**Applying Machine learning to Develop Reliable Models for Calculating the Pressure-Volume-Temperature PVT Properties**

Musa Ahmed, Sathya Lahari Nuka, Hruthika Jinna, Charan Devapatla, Sampath Kumar, Mohammed Ibrahim, University of Houston

**Abstract**

Understanding PVT properties enables engineers and scientists to make informed decisions regarding reservoir management, production optimization, and hydrocarbon recovery. Ideally PVT properties are obtained using laboratory measurements for sample obtain at typical downhole conditions. However, these lab measurements are costly and time consuming; therefore, several correlations were built to predict the PVT properties. These correlations still show deviation from the actual values. In this paper we developed more reliable models to predict these PVT properties using Machine learning and Artificial intelligence.

The new proposed models predict the oil formation volume factor (Bo), bubble point pressure (Pb) and solution gas oil ratio (Rs). The predicted value is calculated based on primary inputs of pressure, temperature, gravity, and solution gas oil ratio. Our data sets include 505 data points generated from 25 PVT reports from 6 oil fields were used to build and test models’ performance. The accuracy of the model was then determined using Average absolute error (AAPE) and coefficient of determination (R-values). We built robust ensemble models by combining the output of Lasso Regression, KNN, Random Forest, ANN, SVM, and Gradient Boosting models. The hyperparameters of each individual model were tuned using grid search with cross-validation, and the best parameters were selected for each model.

Compared to existing empirical correlations the new models are simple and easy to use with good accuracy. The newly developed machine learning models outperformed the existing famous PVT correlations and artificial intelligence models in predicting the oil formation volume factor with average absolute error of 2.28% . In addition to their simplicity the new models shows almost same accuracy of famous PVT correlations (Standing ,Vasques-Beggs, Almarhoun) when predicting the bubble point pressure and solution gas oil ratio. Therefore, using the newly developed models can help in generating fast and cheap PVT properties that have good accuracy and avoid the high cost of lab measurements. Using the new ensemble models to predict accurate PVT properties will enables engineers and scientists to make informed decisions regarding reservoir management, production optimization, and hydrocarbon recovery.

**Key words:** PVT properties, phase behavior, machine learning, PVT models

**Introduction:**

PVT properties are very essential input parameters that required for several application in petroleum industry such as reservoir simulation, material balance calculation, flow assurance and reserve estimation (Ahmed, 2006; Al-Mehaideb, 1997; Elsharkawy and Alikhan, 1997). For oil field PVT properties include bubble point pressure, oil formation factor, solution gas oil ratio, water formation volume factor and oil viscosity(Hassan and Karim, 2016). In general PVT properties is measured in laboratory using sample that obtained from the field and reserved at typical downhole conditions. The fluid samples are then subjected to a series of PVT experiments in a controlled laboratory environment. These experiments are designed to measure the fluid's behavior under various pressure, volume, and temperature conditions. Common PVT experiments include constant composition expansion (CCE), differential liberation, flash liberation, and multi-stage separation tests(Ahmed, 2006).

Several PVT correlations were developed to predict the PVT properties when laboratory measurement is not available. (Standing, 1947) was the first one who developed correlations to estimate PVT properties. His correlation was based on 105 experimentally measured data points from 22 hydrocarbon systems from California oil fields. His correlation showed an average error of less than 5 % when applied to the California oil fields. After standing several authors presented correlations to predict the PVT properties at different geographical locations. For instance, Elam 1957 developed new models for Texas then followed by (Glaso 1980) who have done his work based on data for North Sea. After that several correlations were developed at different geographical locations in the world such as Libya, Nigeria, Angola, Canada, Saudi Arabia, UAE, and Iraq. Some authors such as (Vazquez and Beggs, 1980) developed a generalized correlations that can be applied worldwide. Most of developed statistical correlations were based on nonlinear regression using tool such as IPM SPSS statistical software.

Recently AI and machine learning became popular in oil and gas field applications(Hassan et al., 2020; Ramirez et al., 2017; Sircar et al., 2021; Sola-Aremu, 2019; Uzogor and Akinsete, 2020). (Alimadadi et al., 2011) designed ANN model to predict oil formation volume factor and density for Iranian oil fields. Their model showed a good accuracy compared to PVT correlation however they only predict density and oil formation volume factor with still deviation from lab measurents. (Hassan et al., 2020) also have used ANN to predict the PVT properties based on 250 data sets. His work proves that using ANN provide better estimation of PVT properties compared to old statistical correlations. Despite the huge amount of literure in the topic however, we belive there is still room to improve the PVT properties perdition using the latest machine learning techniques and this work is proposed to fill in the gab.

Despite most of the work that has been done in literatures the model used for PVT properties estimation has some deviation from the actual value measured in the lab. This work has been carried out to provide more reliable models that can predict the PVT properties with better accuracy.

The objective of this paper is developing more reliable models using machine learning and artificial intelligence technique to predict three essential PVT properties. The new models predict the oil formation volume factor, bubble point pressure and solution gas oil ratio at different conditions using the available data at surface. The accurate PVT data are very essential to create accurate reservoir models, design production facilities, and predict equipment requirements.

**Data Acquisition and Data Analysis**

The data sets were collected from 25 PVT report from 6 oil fields 3 fields from Sudan and 3 field from south sudan. We discovered that the dataset contains no duplicate or missing values. Each parameter's distribution looks to be normal which is desired when we use 'distplot' .However, when we use a box plot to visualize, there are some outliers. Using the interquartile range, we filtered the data frame to remove values outside of the lower and upper boundaries for each column. P and Pb have the highest correlation with RsActual, as well as Rsb has highest correaltion with Pb, according to scatterplot, pairplot, and correlation matrix visualizations. Bo has a weak correlation with all of the variables. Finally, we have a dataset free of outliers, duplicate values, and null values.

We standardized the dataset using Min Max Scaler after splitting the data. Our models were created using Lasso Regression, KNN, Random Forest, ANN, SVM, and Gradient Boosting. The hyperparameters of each individual model were tuned using grid search with cross-validation, and the best parameters were selected for each model. An ensemble model was built using all the above algorithms. Ensembling models is a technique used in machine learning to improve the accuracy and stability of predictions by combining the outputs of multiple models. In the context of determining the Pressure-Volume-Temperature (PVT) properties for methodology methods, ensembling can be used to create a more robust and accurate model for predicting these properties.

The workflow involved training individual models using each algorithm, tuning the hyperparameters using cross-validation, and then combining the outputs of all the models to create a final ensemble model.

The models were trained on a training dataset consisting of 70% of the available data. The hyperparameters of each individual model were tuned using grid search, randomized search with cross-validation, and the best parameters were selected for each model.

The final ensemble model was created by combining the outputs of the individual models using a weighted average. The individual models were tested on a testing dataset consisting of the remaining 20% of the available data. The performance of each individual model was evaluated using Average Absolute Percentage error and R-value metrics.

**Formation Volume Factor (*Bo*):**

The dataset contains six parameters related to oil properties: Oil Formation Volume Factor (Bo), Temperature (T), Solution Gas-Oil Ratio (Rs), Gas Gravity, Density, and API Gravity. The dataset includes 224 data points, with the parameters having a range of values, as well as different means, standard deviations, and coefficients of variation. The Bo values range from 0.94 to 1.77 RB/STB, with a mean of 1.11 RB/STB, while the temperature values range from 125.6°F to 244°F, with a mean of 174.27°F. The Rs values have a high coefficient of variation of 153.3%, indicating significant variation within the dataset. The Gas Gravity values range from 0.59 to 1.53, with a mean of 0.94, and the Density values range from 0.81 to 0.94 g/cm³, with a mean of 0.87 g/cm³. The API Gravity values range from 18.87 to 43.93, with a mean of 31.22. The skewness values show a slight deviation from normality, with positive skewness observed for Bo, Rs, and Gas Gravity. The kurtosis values indicate heavy-tailed distributions for Bo and Rs, while the other parameters are almost normally distributed. Figure (1) show the impact of inputs parameters on formation volume factor the solution gas oil ratio has the highest impact on the preicted oil formation volume factor and the gas gravity was the least effective inputs.

**Table (1)** **—Ranges of the parameters used for training and testing the Ensemble model for (*Bo*)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | ***Bo*** | ***T*** | ***Rs*** | **Gas gravity** | **Density** | **API** |
| count | 224 | 224 | 224 | 224 | 224 | 224 |
| mean | 1.110 | 174.269 | 87.536 | 0.939 | 0.870 | 31.218 |
| std | 0.169 | 26.954 | 134.191 | 0.268 | 0.036 | 6.715 |
| min | 0.937 | 125.600 | 0.341 | 0.592 | 0.807 | 18.872 |
| 25% | 1.016 | 157.730 | 16.632 | 0.696 | 0.836 | 27.280 |
| 50% | 1.052 | 171.900 | 38.765 | 0.916 | 0.882 | 28.550 |
| 75% | 1.104 | 194.000 | 83.467 | 1.103 | 0.891 | 37.703 |
| max | 1.768 | 244.000 | 770.170 | 1.530 | 0.941 | 43.928 |
| kurtosis | 5.412 | -0.082 | 8.243 | -0.812 | -1.016 | -1.070 |
| skewness | 2.352 | 0.422 | 2.804 | 0.573 | 0.085 | 0.101 |
| Coefficient of variance | 15.205 | 15.467 | 153.299 | 28.540 | 4.123 | 21.510 |

Chart, waterfall chart

Description automatically generated

**Figure (1) —The feature importance of input parameters on the formation volume factor (Bo):**

**Bubble Point Pressure (*Pb*):**

This data set contains 59 observations of six variables for Pb (Bubble Point). The variables include the Pb formation pressure (Pb), temperature (T), stock-tank oil gravity (Rsb), gas gravity, density, and API gravity. The mean values for Pb, T, Rsb, gas gravity, density, and API are 427.59 psi, 170.75°F, 57.70 scf/STB, 1.00, 0.88 g/cc, and 30.08, respectively. The standard deviations for these variables are 294.82 psi, 28.63°F, 40.19 scf/STB, 0.27, 0.04 g/cc, and 7.30, respectively.

The minimum and maximum values for the variables range from 51.00 psi to 1302.00 psi for Pb, 125.60°F to 244.00°F for temperature, 0.60 scf/STB to 146.32 scf/STB for Rsb, 0.59 to 1.53 for gas gravity, 0.81 g/cc to 0.94 g/cc for density, and 18.87 to 43.93 for API. The kurtosis values indicate that the distributions for Pb, T, and Rsb are approximately normal, while the distributions for gas gravity, density, and API are more flattened compared to a normal distribution. The skewness values suggest that the distributions for these variables are positively skewed, except for density which has a slight negative skew. The coefficient of variance values suggest that the variability of Pb and Rsb is very high, while the variability for density is relatively low. Figure (2) show the impact of inputs parameters on bubble point pressure. Also the solution gas oil ratio at bubble point had the highest impact on the predicted value and the gas gravity was the least effective inputs.

**Table (2)—Ranges of the parameters used for training and testing the Ensemble model for (*Pb*)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | ***Pb* (Actual)** | ***T*** | ***Rs*** | **Gas gravity** | **Density** | **API** |
| count | 59 | 59 | 59 | 59 | 59 | 59 |
| mean | 427.593924 | 170.754746 | 57.697708 | 1.000906 | 0.877475 | 30.080163 |
| std | 294.818123 | 28.633997 | 40.187187 | 0.270360 | 0.039399 | 7.300361 |
| min | 51.000000 | 125.600000 | 1.200000 | 0.591500 | 0.806600 | 18.871945 |
| 25% | 199.000000 | 143.000000 | 25.120000 | 0.776500 | 0.837600 | 24.328673 |
| 50% | 328.000000 | 171.900000 | 51.300000 | 0.969000 | 0.884300 | 28.513570 |
| 75% | 651.385254 | 193.950000 | 88.613500 | 1.210000 | 0.908050 | 37.435055 |
| max | 1302.000000 | 244.000000 | 146.320000 | 1.530000 | 0.941000 | 43.927721 |
| kurtosis | 0.135666 | -0.365983 | -0.913022 | -1.068399 | -1.233452 | -1.216017 |
| skewness | 0.880873 | 0.287824 | 0.448877 | 0.289839 | -0.091240 | 0.194930 |
| Coefficent of variance | 68.948155 | 16.769078 | 69.651271 | 27.011524 | 4.490045 | 24.269687 |

Chart, waterfall chart

Description automatically generated

**Figure (1) —The feature importance of input parameters on the bubble point pressure (Pb).**

**Solution Gas–Oil Ratio (*Rs*):**

This dataset contains 222 observations of various properties related to natural ga (RsActual), temperature (T), density, API gravity, gas gravity, bubble point pressure (Pb), and pressure (P). The mean value of Rs(lab) is 50.18 scf/STB, while the mean temperature is 174.26 degrees Fahrenheit. The density ranges from 0.8066 to 0.9410 g/cm³, with a mean of 0.8700 g/cm³. The API gravity ranges from 18.87 to 43.93, with a mean of 31.20. The gas gravity ranges from 0.5915 to 1.53, with a mean of 0.9368. The bubble point pressure (Pb) ranges from 51 psi to 1985 psi, with a mean of 645.63 psi. The pressure (P) ranges from 40 psi to 1315 psi, with a mean of 363.01 psi. The dataset shows a relatively high coefficient of variance for bubble point pressure, gas gravity, and pressure. The skewness and kurtosis values indicate that the data are approximately normally distributed, with a few exceptions. Overall, the dataset provides useful information for studying the physical properties of natural gas reservoirs. Figure (3) show the impact of inputs parameters on solution gas oil ratio. Unlike the two previous PVT properties the gas gravity and density had the highest impact on the predicted value while the API and temperature were having the lowest feature importance.

**Table (3)—Ranges of the parameters used for training and testing the Ensemble model for (*Rs*)**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Rs(actual)** | **T** | **Density** | **API** | **Gas Gravity** | **Pb(psi)** | **P (psi)** |
| count | 222 | 222 | 222 | 222 | 222 | 222 | 222 |
| mean | 50.1768 | 174.258 | 0.870045 | 31.203 | 0.936758 | 645.628 | 363.009 |
| std | 40.196 | 27.0174 | 0.035949 | 6.72915 | 0.267258 | 447.654 | 286.548 |
| min | 0.795808 | 125.6 | 0.8066 | 18.8719 | 0.5915 | 51 | 40 |
| 25% | 17.0265 | 158 | 0.836025 | 27.2409 | 0.69125 | 278 | 118.5 |
| 50% | 38.8945 | 171.9 | 0.8816 | 28.5498 | 0.916073 | 576.5 | 270.5 |
| 75% | 84.6676 | 194 | 0.8912 | 37.7533 | 1.103 | 964.279 | 538.4 |
| max | 175.66 | 244 | 0.941 | 43.9277 | 1.53 | 1985 | 1315 |
| kurtosis | 0.019079 | -0.084334 | -1.019682 | -1.071858 | -0.788735 | 1.309515 | 0.92425 |
| skewness | 0.869237 | 0.42361 | 0.081689 | 0.104198 | 0.5865 | 1.07529 | 1.169993 |
| Coefficient of variance | 80.1089 | 15.5043 | 4.131911 | 21.56569 | 28.530087 | 69.33629 | 78.93701 |

Chart, bar chart

Description automatically generated

**Figure (2) —The feature importance of input parameters on the Solution Gas–Oil Ratio (Rs)**

**Results and Discussion**

**Formation Volume Factor:**

Figure (4) shows the actual oil formation volume factor versus the predict value using the new ensemble model. The new model input parameters are temperature, solution gas oil ratio, density and API gravity. The new ensemble model was able to predict the oil formation volume factor with an average absolute error of 2.3%. The new model can be used to obtain the oil formation volume factor fortemperature range of125.6°F to 244°F.

**Chart, scatter chart

Description automatically generated**

**Figure (4) —Cross plot of actual against predicted solution oil formation volume factor using Ensemble model, for the testing data set.**

**Bubble Point Pressure (*Pb*):**

Figure (5) displays the actual Bubble Point Pressure (Pb) against the predicted value using the new ensemble model. The new model input parameters for (Pb) are temperature, solution gas oil ratio, density, gas gravity and API gravity. The new ensemble model was able to predict the *Pb* with an average absolute error of 26%.

Chart, scatter chart

Description automatically generated

**Figure (5) —Cross plot of actual against predicted solution gas-oil ratio using Ensemble model, for the testing data set.**

**Solution Gas–Oil Ratio:**

Figure (6) displays the actual Solution Gas–Oil Ratio (*Rs*) against the predicted value using the new ensemble model. The new ensample model outperforms popular model such as Standing (1947) Vasquez-Beggs (1977) and Almarhoun (1988), The new model input parameters for (Rs) are temperature, , density, gas gravity , pressure , bubble point pressure and API gravity. The new ensemble model was able to predict the *Pb* with an average absolute error of 26%.

Chart, scatter chart

Description automatically generated

**Figure (6) —Cross plot of actual against predicted solution gas-oil ratio using Ensemble model, for the testing data set.**

**Comparison Analysis**

The results obtained in this study was compared with famous PVT correlations as well as recently developed AI and machine learning models. The new models outperformed both famous correlations and the machine learning and AI models in predicting the oil formation volume factor (Bo). The average absolute error for the new model was 2.28% which was the lowest among the famous correlations and machine learning and AI Models (Al-Marhoun, 1988; Hassan et al., 2020; Mohamed et al., 2018; Standing, 1947; Vazquez and Beggs, 1980). The popular PVT correlation showed almost similar accuracy with the new ensample models for predicting the bubble point pressure and solution gas oil ratios. While the AI models that developed by Amjed et al 2019 and Mohamed et al 2018 surpassed our newly developed models as well as other correlations in predicting the least two properties.

**Table (4)** **2—The average percentage errors for determining the PVT properties using different approaches**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Determination Approach | | | Formation Volume Factor | Bubble Point Pressure | Solution Gas Oil Ratio |
| 1 | Famous Correlations | Standing (1947) | 6.9% | 23.1% | 24.9% |
| Vazquez-Beggs (1977) | 7.4% | 20.0% | 31.9% |
| Almarhoun (1988) | 6.6% | 25% | 38.1% |
| 2 | SPSS (Nonlinear Regression) | Mohamed et al (2018) | 5.06% | 9.1% | 19.54% |
| 3 | ANN | Amjed et al (2019) | 2.59% | 11.39% | 10.31% |
| 4 | Ensemble Model | This study (2023) | 2.28% | 26.8% | 25.4% |

**Conclusions and Recommendations**

In this study we used machine learning approach to develop new model that can predicted the essential three PVT properties of oil i.e. oil formation volume factor, bubble point pressure and solution gas oil ratio. PVT data were collected from 6 oil fields from south sudan and sudan. By merging the results of the Lasso Regression, KNN, Random Forest, ANN, SVM, and Gradient Boosting models, we were able to create reliable ensemble models. Grid search with cross-validation was used to tweak each model's hyperparameters, and the best values were chosen for each model. The developed new model surpassed the most popular PVT correlations as well as the AI and machine learning models in predicting the oil formation volume factor with average absolute error of only 2.3 %. Furthermore, the ensemble model has predicted the bubble point pressure and solution gas oil ratio with almost same accuracy of the famous PVT equations. The newly developed model are easy to use and has good accuracy to predict the PVT properties when we do not have laboratory results. The finding of the study are important to reservoir engineers, flow assurance engineers, production engineers and geologist for production optimization and resrve estimations.

**References**

Ahmed, T., 2006. Reservoir Engineering Handbook Third Edition, Proceedings - SPE Annual Technical Conference and Exhibition.

Al-Marhoun, M.A., 1988. PVT CORRELATIONS FOR MIDDLE EAST CRUDE OILS. JPT, J. Pet. Technol. 40, 650–666. https://doi.org/10.2118/13718-PA

Al-Mehaideb, R.A., 1997. Improved PVT correlations for UAE offshore crudes. Sekiyu Gakkaishi (Journal Japan Pet. Institute). https://doi.org/10.1627/jpi1958.40.232

Alimadadi, F., Fakhri, A., Farooghi, D., Sadati, S.H., 2011. Using a committee machine with artificial neural networks to predict PVT properties of Iran crude oil. SPE Reserv. Eval. Eng. 14, 129–137. https://doi.org/10.2118/141165-PA

Elsharkawy, A.M., Alikhan, A.A., 1997. Correlations for predicting solution gas/oil ratio, oil formation volume factor, and undersaturated oil compressibility. J. Pet. Sci. Eng. 17, 291–302. https://doi.org/10.1016/S0920-4105(96)00075-7

Hassan, A., Abdulraheem, A., Mohamed, A., 2020. Reliable models for determining the pressure-volume-temperature PVT properties using artificial intelligence technique. Int. Pet. Technol. Conf. 2020, IPTC 2020. https://doi.org/10.2523/iptc-19884-abstract

Hassan, I.G., Karim, M.N., 2016. EFFECT OF PORE SIZE DISTRIBUTION ON MULTIPHASE EQUILI BRIUM OF FLUIDS CONFINED IN POROUS MEDIA.

Mohamed, A., Ahmed, M., Mohamed, I., Abdein, M., 2018. The performance of the popular PVT correlations on Sudanese crude oils: A statistical study. Soc. Pet. Eng. - SPE Kingdom Saudi Arab. Annu. Tech. Symp. Exhib. 2018, SATS 2018. https://doi.org/10.2118/192438-ms

Ramirez, A.M., Valle, G.A., Romero, F., Jaimes, M., 2017. Prediction of PVT Properties in Crude Oil Using Machine Learning Techniques MLT. SPE Lat. Am. Caribb. Pet. Eng. Conf. Proc. 0. https://doi.org/10.2118/185536-MS

Sircar, A., Yadav, K., Rayavarapu, K., Bist, N., Oza, H., 2021. Application of machine learning and artificial intelligence in oil and gas industry. Pet. Res. 6, 379–391. https://doi.org/10.1016/J.PTLRS.2021.05.009

Sola-Aremu, O., 2019. An inferable machine learning approach to predicting PVT properties of Niger delta crude oil using compositional data. Proc. - SPE Annu. Tech. Conf. Exhib. 2019-September. https://doi.org/10.2118/199783-STU

Standing, M.B., 1947. A Pressure-Volume-Temperature Correlation For Mixtures Of California Oils And Gases. Drill. Prod. Pract.

Uzogor, K., Akinsete, O., 2020. Improved correlations and predictive models for nigerian crude oil pvt properties using advanced regression and intelligent techniques. Soc. Pet. Eng. - SPE Niger. Annu. Int. Conf. Exhib. 2020, NAIC 2020. https://doi.org/10.2118/203658-MS

Vazquez, M., Beggs, H.D., 1980. Correlations for Fluid Physical Property Prediction. Softw. - Pract. Exp. 1977-October. https://doi.org/10.2118/6719-MS