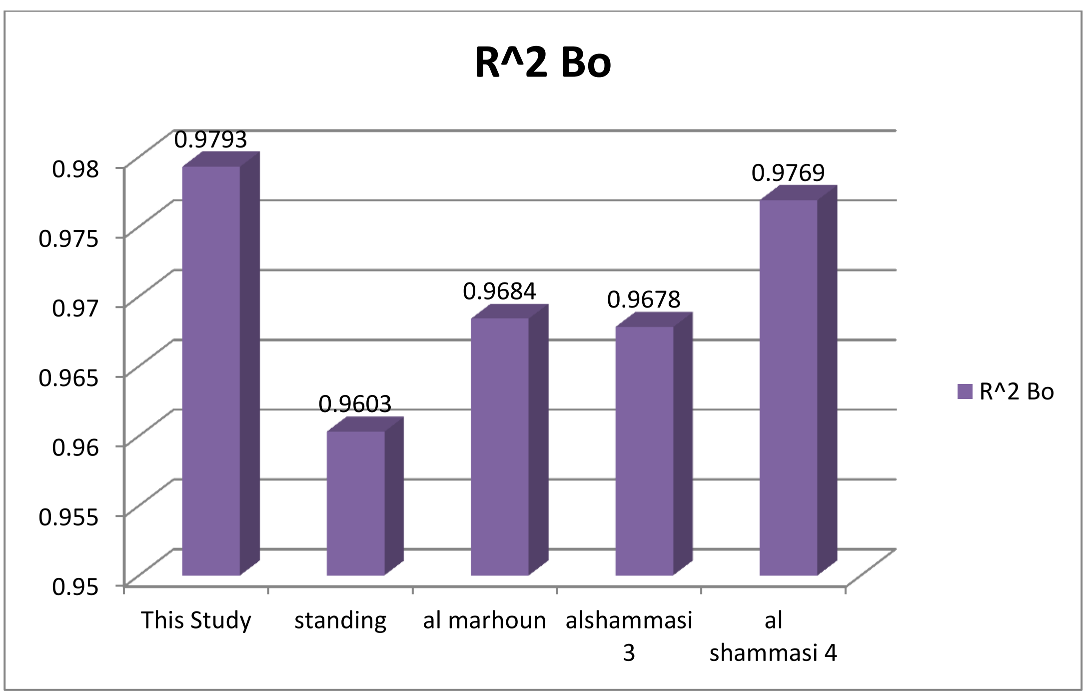

Figure 36 Comparison of R^2 for this study against other correlations

CHAPTER 5

CONCLUSIONS

Correlations for P_b and B_o have been known from the literature review, to be very important and crucial in the early stages of developing an oilfield. Reservoir fluid analysis assist engineers to understand the basic requirement of oil reserve; because the rule of thumb- every well is unique.

This study has achieved all of its objectives set in the earlier chapter of this dissertation. The new correlations for both Bubble point pressure P_b, and Oil Formation Volume Factor B_o, outperform other tested empirical correlations. On top of that, this study also successfully manages to study the effect of reducing the parameters used for the GMDH build correlation.

The author's proposal for his final year project has an impact directly towards the reservoir engineering side.

REFERENCES

[1] E.A Osman and R.E Abdel-Aal, "Abductive Networks: A new modelling tool for the oil and gas industry," 2002.

[2] W. McCain, "Reservoir fluids properties correlations- State of the art," 1991.

[3] Hemmati and Kharrat, "A correlation approach for prediction of crude oil PVT properties," 2007.

[4] T. Ahmed, in *Reservoir Engineering Handbook*, Gulf Professional, 2010.

[5] T. Ahmed, "Oil formation volume factor," in *Reservoir Engineering Handbook*, Gulf Professional, 2010.

[6] Katz, "Prediction of shrinkage of crude oils," pp. 137-147, 1942.

[7] Standing, "A pressure volume temperature correlation for mixture of California Oils and Gas," 1947.

[8] Vasquez and Beggs, "Correlations for fluid physical properties prediction," *Journal of Petroleum Technology,* 1980.

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

[10] S. Elmabrouk, "Prediction of bubble point pressure and bubblepoint oil formation volume factor in the absence of PVT analysis," 2010.

[11] P. Bandyopadhyay, "Improved estimation of bubble point pressure of crude oils: Modelling by regression analysis," 2011.

[12] M. B. Soave, "A pressure volume temperature correlation for mixtures of california oils and gases.," 1972.

[13] D. Y. P. a. D. B. Robinson, "A new two constant equation of state," 1976.

[14] M. I. Omar and A. C. Todd, "Development of New Modified Black Oil Correlations for Malaysian Crudes," 1993.

[15] M. A. Marhoun, "PVT correlations for crude oils," 1988.

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

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

[18] Y. B. L. H. S. a. T. Y. S. Lee, "Abductive Network for Predicting Tool life in drilling," 1995.

[19] M. Abdalla, "Evaluation of Below Bubble Point Viscosity Correlations and Construction of a new neural network model.," 2007.

[20] R. O. A. D. a. A. B. V. A.A. Semenov, "A Hybrid Artificial Intelligence Method for the Optimization of Integrated Gas Production System," 2010.

[21] N. G.-. A. B. a. H. M. A. Mohammad, "“Prediction of Pipeline Scour Depth in Clear-Water and Live-Bed Conditions Using Group Method of Data Handling”.," 2010.

APPENDICES

1. PVT Data Points used for this study.

y	x1	x2	x3	x4	x5	x6	x7
Pb	API	Oil SG	Gas SG	Rs	Temp	Density	Bob
2035	33.3	0.858617	0.815	585	100	47.22	1.272
3279	32.7	0.861754	0.802	898	125	44.45	1.43
1847	39.1	0.829426	0.929	805	100	44.65	1.387
1755	49.5	0.781768	0.79	694	190	38.00	1.48
1630	26.1	0.897843	0.933	347	165	50.23	1.203
2822	46.8	0.793606	0.876	1006	280	36.29	1.695
3160	45.4	0.799887	0.705	1213	186	36.05	1.707
1477	38.6	0.831864	1.002	560	150	44.87	1.327
3840	33.9	0.8555	0.838	1408	216	38.55	1.801
2636	39.4	0.82797	0.951	1143	200	40.35	1.647
1988	41.9	0.816032	0.876	692	150	43.03	1.375
1769	49.1	0.783499	0.765	585	204	39.24	1.401
2639	32.2	0.864386	0.774	700	100	46.34	1.323
1141	35.4	0.8478	0.98	446	190	44.08	1.335
1620	42.9	0.811353	0.847	404	188	43.70	1.265
2360	48.4	0.786548	1.014	993	267	36.58	1.716
2692	38.6	0.831864	0.631	393	179	44.94	1.23
1378	31.2	0.869699	0.98	417	160	47.86	1.25
1153	31.2	0.869699	0.98	417	100	49.53	1.208
1472	31.2	0.869699	0.98	417	185	47.22	1.267
1928	28.8	0.88272	0.824	469	100	49.13	1.228
2365	30.1	0.875619	0.798	498	175	46.95	1.279
2402	40.7	0.821719	0.919	844	242	38.19	1.619
3354	34.2	0.853953	0.779	825	185	43.35	1.431
1225	38	0.834808	1.168	260	211	48.05	1.17
1180	28.4	0.884928	0.921	331	100	51.35	1.156
3066	35.4	0.847813	0.799	867	140	43.89	1.42
3405	42.8	0.811819	0.93	1579	235	35.37	1.997
2132	30.1	0.875619	0.801	521	110	48.64	1.24
1265	31.2	0.869699	0.98	417	130	48.68	1.229
1085	29.1	0.881071	0.638	169	187	50.04	1.128
1480	31	0.870769	0.973	412	180	46.71	1.28
4627	37.4	0.8378	0.825	2217	252	30.95	2.493

2896	42.8	0.811819	0.93	1579	145	38.14	1.852
1824	41.9	0.816032	0.876	692	115	44.02	1.344
2344	39.4	0.82797	0.951	1143	150	41.56	1.599
2617	36.5	0.842262	0.812	811	100	44.87	1.371
2521	36.1	0.844272	0.907	746	200	42.98	1.44
1230	28.9	0.88217	0.931	302	160	49.55	1.188
1271	29.2	0.880523	0.775	198	187	50.07	1.139
1631	36.2	0.843769	1.013	803	100	45.61	1.397
1325	32.1	0.8649	1.145	439	213	45.21	1.345
1345	36.3	0.8433	0.923	390	254	42.17	1.364
2350	37	0.839763	0.818	680	169	44.35	1.352
790	39.8	0.826036	1.005	274	150	47.34	1.168
2133	39.1	0.829426	0.929	805	150	43.25	1.432
2061	34.5	0.8524	0.936	737	234	40.82	1.533
1195	31.9	0.865973	0.664	214	180	48.58	1.152
3142	33.3	0.858617	0.723	761	247	41.15	1.484
1492	37.4	0.837774	0.716	341	159	46.29	1.201
2172	43.6	0.80811	1.008	1493	100	40.88	1.734
966	31.2	0.869699	1.188	433	150	49.21	1.245
2751	32	0.865443	0.8	750	100	46.63	1.333
1810	50.5	0.777473	0.77	606	189	38.55	1.423
2254	31.8	0.8665	0.923	765	243	40.92	1.556
1838	34.8	0.850872	0.664	366	153	46.69	1.208
3198	44.6	0.803521	0.96	1602	230	35.78	1.986
3057	32	0.865443	0.778	679	175	44.63	1.371
2836	36.5	0.842262	0.812	811	140	43.84	1.403
2413	40.2	0.824112	0.925	1203	100	42.23	1.576
2445	33.3	0.858617	0.815	585	180	45.19	1.329
2344	40.4	0.823153	0.743	791	184	41.54	1.429
2925	33.2	0.859138	0.774	693	175	43.32	1.406
2256	33.3	0.858617	0.815	585	140	46.20	1.3
2145	47.9	0.78874	1.045	1022	216	37.56	1.697
1912	33.3	0.858617	0.815	585	80	47.78	1.257
854	32.1	0.864914	0.942	196	175	49.50	1.141
1095	31.2	0.869699	1.188	433	190	48.32	1.268
1405	31	0.870769	0.973	412	160	47.49	1.259
2900	34.2	0.853953	0.789	818	100	45.47	1.365
1641	41.9	0.816032	0.876	692	80	45.06	1.313
1805	48.1	0.787862	0.767	599	204	38.91	1.424
1370	38.2	0.833824	1.174	313	205	47.84	1.192
1030	28.2	0.886	1.055	333	230	45.43	1.322

2568	36.6	0.8416	1.036	941	230	39.22	1.677
3127	32.7	0.861754	0.802	898	100	45.05	1.411
3184	31.2	0.8697	0.865	1018	226	40.22	1.647
2274	45.2	0.800792	0.689	546	245	37.96	1.451
1367	39.3	0.828454	1.004	755	80	46.03	1.347
2500	48.8	0.784803	0.877	1355	228	35.34	1.843
1603	39.3	0.828454	1.004	755	125	44.70	1.387
3954	36.6	0.842	0.844	1325	218	37.31	1.816
3228	34.4	0.852923	0.783	775	175	43.51	1.413
2392	39.1	0.829426	0.929	805	200	41.87	1.479
1179	34.5	0.8524	1.048	406	220	44.21	1.334
2652	39.3	0.828454	0.951	1507	100	41.44	1.718
2609	40.4	0.823153	1.038	1019	198	40.54	1.622
2540	50.4	0.7779	0.73	1020	239	34.27	1.712
1180	31	0.870769	0.973	412	100	49.17	1.216
1390	33.4	0.858096	0.718	287	141	48.83	1.154
2687	29.7	0.877792	0.755	680	100	47.36	1.304
1110	29.5	0.8789	1.087	409	234	45.85	1.328
2588	30.8	0.871842	0.766	665	100	47.77	1.284
1910	32.6	0.862279	0.733	384	152	46.55	1.238
2417	39.6	0.827	0.899	889	220	39.00	1.602
3297	35.4	0.847813	0.799	867	180	42.75	1.458
601	29.0	0.8816	1.29	209	218	48.26	1.216
2616	37.3	0.83827	0.842	667	177	43.72	1.371
508	27.5	0.889937	1.072	141	130	51.88	1.11
2177	21.9	0.922425	0.799	421	145	51.22	1.213
1660	37.1	0.839265	1.298	421	203	48.98	1.221
2350	44.6	0.803521	0.96	1602	100	39.72	1.789
2310	35.2	0.8488	1.063	882	229	40.57	1.62
1981	30.1	0.875619	0.798	498	100	48.97	1.226
3798	36.6	0.842	0.851	1260	218	39.23	1.711
2504	39.9	0.825554	0.894	1151	100	42.32	1.548
3187	40.3	0.8237	0.861	1102	228	37.67	1.707
2290	43.1	0.810424	0.801	990	208	37.12	1.653
2111	53.2	0.766107	0.74	692	220	37.23	1.471
2425	31.3	0.8693	0.873	816	250	40.70	1.571
1530	45.2	0.800792	0.817	566	185	42.17	1.334
2804	35.4	0.847813	0.799	867	100	45.03	1.384
2090	48.2	0.787423	1.05	1011	210	37.84	1.68
3399	38.0	0.8348	0.851	1561	268	34.26	2.048
3573	39.3	0.828454	0.951	1507	225	37.97	1.875

952	26.9	0.893308	0.667	142	146	52.23	1.092
3201	42.8	0.811819	0.93	1579	190	36.79	1.92
1094	22.8	0.917045	1.058	265	185	51.73	1.18
1405	26.1	0.897843	0.933	347	100	51.87	1.165
1982	36.1	0.844272	1.14	415	224	47.45	1.246
1261	28.4	0.8849	0.987	364	215	46.59	1.29
2970	34.6	0.851896	0.707	737	239	41.69	1.445
3420	42.3	0.814154	0.685	1212	194	36.89	1.683
804	31.2	0.869699	1.188	433	100	50.42	1.215
584	25.1	0.903576	1.025	127	160	52.20	1.114
1591	32.2	0.8644	1.054	583	239	42.23	1.475
2470	40	0.825073	0.758	760	166	41.51	1.429
3030	39.9	0.825554	0.894	1151	180	40.04	1.636
2020	39.2	0.82894	1.051	491	211	44.47	1.321
3311	34.2	0.853953	0.779	825	175	43.53	1.425
2172	33	0.860182	0.803	602	100	47.33	1.273
696	32.1	0.864914	0.942	196	100	51.49	1.097
3101	32.2	0.864386	0.774	700	175	44.55	1.376
1962	36.1	0.844272	0.907	746	100	45.71	1.354
2550	48.9	0.784368	0.858	1170	231	33.23	1.884
2165	46.6	0.794497	0.916	856	211	39.71	1.517
1104	30.2	0.8751	1.069	408	232	44.98	1.346
1758	48.4	0.786548	0.762	628	199	38.55	1.442
874	27.2	0.891619	0.989	232	160	51.00	1.152
1205	28.2	0.886036	1.002	389	80	51.48	1.177
1818	26.6	0.895003	0.704	285	152	50.80	1.153
601	37.3	0.83827	1.192	266	145	47.54	1.191
1137	38.6	0.831864	1.002	560	74	46.92	1.269
1430	35.8	0.8456	0.958	554	226	40.58	1.478
1951	37.5	0.837278	0.627	367	173	45.02	1.23
2946	36.9	0.8403	0.924	1439	240	36.24	1.946
1750	48.7	0.785239	0.82	714	189	37.97	1.5
3204	32.6	0.862279	0.752	742	160	44.75	1.372
1159	37	0.839763	1.01	512	100	47.10	1.262
3250	40.2	0.824112	0.925	1203	240	38.10	1.747
1062	32.0	0.8654	1.09	393	234	44.65	1.34
2558	33	0.860182	0.803	602	170	45.54	1.323
697	27.9	0.887704	1.031	189	80	52.67	1.102
515	25.1	0.903576	1.025	127	120	53.06	1.096
1760	31	0.870769	1.195	372	211	49.41	1.222
1834	39.3	0.828454	1.004	755	170	43.51	1.425

1302	31.4	0.868631	0.824	242	180	48.64	1.17
2194	42.9	0.811353	0.75	664	214	39.92	1.438
3851	34.1	0.854469	0.663	819	243	41.41	1.466
1790	47.1	0.792273	0.8	686	224	38.04	1.496
2562	42	0.815562	0.795	741	234	39.51	1.491
1510	47.8	0.78918	0.73	522	189	39.87	1.365
2081	44.5	0.803977	0.677	494	230	41.61	1.315
2058	48.8	0.784803	0.939	765	205	38.65	1.52
2480	38.2	0.833824	0.737	686	171	43.41	1.357
3063	32.2	0.864386	0.628	586	180	45.80	1.287
2310	38.3	0.833333	0.801	636	161	43.81	1.345
847	22.8	0.917045	1.058	265	100	53.92	1.132
710	29.4	0.8794	1.144	265	216	47.12	1.252
1744	40.5	0.822674	0.727	524	190	42.65	1.325
3148	50.3	0.778328	0.788	1440	250	32.75	1.954
2831	40.2	0.824112	0.925	1203	160	40.53	1.642
3647	34.0	0.855	0.831	1295	218	39.48	1.722
3212	40.3	0.8236	0.806	886	219	39.78	1.536
3220	36.4	0.8428	0.798	1184	238	36.78	1.779
3160	33.1	0.85966	0.757	730	175	43.94	1.392
2408	38.6	0.831864	0.821	683	166	43.02	1.384
1061	28.9	0.88217	0.931	302	100	51.10	1.152
518	37.3	0.83827	1.192	266	105	48.68	1.163
1766	38	0.834808	1.056	1087	100	44.16	1.533
3426	32.7	0.861754	0.802	898	150	43.81	1.451
2193	45.3	0.800339	0.717	634	214	39.38	1.425
3780	40.2	0.824112	0.658	1023	209	38.32	1.581
2368	32.5	0.862805	0.756	440	235	45.52	1.282
3172	37.6	0.8368	0.825	1186	230	37.38	1.753
1719	31.7	0.867	0.975	554	216	43.39	1.416
1593	39.8	0.826036	1.181	421	203	45.98	1.268
1728	41.8	0.816503	0.941	397	215	44.50	1.259
2390	43.2	0.80996	0.811	956	226	39.72	1.538
2611	39.6	0.827002	0.789	810	225	39.54	1.525
1570	39	0.829912	1.315	366	207	47.00	1.241
1450	35.4	0.847813	1.25	359	208	48.61	1.214
1197	36.0	0.8448	1.05	457	220	41.96	1.412
3223	32	0.865443	0.8	750	175	44.82	1.387
1401	31.7	0.867	0.959	490	212	45.08	1.342
2259	30.1	0.875619	0.801	521	135	47.98	1.257
2360	40	0.825073	0.765	694	167	45.19	1.299

2871	34.2	0.853953	0.779	825	100	45.34	1.368
1225	38	0.834808	1.263	267	211	48.20	1.176
2231	36.1	0.844272	0.907	746	150	44.27	1.398
3155	34.2	0.853953	0.789	818	170	43.49	1.427
4004	33.6	0.8571	0.861	1417	219	37.82	1.853
2607	32	0.865443	0.778	679	100	46.53	1.315
1282	36.5	0.842262	0.96	469	155	45.45	1.291
2221	45.3	0.800339	0.693	547	238	40.45	1.362
2559	42.8	0.811819	0.93	1579	100	39.55	1.786
1414	41	0.82029	1.155	425	185	46.33	1.249
3003	30.8	0.871842	0.766	665	175	45.77	1.34
3218	39.9	0.825554	0.894	1151	220	38.85	1.686
2106	28.9	0.88217	0.648	344	161	48.64	1.194
3090	29.7	0.877792	0.755	680	175	45.41	1.36
935	31.9	0.865973	0.612	150	125	50.95	1.085
2124	41.9	0.816032	0.876	692	185	42.08	1.406
2789	34.4	0.852923	0.783	775	100	45.47	1.352
590	35.0	0.8498	1.278	181	220	45.37	1.238
2423	40	0.825073	0.765	713	169	42.10	1.399
1292	31	0.870769	0.973	412	130	48.29	1.238
2359	30.1	0.875619	0.801	521	160	47.34	1.274
3449	39.3	0.828454	0.769	899	195	40.65	1.503
2944	37.5	0.8373	0.841	1008	230	38.65	1.65
1698	40	0.825073	0.964	646	193	42.58	1.408
1920	35.6	0.8468	0.838	523	250	41.35	1.422
3571	32.7	0.861754	0.802	898	175	43.21	1.471
3796	36.8	0.8408	0.849	2266	255	32.46	2.422
2148	33.3	0.858617	0.815	585	120	46.70	1.286
1890	38.1	0.834316	0.802	580	100	46.38	1.259
1207	29.7	0.878	1.079	405	212	45.94	1.322
1437	28.2	0.886036	1.002	389	150	49.42	1.226
1377	28.4	0.884928	0.921	331	160	49.06	1.21
1058	32.3	0.863858	0.79	220	127	49.80	1.13
1990	30.1	0.875619	0.801	521	85	49.36	1.222
1193	36.5	0.842262	0.96	469	130	47.10	1.246
3200	39.6	0.827	0.91	1246	250	36.19	1.852
545	27.5	0.889937	1.072	141	155	51.19	1.125
1700	36.6	0.841761	1.028	364	206	46.77	1.232
3063	31.2	0.869699	0.737	577	180	46.16	1.301
642	37.3	0.83827	1.192	266	165	46.41	1.22
994	30.6	0.8729	1.16	343	230	46.03	1.301

4640	36.2	0.8438	0.827	1784	231	35.39	2.055
1562	38.9	0.830399	1.281	463	196	47.49	1.261
1530	35	0.84985	1.228	355	209	47.55	1.24
2016	36.2	0.843769	1.013	803	160	43.88	1.452
2845	39.4	0.82797	0.951	1143	240	39.51	1.682
1741	48.4	0.786548	0.759	563	217	38.96	1.409
2768	36.8	0.8407	0.942	1016	218	38.84	1.686
2168	37.1	0.839265	0.789	544	164	44.88	1.297
3155	32.2	0.864386	0.774	700	185	44.30	1.384
2632	49.3	0.782633	0.73	888	228	36.54	1.578
2401	34.5	0.85241	0.782	567	175	44.93	1.318
1765	34	0.854985	0.695	345	151	47.81	1.184
1658	41.4	0.818392	0.865	368	186	45.71	1.212
1780	37.8	0.835794	0.853	509	205	42.63	1.362
2941	36.5	0.842262	0.812	811	160	43.29	1.421
1065	34.2	0.8541	1.061	392	213	45.17	1.305
2530	33.2	0.859138	0.774	693	100	45.15	1.349
2230	38.1	0.834316	0.802	580	175	44.37	1.316
1625	33.5	0.8576	1.047	631	244	41.97	1.489
2865	32.6	0.862279	0.752	742	100	46.27	1.327
3440	37.4	0.837774	0.764	863	192	42.09	1.455
901	30.1	0.8756	1.12	242	235	47.04	1.24
1360	39.6	0.827	1.116	587	275	39.73	1.523
2482	37.2	0.8388	1.061	948	229	40.78	1.619
3057	36.5	0.842262	0.812	811	185	42.57	1.445
3387	41.4	0.818392	0.673	919	194	39.52	1.505
2509	36.8	0.8408	0.865	963	220	40.58	1.572
2249	28.8	0.88272	0.824	469	165	47.44	1.272
1490	29.4	0.8794	0.989	537	239	43.61	1.424
1220	31.4	0.868631	0.884	267	174	48.95	1.173
2901	32.2	0.864386	0.774	700	140	45.34	1.352

2. Other Empirical Correlations for Bubble point Pressure

Standing:

Pb =18.2*((E3/D3)^0.83*10^((F3*0.00091)-(B3*0.0125))-1.4)

Al Marhoun :

Pb =5.38088*10^-3*(E3^0.715082)*(D3^ -1.87784)*(C3^3.1437)*((F3+460)^1.32657)

Al Shammasi:

Pb =(C3^5.527215)* (EXP(-1.841408*ABS(C3*D3)))*((E3*(460+F3)*D3)^0.783716)

Dokla and Osman:

Pb =0.836386*10^4*(E3^0.724047)*(D3^-1.01049)*(C3^0.107991)*((F3+460)^-0.952584)

3. Other Empirical correlations for Oil Formation Volume Factor

Standing:

Bo =0.972+(1.472*10^-4)*(((E3*(D3/C3)^0.5)+(1.25*F3))^1.175)

Al Marhoun

Bo =0.497069+((0.862963*10^-3)*(F3+460))+((0.182594*10^-2)*((E3^0.74239)*(D3^0.323294)*(C3^-1.20204)))+((0.318099*10^-5)*(((E3^0.74239)*(D3^0.323294)*(C3^-1.20204))^2))

Al Shammasi 3 parameters

Bo = 1+0.000412*(E3/C3) +0.00065*((F3-60)/C3)

Al Shammasi 4 parameters

Bo =1+(5.53*10^-7)*(E3*(F3-60))+(0.000181*(E3/C3))+(0.000449*(F3-60)/C3)+(0.000206*E3*D3/C3)