Support Vector Machine Model for Predicting Gas Saturated and Undersaturated Crude Oil Viscosity of Niger Delta Oil Reservoir

Isemin Isemin[1,2], Akinsete Oluwatoyin[1*] and Akpabio Julius[2]

[1]*Department of Petroleum Engineering, University of Ibadan, Ibadan, Nigeria.*
[2]*Chemical and Petroleum Engineering Department, University of Uyo, Uyo, Nigeria.*

Authors' contributions

This work was carried out in collaboration among all authors. Author AO designed the study, performed the statistical analysis, wrote the protocol, and wrote the first draft of the manuscript. Authors AO and AJ managed the analyses of the study. Author AO managed the literature searches. All authors read and approved the final manuscript.

Original Research Article

ABSTRACT

Oil viscosity is one of the most important physical and thermodynamic property used when considering reservoir simulation, production forecasting and enhanced oil recovery. Traditional experimental procedure is expensive and time consuming while correlations are replete however they are limited in precision, hence need for a new Machine Learning (ML) models to accurately quantify oil viscosity of Niger Delta crude oil.

This work presents use of ML model to predict gas-saturated and undersaturated oil viscosities. The ML used is the Support Vector Machine (SVM), it is applicable for linear and non-linear problems, the algorithm creates a hyperplane that separates data into two classes. The model was developed using data sets collected from the Niger Delta oil field. The data set was used to train, cross-validate, and test the models for reliability and accuracy. Correlation of Coefficient, Average Absolute Relative Error (AARE) and Root Mean Square Error (RMSE) were used to evaluate the developed model and compared with other correlations.

Result indicated that SVM model outperformed other empirical models revealing the accuracy and advantage SVM a ML technique over expensive empirical correlations.

Corresponding author: Email: oo.akinsete@mail1.ui.edu.ng, oo.akinsete@ui.edu.ng;

ABBREVIATIONS

AARE : *Average Absolute Relative Error*
ANN : *Artificial Neural Network*
EOS : *Equations of State*
ML : *Machine Learning*
SVM : *Support Vector Machine*
PVT : *Pressure-Volume-Temperature*
R : *Correlation Coefficient*
RMSE : *Root Mean Square*

1. INTRODUCTION

With the advancement of technology today and the revolution of data, machine learning (ML) has become a fast, accurate and effective means of predictions. Machine learning is referred to as the scientific study of algorithms and statistical models which computer systems deploy to perform specific task that depends on patterns and interference without introducing any explicit instructions. Machine Learning (ML) is based on algorithms and gives better performance when supplied with enough data; the more data feed into the machine the more accurate the predictions will be hence it is a means of building mathematical models to understand data. The algorithm types are different in their individual approach, in the data type they input and output, and also in the type of task or problem they intended to solve. They combine many features of the data together to produce a model. Before Machine Learning techniques were deployed in predicting crude oil properties, laboratory measurements, empirical correlations, and Equation of States (EOS) have been used extensively.

Pressure-Volume-Temperature (PVT) analysis is the process of determining and predicting fluid behaviors and properties of oil and gas samples of an existing well and it is an integral part in understanding flow of hydrocarbon fluids from the well. The Laboratory measurements have been known to be the most reliable measurement of PVT analysis of reservoir fluid properties although the measurements are often not available or too expensive to obtain. The empirical correlation is used for predicting fluid properties where there is no enough experimental information. This paper presents the use of support vector machine learning model to predict saturated viscosity and undersaturated viscosity. The crude oil viscosity is an important physical property that describes fluid's resistance to flow through porous media

and pipes. Above bubble point pressure, the viscosity of the oil in the reservoir decreases as pressure decreases. At lower pressure the molecules are number further apart and therefore move past each other easily. A typical viscosity diagram as a function of pressure at constant reservoir temperature is shown in Fig. 1.

1.1 The Niger Delta Crude Oil

There are two types of crude oil in the Niger Delta basin: light and comparatively heavy. The Bonny Light oil produced here in Nigeria is a light-sweet crude oil grade with a good quality of 34.5° API with a low Sulfur content of 0.14%, and are volatile with gas-oil-ratio (GORs) ranging from 180 to 1600 ft^3/bbl. The heavier crude oil has API of 20° – 25° gravity also having a low Sulfur content. The Bonny Light crude oil is a very high grade due to the very low content sulfur resulting to low corrosiveness to refinery infrastructure and the lower environmental impacts of its byproducts in refinery effluent. Other than the Bonny Light, the Nigerian crude oil also comprises of the Qua Iboe crude oil, Brass River crude oil, Pennington Anfan and Forcados crude oil.

There are three factors used to categorize the crude oil in the oil and gas industry, they include viscosity, volatility, and toxicity. While viscosity refers to the internal resistance of fluid to flow, a higher viscosity will not allow flow of fluid easily thereby requiring more energy to pump or produce from the ground. The volatility accounts for the rapid oil evaporation into the air and high volatile crude will require additional processes to control environments during extraction. Toxicity describes the poisonous and harmful state the oil will be to the environment and humans during extraction and refining processes.

There are several regional and global empirical PVT correlations that have been developed for fluid properties to determine reservoir performance, estimate reserves, make real-time decision etc. this developed empirical correlations are based on data from different parts of the world and are more accurate for use with crudes from the same basins for which the data was correlated [2].

In 2008, Ikiensikimama and Ogboja [3] stated that the accuracy of empirical correlations depends on the region from which the crudes

were obtained. The data used in the correlations are from specific geographical areas also that the paraffinicity, which affects properties of crude oil, differs from region to region. They also showed in their work that correlations developed from various regions were concerned with crudes of different characteristics and they would not provide best approximation of PVT properties elsewhere. The following authors Glaso [4], Al-Marhoun [5], Labedi [6], Uzogor and Akinsete [7] were in support of the foregoing assertion. Also no one particular crude from the regions match the crude from the Niger Delta. From the foregoing, using any of the correlations from these regions would have implications such as poor fluid property estimations, poor reservoir performance studies, as well as uncertainty in reserve estimations [3,7]. The universal correlations are less accurate than regional correlations. Most widely used correlation is the ones by Vasquez and Beggs [8], other early works to predict reservoir fluid properties are Glaso [4], Al-Marhoun [5], Standing [9], Petrosky and Farshad [10], Hanafy et al. [11].

Existing PVT studies for gas saturated and undersaturated viscosity are as follows: in 1975, Beggs and Robinson [12] developed an empirical correlation for determining the viscosity of dead oil by analyzing 460 dead oil viscosity measurements and the data set from which the results were obtained ranges from 16°API to 58°API and 70°F to 295°F. From their correlation, it tends to overstate the viscosity of the crude oil in ranges from 100°F to 150°F.

$$\mu_{od} = 10^x - 1 \qquad\qquad 1$$

In 1987, Khan et al. [13] gave a correlation that was developed for Saudi Arabian oils for determination of viscosity at, above, and below the bubble point pressure. The result of their correlation gives the most accurate predictions for Saudi Arabian crude oils as when compared to the Beggs and Robinson [12], Beal [14] and Chew and Connaly [15] correlations. The oil gravity for the correlation must be less than 1 (10°API).

$$\mu_{os} = \mu_{ob} \left(\frac{p}{p_b}\right)^{-0.14} e\left(-2.5 * 10^{-4}(p - p_b)\right) \qquad 2$$

Egbogah and Ng [16] in 1990 modified Beggs and Robinsons [12] viscosity correlation using the pour point temperatures. The purpose of introducing the pour point temperature into the correlation is to reflect the chemical composition of crude oil into the viscosity correlation. To obtain the viscosity of the live oils, the dead oil correlations are used with the Beggs and Robinson [12] viscosity correlation. The data used to derive the correlation was taken from the reservoir fluids analysis lab using a total of 394 oil systems. Correlation for saturated live oil is:

$$\mu_{os} = \mu_{ob} \left(\frac{p}{p_b}\right)^x \qquad\qquad 3$$

In 1997, Hanafy et al. [11] used a total of 324 fluid samples taken from 123 reservoirs in 75 fields for the PVT measurements for estimating bubble point pressure, solution gas oil ratio, oil formation volume factor, oil compressibility, oil viscosity, and oil density for the Egyptian oils. While in 2013, El-Hoshoudy et al [17] developed a new correlation for the prediction of density and viscosity of Egyptian oil system containing both dead and live crude.
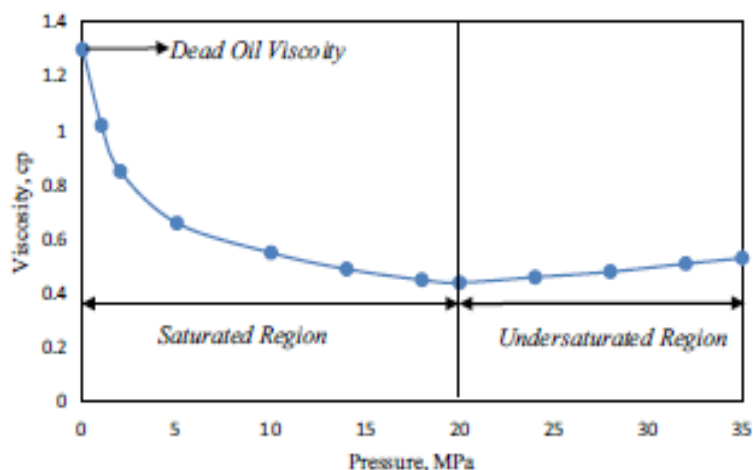

Fig. 1. Viscosity trend as function of Pressure [1]

In the work of Chew and Connally [15], where they presented oil viscosity at the bubble point as a function of the solution gas–oil ratio, they used 457 data points which covered samples from South America, Canada, and the U.S. Abu-Khamsin and Al-Marhoun [18] in their work developed a correlation based on Canadian and Middle Eastern oil data, their correlation gave an average absolute error and standard deviation of 4.91% and 5.76, respectively. Also the authors Kartoatmodjo and Schmidt [19] developed oil viscosity correlation based on data bank consisting of 5321 data points while De Ghetto et al. [20] developed the saturated oil viscosity correlation based on data bank ranging from 0.07 to 295.9 cp. Almehaideb [2] in his published work on oil viscosity at bubble point, developed a correlation by using 57 PVT data points he collected from 15 different reservoirs in UAE. The average absolute error and standard deviation of his correlation gave 13.0% and 16.26%, respectively.

Artificial Neural Network (ANN) is the most common machine learning technique used in recent times to predict PVT properties, other models do exist such as the Support Vector Machine although not commonly used yet but few works where SVM has been utilized has shown better predictions and performances. Some of the PVT work carried out using ANN are credited to the following authors: Ikiensikimama and Ogboja [3], Uzogor and Akinsete [7], Gharbi et al. [21], Osman [22] and Alakbari [23].

1.2 Theoretical Concept of Support Vector Machine (SVM)

SVM is a supervised machine learning algorithm used mainly for classification purposes. It is also used for regression, outliner detection, and clustering. It is regarded as one of the most popular classifier, otherwise known as a large margin classifier. As a supervised ML algorithm, SVM trains label data, studies the data, then classify input data. SVM works by drawing a decision boundary (also known as a hyperplane) to separate between two classes with the optimum hyperplane having a maximum distance from each of the support vectors. The largest margin between any two classes is the optima or best line. The support vectors are the data points that are close to the hyperplanes, they influence positioning of the hyperplane. The margin is the distance between the hyperplane and the support vectors.

Machine Learning (ML) enables analysis of massive quantities of data, and this contributes in making it to deliver a faster, more accurate result for ease of predictions. Algorithms built using machine learning techniques are based on sample data referred to as training data which in turn make predictions or decisions in the task it performed having been explicitly programmed. Some of the types of machine learning models are Artificial Neural Network, Decision trees, Support vector machines, Regression analysis, Bayesian networks, and Genetics algorithm etc. This work presents use of ML model to predict saturated and undersaturated oil viscosities. The ML model used is the Support Vector Machine (SVM).

1.3 SVM Kernels

SVM efficiently perform nonlinear classifications using the kernel tricks, as it transforms a non-linear space to a linear space in other words it transforms a low-dimensional input space into a higher-dimensional input space.

2. METHODOLOGY

2.1 Model Description

The SVM model was built using Python, an interpreted, object-oriented, high level programming language with dynamic semantics, and it is used for machine learning via the machine learning libraries and its framework such as the Scikit-learn, Panda, and NumPy. While the Scikit learn is a powerful python library for machine learning and predictive modeling, the Pandas is a high level Python package which helps in providing a fast and expressible data structures, and the Numpy provides high performance multidimensional arrays processing in Python.

2.2 Data Processing

The data was first divided into attributes and labels, then finally divided into training set (70%) and testing set (30%). The train data is for model building and fitting, while the test data is for prediction and evaluation. With the model selection library of the scikit-learn which contains the train_test_split, the script was executed.

2.3 Input Data

The dataset used comprises of 450 data points from existing literatures for the Niger Delta

region. The datasets are divided into train set and test set. 70% of the dataset was used to train the model for which the algorithm learns the pattern and the other 30% of the data set was used to test the model for prediction.

Table 1. Saturated and undersaturated viscosity dataset

	Undersaturated Viscosity	Saturated Viscosity
Data points	450	400
Training set	315	280
Testing set	135	120

2.4 Training the Algorithm

After splitting the data, the model was built by first importing the SVM module and creating the support vector classifier (SVC) using the kernel function argument. The model was fit on the training data set before performing predicting on the testing data set. Model evaluation was done to know the accuracy of the model prediction. This was done by comparing the actual values and predicted values. The linear kernel function was used to transform the given data set into the required form. A regularization was done to represent the error term which indicates for the SVM optimization the amount of error that is allowed. However, SVMs are limited in scalability, and are challenged in selecting the best parameters of C and gamma to run the model (Elite Data Science, 2019) [24].

3. RESULTS AND DISCUSSION

To describe the accuracy of the developed model, statistical error analyses used are Correlation Coefficient (CC), Average Absolute Relative Error (AARE), and Root Mean Square Error (RMSE).

The gas saturated oil viscosity and undersaturated oil viscosity models were both evaluated using the train dataset and predictions were made using the test dataset. As stated earlier, a total number of 450 data points and 400 data points were used for the training and testing data respectively. The performance of machine learning SVM algorithm for undersaturated and saturated oil viscosities were given in Table 2. These statistical parameters were used for comparison with existing empirical models (Table 3).

Table 3 showed the SVM performance considering other existing models. The developed model gave a Correlation Coefficient of 96% which showed a good performance when compared to other empirical models. The correlation coefficient represents the degree of success in reducing the standard deviation, however there is only a marginal reduced error in terms of RMSE (0.42), AARE (5.6) for the developed model compared to other models. The RMSE measures the data dispersion around zero deviation.

Table 2. SVM Model Performance

	Training Dataset	Testing Dataset	Training Dataset	Testing Dataset
	Undersaturated Oil Viscosity		Saturated Oil Viscosity	
Correlation Coefficient	0.97	0.96	0.96	0.95
RMSE	0.45	0.42	0.47	0.41
AARE	5.70	5.60	5.50	5.40

Table 3. Comparison of developed SVM model with other empirical and models

	Undersaturated Oil Viscosity			Saturated Oil Viscosity		
Models	AARE	CC	RMSE	AARE	CC	RMSE
Vazques and Beggs	4.70	0.85	0.45	4.45	0.82	0.44
Standing	4.80	0.86	0.41	4.60	0.87	0.42
Khahn	4.55	0.50	0.47	4.50	0.55	0.47
Petrosky and Farshad	3.50	0.77	0.41	3.51	0.72	0.42
Labedi	3.52	0.71	0.31	3.30	0.79	0.28
Almehideb	5.20	0.80	0.35	5.20	0.80	0.37
Present SVM Model	5.60	0.96	0.42	5.40	0.95	0.41

4. CONCLUSION

Support Vector Machine (SVM) has shown tremendous performance for model development in predicting reservoir fluid properties compared to existing PVT correlations and other machine learning algorithms such as the artificial neural network (ANN), however the usage of SVM is not yet popular. To determine accurate PVT properties, data from specific region have to be used when developing models.

The present SVM viscosity model compares well with other empirical models, hence viscosity determined by ML can be included as an integral part of all reservoir simulators, PVT simulators and fluid properties prediction packages for reservoir studies.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Alqahtani NB, AlQuraishi AA, Al-Baadani W. New correlations for prediction of saturated and undersaturated oil viscosity of Arabian oil fields. J Petrol Explor Prod Technol. 2018;8:205–215.
2. Almehaideb RA. Improved PVT correlations for UAE crude oils. Paper SPE 37691. Presented at the SPE Middle East Oil Conference and Exhibition, Bahrain; 1997.
3. Ikiensikimama S, Ogboja O. Review of PVT Correlations for Crude Oils. Global Journal of Pure and Applied Sciences. 2008;14.
    DOI: 10.4314/gjpas.v14i3.16816.
4. Glasø, Ø. Generalized Pressure-Volume-Temperature Correlations. Journal of Petroleum Technology. 2008;32(5):785–795. SPE-8016-PA.
    DOI: 10.2118/8016-PA.
5. Al-Marhoun MA. PVT Correlation for Middle East Crude Oils. Journal of Petroleum Technology. 1988;650–665.
6. Labedi R. Use of Production Data to Estimate Volume Factor Density and Compressibility of Reservoir Fluids; 1990.
7. Uzogor K, Akinsete O. Improved Correlations and Predictive Models for Nigerian Crude Oil PVT Properties Using Advanced Regression and Intelligent Techniques. Paper presented at the Nigeria Annual International Conference and Exhibition (NAICE); 2020.
8. Vazquez M, Beggs HD. Correlations for fluid physical property prediction. Journal of Petroleum Technology. 1980;32(06):968–970.
9. Standing MB. Oil-system correlations. Petroleum Production Handbook systems; 1947.
10. Petrosky GE, Farshad F. Pressure-Volume-Temperature Correlations for Gulf of Mexico Crude Oils. SPE paper 26644, presented at the 68th Annual Technical Conference of Petroleum Engineers, Houston, October 3–6; 1993.
11. Hanafy HH, Macary SM, ElNady YM, Bayomi AA, El Batanony MH. A new approach for predicting the crude oil properties. In SPE Production Operations Symposium. Society of Petroleum Engineers; 1997.
12. Beggs HD, Robinson JR. Estimating the viscosity of crude oil systems. Journal of Petroleum Technology. 1975;27(09):1140–1149.
13. Khan SA, Al-Marhoun MA, Duffuaa SO, Abu-Khamsin SA. Viscosity correlations for Saudi Arabian crude oils. In Middle East Oil Show. Society of Petroleum Engineers; 1987.
14. Beal C. Viscosity of air, water, natural gas, crude oil and its associated gases at oil field temperature and pressures. Trans he AIME. 1946;165(114):127.
15. Chew J, Connally CA. Viscosity correlation for gas saturated crude oil. Trans AIME. 1959; 216:23–25.
16. Egbogah EO, Ng JT. An improved temperature-viscosity correlation for crude oil Journal of Petroleum Science and Engineering. 1990;4(3):197–200.
17. El-Hoshoudy AN, Farag AB, Ali OIM, El-Batanoney MH, Desouky SEM, Ramzi M. New correlations for prediction of viscosity and density of Egyptian oil reservoirs. Fuel. 2013;112: 277-282.
18. Abu-Khamsim A, Al-Marhoun M. Development of a new correlation for bubble-point viscosity. Arab J Sci. Eng. 1991;16(2A):99.
19. Kartoatmodjo F, Schmidt Z. Large data bank improves crude physical property correlation. Oil Gas J. 1994;4:51–55.
20. De Ghetto G, Paone F, Villa M. Reliability analysis on PVT correlations. Paper SPE 28904. Presented at the European Petroleum Conference London; 1994.

21. Gharbi RB, Elsharkawy AM, Karkoub M. Universal neural-network-based model for estimating the PVT properties of crude oil systems. Energy & Fuels. 1999;13(2):454–458.

22. Osman EA, Abdel-Wahhab OA, Al-Marhoun MA. Prediction of Oil PVT Properties Using Neural Networks. Paper presented at the SPE Middle East Oil Show, Manama, Bahrain; 2001.

23. Alakbari FS, Elkatatny S, Baarimah SO. Prediction of Bubble Point Pressure Using Artificial Intelligence AI Techniques. Paper presented at SPE Middle East Artificial Lift Conference and Exhibition; 2016.

24. Elite Data Science Modern Machine Learning Algorithms: Strengths and Weaknesses; 2019.
Retrieved from Elite Data Science:
Available:https://elitedatascience.com/machine-learning-algorithms#classification

---

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sdiarticle4.com/review-history/70264