International Journal of Frontiers in
Engineering and Technology Research
Journal homepage: https://frontiersrj.com/journals/ijfetr/
ISSN: 2783-0497 (Online)

(RESEARCH ARTICLE)

Check for updates

Machine learning prediction of bottomhole flowing pressure as a time series in the volve field

Olugbenga Olamigoke * and David Chinweuba Onyeali

Department of Chemical and Petroleum Engineering, University of Lagos, Lagos, Nigeria.

International Journal of Frontiers in Engineering and Technology Research, 2022, 02(02), 020–029

Publication history: Received on 28 April 2022; revised on 01 June 2022; accepted on 03 June 2022

Article DOI: https://doi.org/10.53294/ijfetr.2022.2.2.0039

Abstract

Bottomhole flowing pressure (BHFP) is a critical parameter in analyzing oil and gas well performance, production forecasting and reservoir management. This study is focused on obtaining feature combinations towards low-error prediction of time-series BHFP in two wells in the Volve field. Three machine learning (ML) models (support vector regression (SVR), a distance-based model; random forest (RF), a tree-based ensemble model and Long Short-Term Memory (LSTM), a type of recurrent neural network) are used for BHFP prediction in two wells of the Volve field. The data for each well was split such that the first 70% is used in training the model, the next 15% as validation data for selecting the optimal hyperparameters and the last 15% for testing the models. The train and validation sets were used to train the models before making predictions on the test sets. While the SVR and RF models reasonably predicted the BHFP in both wells with a maximum Mean Absolute Percentage Error (MAPE) of 5.0% and 4.3% respectively, the LSTM model performed best across both wells with the MAPE less than 2.9% in both wells. ML model performance was superior for the well with the data distributed more uniformly. The three feature combinations with superior ML model performance for BHFP prediction all have five features in common namely: bottomhole temperature, oil flow rate, gas flow rate, choke size, onstream hours. The workflow in this work can be adopted for fieldwide BHFP prediction.

Keywords: Bottomhole flowing pressure; Well performance; Support vector regression; Random Forest; Long Short-Term Memory; Volve Field

1. Introduction

Large amounts of data are generated on the oil field daily, to which various machine learning techniques have been applied. Ensemble learning algorithms such as Random Forest (RF), Gradient Boosting (GB), Adaptive Boosting, and K-Neighbors were used to predict the time-lapse oil saturation profile at four well locations using data from the Volve field with synthetically generated time-lapse oil saturation data from a history-matched reservoir simulation model. It was found that the RF model performs better than the other models with the K-Neighbors model being the worst performing model [1]. The application of RF was proposed for testing the performance of two classification methods in detecting anomalies due to flow assurance, integrity, and mechanical problems to mitigate against loss of oil and gas production based on data obtained from bottomhole pressure and temperature sensors in wells of an offshore field [2]. Non-linear Support Vector Regression (SVR) was shown to be superior to segmented linear and polynomial regression in describing the time-series well production data and computing the deviation of production test parameters with time. This was critical to removing outliers from a database. The deviations obtained were then stochastically modelled to assess the uncertainty in production forecasts in a Brazilian oilfield [3]. It has been suggested that predictions obtained from machine learning models and deep learning models can be used to infer reservoir behavior as a substitute to flowing or shut-in well tests [4].

* Corresponding author: Olugbenga Olamigoke
Department of Chemical and Petroleum Engineering, University of Lagos, Lagos, Nigeria.

One parameter critical to the history matching, production forecasting or even well and reservoir management is bottomhole flowing pressure (BHFP). Its accurate prediction impacts on production optimization: attaining the balance between high technical hydrocarbon recovery while minimizing the unit cost per barrel [5]. Traditionally, the estimation of BHFP is done using published empirical correlations [6-11] and mechanistic models [12-17] which are applied taking into cognizance the flow regimes and wellbore inclination.

Machine learning models, and deep learning models, have shown potential in estimating BHFP over the last few years. Tariq et al. [5] used a feed-forward artificial neutral network (ANN) model with three layers to relate BHFP to ten variables such as production rates for the three fluid phases, wellhead pressure, wellhead temperature, bottom-hole temperature, tubing diameter, depth, and oil gravity. Published bottom-hole survey data taken in vertical wells producing without the aid of artificial lift was sourced from Govier and Forgarasi [18] and Asheim [19]. The weights and biases of the model were tuned and updated using a particle swarm optimization method. The developed model outperformed established empirical and mechanistic models for estimating BHFP thus providing a cost-effective alternative to running bottom-hole surveys. Akinsete and Adesiji [20] built an ANN model for predicting BHFP in the Volve field. Data was obtained from 7 wells with 70% used for training and 30% used to test the ANN model. The features used for training the ANN model are like those employed in Tariq et al. [5] differing in the exclusion of gas volume, water volume and oil gravity. Additional features used are annulus and differential pressure. The ANN model performance was compared to other machine learning models – SVR, RF and Decision Trees (DT). The ANN model had the least mean absolute error (MAE) of 0.02657 while SVR was the worst performing model with MAE of 7.22810. The MAE for RF and DT was similar yielding 0.52584 and 0.60121 respectively. Details of how ensemble models were trained was not discussed. Four machine learning algorithms (RF, GB, SVR and ANN) were compared by Kanin et al. [21] in predicting the pressure gradient along a pipe segment in inclined wells with two-phase gas-liquid flow. A set of dimensionless variables were used as inputs to the regression models to ensure similarity for different fluids and reduce the number of inputs. In addition to BHFP estimation, the liquid holdup and flow pattern were predicted as regression and classification tasks respectively. This necessitated building three models per algorithm. Of the four algorithms GB performed best with high accuracy while the worst performing model was ANN. The performance of the machine learning models was comparable to the empirical models tested. Baryshnikov et al. [22] improved the predictive capacity of an ANN model for BHFP estimation proposed by Kanin et al. [21] by training it on both synthetic data generated using mechanistic model and laboratory multiphase flow data and then tuning same to actual field data. The wellbore was divided into segments in line with the well survey data and the pressure drop along the wellbore calculated via an iterative procedure which computes fluid properties using correlations assuming steady-state. The ANN model performed slightly better than the empirical models to which it was compared. The model was then applied to time-dependent datasets obtained from both surface and bottom-hole pressure gauges during a flowback period. The train to test data split was 60% to 40% and 70% to 30% for well#1 and well#2 respectively. The accuracy of the model was limited by the approximation of quasi-steady-state behavior to a transient flow.

BHFP prediction has been approached as a time-series forecasting problem in some studies. Ignatov et al. [23] proposed tree-based models (RF and GB) for BHFP prediction, in multiphase wellbore flow which captured transient behavior. The training and testing datasets were generated by a flow simulator capable of handling transient multiphase flow. Time-dependent BHFP is considered a function of wellhead pressure, surface production rates and well geometry (trajectory and diameter). Rates were generated as exponential functions of time. The dataset was split for training and testing by 80% to 20% respectively. The normalized root mean squared error for the worst scenarios was GB, ANN and RF was 15.3%, 17.0% and 17.3%. Li et al. [4] explored the use of Recurrent Neural Network (RNN) models to analyze well behavior and predict BHFP in the Volve field as a time series; Simple RNN, Gated Recurrent Unit (GRU), LSTM and Long- and Short-term Time-series Network (LSTNet) which is an innovative combination of RNN and Convolutional Neural Network. The production data from well NO 15/9-F-1 C from April 2014 to April 2016 was used for both training and testing the models. The data was split into training, testing and validation sets in the ratio 70:15:15. Relative squared error (RSE) was used as the evaluation metric for the models, with relative absolute error (RAE) used as a reference. Single parameters were used to train a single layer GRU model to predict the BHFP, the best parameter (i.e., the one that gave the least RSE) was found to be temperature, followed by the choke opening percentage and then water production. Hybrid features formed from different combinations of rate and time difference were tested, their performance showed that the model was able to extract the relationship between flow rate and pressure more effectively. The combination of features with the least errors includes the following: bottomhole temperature, time interval, production rate for each phase and the rates over each time interval. The GRU and LSTM models were found to be more accurate than the simple RNN model with RSE for the test dataset yielding 0.1394, 0.2008 and 0.5739 respectively. The LSTNet showed significant improvement in performance with the least RSE of 0.0884. Abdrakhmanov et al. [24] used transformer-based neural networks for the forecasting of bottomhole pressure in the Volve field. Learning from training the transformer-based model to predict BHFP based production parameters (daily production rates for oil, gas, and water, bottomhole temperature, and percentage choke opening) from one well was transferred to the training procedure of a second well by tuning of the weights of the initial model to obtain the bottomhole pressure predictions for the second well. Generalization of the transformer-based model to handle well interference in multiple well flow test was carried out which resulted in more accurate downhole pressure prediction. While the performance of the transformer model was comparable to the RNN, LSTM and GRU models on the initial dataset, it outperformed the other models on the second well with RSE of 0.12 as compared to 0.15, 0.46 and 0.62 for LSTM, GRU and RNN respectively.

In this study, optimal feature selection for the prediction of time dependent BHFP is addressed. The performance of machine learning algorithms (RF and SVR) is compared to a sequential deep learning model – LSTM. Recommendations to BHFP prediction as a time series are provided.

2. Methodology

2.1. Description of Volve Field Datasets

The Volve field was developed with pressure support provided by two water injection wells, F-4 and F-5, both drilled in 2007. Production started with two production wells, F-12 and F-14, both drilled in 2008. The field was expected to produce for 5 years. Wells 15/9-F-11, 15/9-F-15 D and 15/9-F-1 C were drilled after 2013. The daily production data from the Volve field used for this work, specifically for wells 15/9-F-15 D and 15/9-F-1 C, are shown in Figure 1 and Figure 2 respectively. Cyclical shut-in and flow periods can be seen in both figures which could be because of maintenance operations on the platform [25].

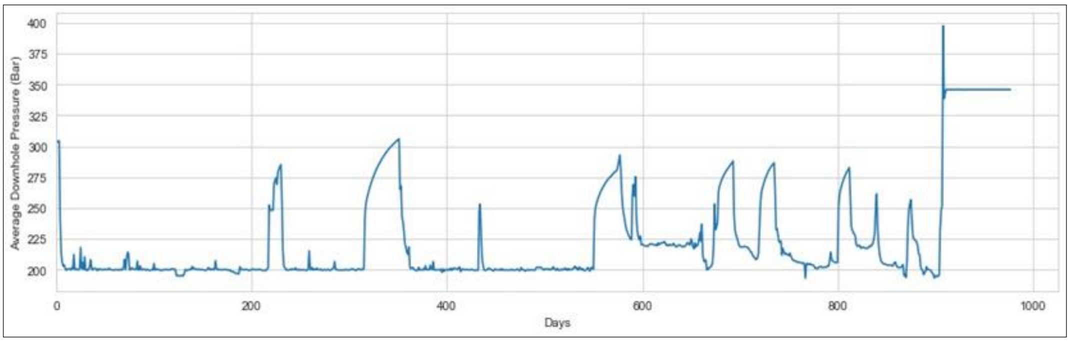

Figure 1 Average bottomhole pressure in Well 15/9-F-15 D

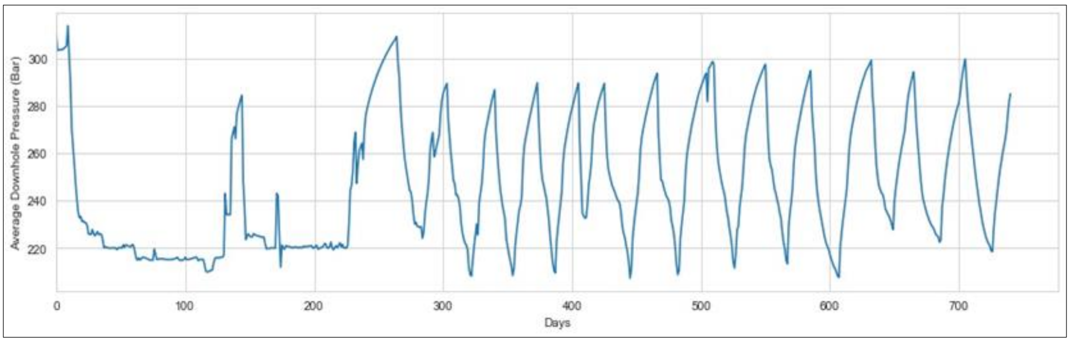

Figure 2 Average bottomhole pressure in Well 15/9-F-1 C

The selected features, which are all numerical, from the data for the selected wells as extracted from the full dataset are summarized in Table 1. There were no missing values in well 15/9-F-15 D. However, it was observed that the average bottomhole pressure data was found to be abnormally high from 2016-07-10 (after 900 days). These data points were considered outliers and were consequently removed. Data from the Well 15/9-F-1 C is well distributed with no outliers, and slightly skewed to the left. Missing values were removed from well 15/9-F-1 C data. The correlation of the features with the bottomhole pressure for both wells is shown in Table 2. Three models are employed in this work - Support Vector Machine (SVM), Randon Forrest and Long Short-Term Memory (LSTM) - to predict the bottomhole pressure for wells in the Volve field.

Table 1 Features definition, units, and data types

Feature	Definition	Unit
ONS_HR	Number of hours of production	Hours
AVG_BHP	Bottomhole pressure	Bar
AVG_BHT	Bottomhole temperature	^{o}C
*ΔP_TUB	Differential tubing pressure	Bar
%CHK_OP	Percentage of choke open	%
AVG_WHP	Wellhead pressure	Bar
AVG_WHT	Wellhead temperature	^{o}C
CHK_SZ	Choke size	mm
OIL_VOL	Average oil flow rate	Sm^3/day
GAS_VOL	Average gas flow rate	Sm^3/day
WAT_VOL	Average water flow rate	Sm^3/day

* Differential tubing pressure, is a derived term (ΔP_TUB = AVG_BHP – AVG_WHP)

Table 2 Correlation of features with bottomhole pressure

Feature	15/9-F-15 D	15/9-F-1 C
AVG_BHT	-0.927	-0.795
ONS_HR	-0.883	-0.694
OIL_VOL	-0.682	-0.599
GAS_VOL	-0.680	-0.611
AVG_WHT	-0.615	-0.675
%CHK_OP	-0.537	-0.675
WAT_VOL	-0.088	-0.324
ΔP_TUB	0.594	0.471
AVG_WHP	0.703	0.355
CHK_SZ	0.818	0.566

3. Machine Learning Models

3.1. Support Vector Machine

Support Vector Machine, as applied to regression, works by trying to minimize the error margin, $\pm\epsilon$, between the predicted value and the actual value for each data point. It uses a kernel for nonlinear regression by projecting the nonlinear data to a space where it can be represented linearly [26]. The kernels include “linear”, “sigmoid”, “radial basis function”, and “polynomial”. For well 15/9-F-15 D, the data was normalized such that each feature is between 0 and 1 (min-max normalization), while for well 15/9-F-1 C, the data was standardized such that data has a mean of 0 and standard deviation of 1.

3.2. Random Forest

Random Forest is an ensemble model; it is made up of decision trees. A decision tree is a structure that is built up by a rule based hierarchical structure made up of nodes or leaves. A condition is imposed on a feature or features at each node. The prediction is obtained by averaging the values in the terminal nodes. A RF model works by training multiple decision trees in parallel and uses a bagging technique to obtain a robust model. For both wells min-max normalization was employed on each feature.

3.3. Long Short-Term Memory

Long Short-Term Memory (LSTM) is a sophisticated type of RNN, a type of deep neural network designed for use with sequential data. LSTM models are used to overcome problems associated with simple RNNs, such as the vanishing gradient problem. The LSTM model retains relevant information using three gates: the input gate which evaluates the importance of the input data and carries it to the next cell, the forget gate which is used to determine whether information should be discarded or retained and the output gate that controls the content for the next hidden state. A network with a single LSTM layer and a linear output layer was used in training. The LSTM layer had 512 units and a sequence length of 5 was used, i.e., the previous 5 values of each of the three features are used in predicting the next value of the target. For both well 15/9-F-15 D and well 15/9-F-1 C, the data was normalized using the robust scaling method. The target value was also scaled, and the result inverted to obtain the unscaled bottomhole pressure.

3.4. Optimization of model hyperparameters

The dataset for each well was split such that the first 70% is used in training, the next 15% as validation data for selecting the optimal hyperparameters and the last 15% for testing the models, for wells 15/9-F-15 D and 15/9-F-1 C respectively. The validation results were obtained by training on the train data and using the resulting model to predict the validation set. The test results were obtained by training the model on the train and validation sets before predicting the test set. An optimization algorithm [27] is used to select optimal hyperparameters for the SVM and RF models for each well. For the LSTM model, Keras tuner [28] was used to select optimal hyperparameters using Adam optimizer which combines the advantages of Adaptive Gradient Algorithm and Root Mean Square Propagation with a learning rate of 1e-4. The performance of each model in this work has been evaluated using six metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Coefficient of Determination (R^2 Score), Relative Squared Error (RSE) and Relative Absolute Error (RAE).

4. Results and discussion

Table 3 Performance (RMSE) of single features in predicting BHFP

	15/9-F-15 D			**15/9-F-1 C**			**Average**		
Feature/Method	**SVR**	**RF**	**LSTM**	**SVR**	**RF**	**LSTM**	**SVR**	**RF**	**LSTM**
AVG_BHT	12.126	11.614	11.704	19.853	16.131	12.160	15.990	13.872	11.932
OIL_VOL	12.603	12.923	12.557	15.644	19.649	11.314	14.123	16.286	11.936
GAS_VOL	12.694	13.753	12.307	15.677	20.055	11.663	14.185	16.904	11.985
%CHK_OP	12.758	15.688	12.621	23.995	25.725	21.167	18.377	20.706	16.894
ONS_HR	13.308	13.032	12.323	24.190	21.462	15.177	18.749	17.247	13.750
CHK_SZ	15.438	16.328	14.677	24.949	26.487	20.729	20.193	21.407	17.703
AVG_WHT	16.782	16.106	14.401	19.868	25.029	15.248	18.325	20.568	14.824
AVG_WHP	19.966	43.827	16.692	22.157	25.220	13.337	21.061	34.524	15.015
ΔP_TUB	22.615	36.025	24.854	28.787	25.356	25.504	25.701	30.690	25.179
WAT_VOL	27.022	18.588	20.237	20.704	22.673	19.257	23.863	20.631	19.747

The performance of the three machine learning models for BHFP prediction based on each single feature independently was tested, with the RMSE shown in Table 3 for both wells. Though the models perform differently, the three features which best estimate BHFP across both wells are bottomhole temperature, oil production rate, and gas production rate. However, the wellhead pressure, which is a key parameter in bottomhole pressure prediction in both empirical and mechanistic models, performs poorly in the time-series FBHP prediction independently. To minimize the error in prediction, different combinations of features were tested with the three models for BFHP prediction. These feature combinations include at least one of the prior identified parameters: bottomhole temperature, oil production rate, and gas production rate. Table 4 shows the results of the feature combinations in predicting the test set with RMSE as the performance metric.

Table 4 RMSE for models based on different feature combinations on prediction of the test set

	15/9-F-15 D			**15/9-F-1 C**		
Features	**SVR**	**RF**	**LSTM**	**SVR**	**RF**	**LSTM**
AVG_BHT, CHK_SZ	10.027	11.677	11.622	16.300	16.162	10.984
AVG_BHT, OIL_VOL	10.027	10.942	11.480	12.295	14.571	8.767
AVG_BHT, OIL_VOL, %CHK_OP	10.130	9.706	12.862	9.749	14.026	7.266
AVG_BHT, OIL_VOL, CHK_SZ	10.066	11.143	11.269	13.687	13.606	7.322
AVG_BHT, OIL_VOL, WAT_VOL	14.672	9.584	9.367	9.992	13.782	9.516
AVG_BHT, OIL_VOL, %CHK_OP, AVG_WHP	9.922	9.985	11.965	11.501	12.927	8.492
AVG_BHT, OIL_VOL, GAS_VOL, %CHK_OP	10.317	9.923	12.751	10.005	11.504	7.108
AVG_BHT, OIL_VOL, GAS_VOL, CHK_SZ	9.868	11.293	10.474	12.544	14.078	7.352
AVG_BHT, OIL_VOL, GAS_VOL, CHK_SZ, AVG_WHT	11.862	10.381	11.129	13.169	13.710	6.925
AVG_BHT, OIL_VOL, GAS_VOL, CHK_SZ, ONS_HR	**9.854**	**11.259**	**8.919**	**9.451**	**12.917**	**6.471**
AVG_BHT, OIL_VOL, GAS_VOL, CHK_SZ, AVG_WHP, ONS_HR	**11.906**	**9.870**	**8.263**	**9.827**	**12.171**	**6.567**
AVG_BHT, OIL_VOL, GAS_VOL, CHK_SZ, AVG_WHP, AVG_WHT, ONS_HR	**11.663**	**9.919**	**7.855**	**10.188**	**12.568**	**6.679**
CHK_SZ, ΔP_TUB, AVG_WHT	**6.620**	**8.423**	**10.724**	**5.132**	**7.965**	**4.961**
GAS_VOL, AVG_WHT, ONS_HR	14.828	11.035	11.548	18.023	21.602	13.607
OIL_VOL, CHK_SZ	13.283	13.465	10.410	13.735	19.297	12.859
OIL_VOL, AVG_WHP	14.727	11.051	14.783	16.192	20.038	14.401
OIL_VOL, AVG_WHP, AVG_WHT	17.676	17.506	15.159	16.030	16.594	14.096
OIL_VOL, AVG_WHP, AVG_WHT, CHK_SZ, WAT_VOL	10.440	11.118	8.695	11.543	15.254	12.911
OIL_VOL, AVG_WHP, AVG_WHT, %CHK_OP, WAT_VOL	10.824	10.454	9.394	15.587	17.150	14.135

Models using feature combinations which exclude bottomhole temperature tend perform poorly in BFHP prediction as shown in Table 4. An exception is seen with the combination of three features: *choke size, tubing differential pressure* and *wellhead temperature*. This feature combination resulted in superior performance for both the SVR and RF models over other feature combinations for both wells 15/9-F-15 D and 15/9-F-1 C. The SVR model had the best performance in predicting FBHP for the test data for well 15/9-F-15 D with highest coefficient of determination of 0.914 and the least error metrics as regards MSE and RSE with 5.895 and 0.00864 respectively. With respect to RMSE, the SVR model error was 21% and 38% lower than the RF and the LSTM error respectively; the RAE associated with the SVR model of 0.3416 was also the least compared to 0.3984 and 0.4916 associated with the RF and LSTM models respectively. The LSTM had the best performance on the well 15/9-F-1 C test data with RMSE of 4.961 as compared to 5.132 and 7.965 for SVR and RF respectively and MAE of 3.842 to 4.344 and 6.723 for SVR and RF respectively. Overall, the feature combination of choke size, tubing differential pressure and wellhead temperature was the most concise and accurate in predicting BFHP. It also reflects three elements in tandem with bottomhole pressure estimation – choke size for rate, pressure drop along the tubing and wellhead temperature. However, there is a drawback of data leakage in this feature combination as the tubing differential pressure is the difference between the predicted BFHP and wellhead pressure. Thus, alternative feature combinations which include bottomhole temperature must be considered. Of these feature combinations three combinations stand out with low RMSE values (as highlighted in Table 4). The performance of the SVR, RF and LSTM models based on these highlighted feature combinations (C#1, C#2, and C#3) is shown in Table 5 with their evaluation metrics.

Table 5 Results for the three best performing feature combinations with bottomhole temperature

	15/9-F-15 D			15/9-F-1 C		
	C#1: AVG_BHT, OIL_VOL, GAS_VOL, CHK_SZ, ONS_HR					
	SVR	RF	LSTM	SVR	RF	LSTM
RMSE	9.854	11.259	8.919	9.451	12.917	6.471
MAE	8.215	8.209	6.458	7.226	10.717	5.178
MAPE	3.780	3.750	2.910	2.880	4.280	2.010
R^2	0.809	0.750	0.843	0.833	0.688	0.922
RSE	0.191	0.250	0.157	0.167	0.313	0.079
RAE	0.476	0.486	0.374	0.361	0.536	0.259
C#2: AVG_BHT, OIL_VOL, GAS_VOL, CHK_SZ, ONS_HR, AVG_WHP						
	SVR	RF	LSTM	SVR	RF	LSTM
RMSE	11.906	9.871	8.263	9.827	12.171	6.567
MAE	10.722	6.847	5.827	8.020	9.794	5.197
MAPE	4.987	3.086	2.595	3.183	3.814	1.985
R^2	0.721	0.808	0.865	0.819	0.722	0.919
RSE	0.279	0.192	0.135	0.181	0.278	0.088
RAE	0.621	0.397	0.338	0.401	0.490	0.260
C#3: AVG_BHT, OIL_VOL, GAS_VOL, CHK_SZ, ONS_HR, AVG_WHP, AVG_WHT						
	SVR	RF	LSTM	SVR	RF	LSTM
RMSE	11.663	9.919	7.855	10.188	12.568	6.679
MAE	10.216	7.100	5.227	7.754	9.925	5.246
MAPE	4.726	3.228	2.329	3.016	3.871	2.003
R^2	0.732	0.806	0.878	0.806	0.704	0.916
RSE	0.268	0.194	0.122	0.194	0.296	0.084
RAE	0.592	0.411	0.303	0.388	0.496	0.262

The SVR model performs best across both wells for the combination with five features (C#1) with RSE of 0.191 and 0.167 for wells 15/9-F-15 D and 15/9-F-1 C respectively as shown in Table 5 as compared to the combinations with six-features (C#2) and seven-features (C#3). The relative square error for the SVR model is similar for the six and seven feature combinations. The RSE obtained with the six-feature combination (C#2) was 0.279 and 0.181 for wells 15/9-F-15 D and 15/9-F-1 C respectively while 0.268 and 0.194 was obtained when seven features (C#3) were employed for wells 15/9-F-15 D and 15/9-F-1 C respectively. The RF model generated the least error across both wells for the six-feature combination with the RSE of 0.192 and 0.278 for wells 15/9-F-15 D and 15/9-F-1 C respectively. The RSE is quite similar for the seven-feature combination with 0.194 and 0.296 for wells 15/9-F-15 D and 15/9-F-1 C respectively. The relative square error was highest when five features were used with 0.250 and 0.313 for wells 15/9-F-15 D and 15/9-F-1 C respectively.

It can be seen from Table 5 that the LSTM model best predicts the BHFP in both wells irrespective of the feature combination and regardless of the evaluation metric used with RAE ranging from 0.303 to 0.374 and 0.259 to 0.262 for wells 15/9-F-15 D and 15/9-F-1 C respectively. While the performance of the LSTM improves with increasing number of features for well 15/9-F-15 D the converse is the case for well 15/9-F-1 C. Judging based on performance, the seven-feature combination i.e., C#3 *(bottomhole temperature, oil flow rate, gas flow rate, choke size, onstream hours, wellhead pressure and wellhead temperature*) has been selected for the prediction of BHFP. This is because while the RAE increases marginally from 0.259 to 0.262 for well 15/9-F-1 C while it declines from 0.374 to 0.303 for well 15/9-F-15 D which results in 1.2% and 19.0% RAE decline in with the inclusion of wellhead pressure and temperature.

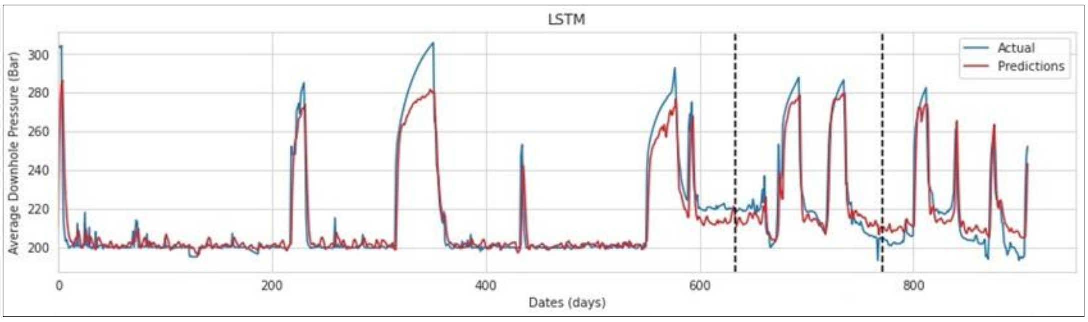

Figure 3 LSTM model performance- well 15/9-F-15 D.

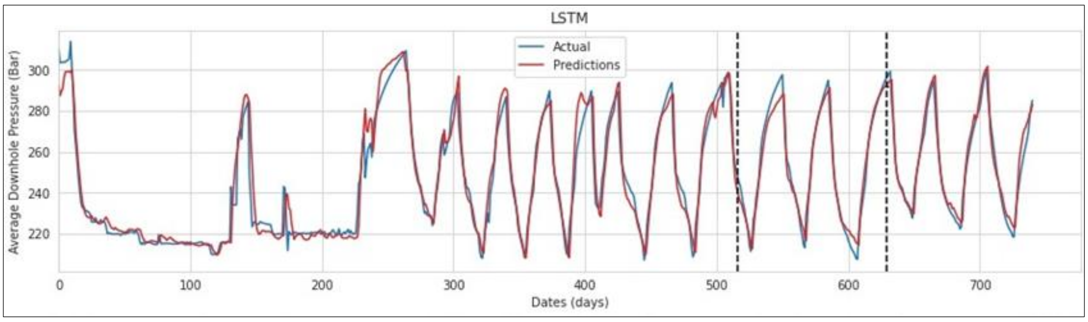

Figure 4 LSTM model performance- well 15/9-F-1 C

The performance metrics for the LSTM model are shown in Table 6. The actual to predicted time-series BHFP data is shown in Figure 3 and Figure 4 for wells 15/9-F-15 D and 15/9-F-1 C respectively. A comparison between the train results and the test results for the LSTM model for the seven-feature combination (Table 6) shows that the MAPE increases by 51% from 1.149% to 2.329% for the training data and test data based on well 15/9-F-15 D. In contrast, the MAPE declines 14% from 2.319% to 2.003% for well 15/9-F-1 C indicating the LSTM model performed better in predicting BHFP as compared to well 15/9-F-15 D. The three models (SVR, RF and LSTM) generally performed better on well 15/9-F-1 C than on well 15/9-F-15 D. This is attributed to the difference in pressure distributions for both wells. The bottomhole pressure data for well 15/9-F-1 C is more uniformly distributed compared to the well 15D pressure data.

The models were all able to extract meaningful information from the data. It is worth noting that similar accuracy in BHFP prediction in this study has been achieved with five parameters, six parameters and seven parameters, as compared to nine parameters in an earlier study [4]. A comparison of the error generated by models used in predicting BHFP in this work shows that the LSTM model in this study performed marginally better in BHFP prediction for well 15/9-F-1 C with RSE and RAE of 0.08 and 0.26 respectively as compared to RSE and RAE of 0.084 and 0.267 respectively obtained using the LSTNet – an innovative combination of Recurrent Neural Networks with Convolutional Neural Networks [4]. The LSTM model implemented in this work is also superior in performance for both wells to the Transformer model in Abdrakhmanov et al. [24] with a 22% and 28% decrease in RMSE and MAPE respectively. The application of machine learning methods to time-series bottom hole flowing pressure data over two different wells give further credence to the consistency of the results obtained.

Table 6 Model Performance metrics for train and test data using C#3 for both wells

	15/9-F-15 D						**15/9-F-1 C**					
	SVR		**RF**		**LSTM**		**SVR**		**RF**		**LSTM**	
	Train	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**
RMSE	9.18	11.7	1.69	9.92	6.35	7.86	7.39	10.2	3.61	12.6	7.26	6.68
MAE	7.57	10.2	0.77	7.10	2.68	5.23	4.46	7.75	2.32	9,93	5.64	5.25
MAPE	3.50	4.73	0.34	3.23	1.15	2.33	1.82	3.02	0.94	3.87	2.32	2.00
R^2	0.89	0.73	0.99	0.81	0.95	0.88	0.92	0.81	0.98	0.70	0.94	0.92
RSE	0.11	0.27	0.01	0.19	0.05	0.12	0.08	0.19	0.02	0.30	0.06	0.08
RAE	0.35	0.59	0.04	0.41	0.13	0.30	0.18	0.39	0.09	0.50	0.23	0.26

5. Conclusion

Three machine learning models (Support Vector Regression, Random Forest, and Long Short-Term Memory) have been employed for predicting flowing bottomhole pressure as time series data for two wells in the Volve field. Three different feature combinations have been identified with similar performance metrics for the three machine learning models. The common features to the three combinations are bottomhole temperature, oil flow rate, gas flow rate, choke size, onstream hours, wellhead pressure and wellhead temperature. Of the three models, the LSTM model is the best performing model across both well. The SVR model appears to produce favorable performance metrics for fewer number of features. This contrasts with the LSTM model which posts improved performance metrics with increasing number of features. The RF model tends to overfit training data thus resulting in poor performance on the test data as compared to LSTM models. Consistency in BHFP estimates obtained via machine learning models in two wells in the Volve field have been demonstrated. The performance of each model however depends on effective tuning of the model hyperparameters.

Compliance with ethical standards

Acknowledgments

We are profoundly grateful to Dr. (Mrs.) A.O. Ogunbayo and Mr. Joseph Olaide for their support towards the completion of this work.

Disclosure of conflict of interest

The authors state that there is no conflict of interest in this work. No funding grant was used in carrying out this work.

References

[1] Wang B, Sharma J, Chen J, Persaud P. Ensemble Machine Learning Assisted Reservoir Characterization using Field Production Data-An Offshore Field Case Study. Energies. 2021; 14(4): 1052.

[2] De Salvo Castro AO, De Jesus Rocha Santos M, Leta FR, Lima CBC, Lima GBA. Unsupervised Methods to Classify Real Data from Offshore Wells. American Journal of Operations Research. 2021; 11: 227-41.

[3] Monteiro DD, Duque MM, Chaves GS, Ferreira Filho VM, Baioco JS. Using data analytics to quantify the impact of production test uncertainty on oil flow rate forecast. IFP Energies Nouvelles. 2020; 75(7): 1-15.

[4] Li Y, Sun R, Horne R. Deep Learning for Well Data History Analysis. In: SPE Annual Technical Conference and Exhibition, Calgary, Alberta, Canada. 2019; 1-16.

[5] Tariq Z, Mahmoud M, Abdulraheem A. Real-time prognosis of flowing bottom-hole pressure in a vertical well for a multiphase flow using computational intelligence techniques. J Petrol Explor Prod Technol. 2020; 10: 1411-28.

[6] Duns H, Ros NCJ. Vertical flow of gas and liquid mixtures in wells. In: Proceedings of the 6th World Petroleum Congress, Frankfurt, Germany. Paper Number: WPC-10132. 1963; 451-65.

[7] Orkiszewski J. Predicting two-phase pressure drops in vertical pipe. J Pet Technol. 1967; 19(06): 829-38.

[8] Aziz K, Govier GW. Pressure drop in wells producing oil and gas. J Can Pet Technol. 1972; 11(03): 7-9.

[9] Beggs DH, Brill JP. A study of two-phase flow in inclined pipes. J Pet Technol. 1973; 25: 607-17.

[10] Chierici GL, Ciucci GM, Sclocchi G. Two-Phase Vertical Flow in Oil Wells – Prediction of Pressure Drop. J Pet Technol. 1974; 26(08): 927-38.

[11] Mukherjee H, Brill JP. Pressure Drop Correlations for Inclined Two-Phase Flow. J Energy Resour. Technol. 1985; 107(4): 549-54.

[12] Ansari AM, Sylvester ND, Sarica C, Shoham O, Brill JP. A comprehensive mechanistic model for upward two-phase flow in wellbores. SPE Prod & Fac. 1985; 9(02): 143-51.

[13] Hasan AR, Kabir CS. Fluid Flow and Heat Transfer in Wellbores. 2nd ed. Richardson: Society of Petroleum Engineers; 2018.

[14] Guo B, Ghalambor A. Natural Gas Engineering Handbook. 2nd ed. Houston: Gulf Publishing Company; 2012.

[15] Bendiksen KH, Maines D, Moe R, Nuland S. The Dynamic Two-Fluid Model OLGA: Theory and Application. SPE Prod Eng. 1991; 6(02): 171-80.

[16] Gomez LE, Shoham O, Schmidt Z, Chokshi RN, Northug T. Unified Mechanistic Model for Steady-State Two-Phase Flow: Horizontal to Vertical Upward Flow. SPE Journal. 2000; 5(3): 339-50.

[17] Zhang H-Q, Wang Q, Sarica C, Brill JP. A unified mechanistic model for slug liquid holdup and transition between slug and dispersed bubble flows. International Journal of Multiphase Flow. 2003; 29(1): 97-107.

[18] Govier GW, Forgarasi M. Pressure drop in wells producing gas and condensate. J Can Pet Technol. 1975; 14(04): 1-3.

[19] Asheim H. MONA, an accurate two-phase well flow model based on phase slippage. SPE Prod Eng. 1986; 1(03): 221-30.

[20] Akinsete O, Adesiji BA. Bottom-Hole Pressure Estimation from Wellhead Data Using Artificial Neural Network. In: SPE Nigeria Annual International Conference and Exhibition, Lagos, Nigeria. 2019; 1-15.

[21] Kanin EA, Osiptsov AA, Vainshtein AL, Burnaev EV. A predictive model for steady-state multiphase pipe flow: Machine learning on lab data. Journal of Petroleum Science and Engineering. 2019; 180: 727-46.

[22] Baryshnikov ES, Kanin EA, Osiptsov AA, Vainshtein AL, Burnaev EV, Paderin GV, Prutsakov AS, Ternovenko SO. Adaptation of Steady-State Well Flow Model on Field Data for Calculating the Flowing Bottomhole Pressure. In: Proceedings of the SPE Russian Petroleum Technology Conference. 2020; 1-17.

[23] Ignatov DI, Sinkov K, Spesivtsev P, Vrabie I, Zyuzin V. Tree-Based Ensembles for Predicting the Bottomhole Pressure of Oil and Gas Well Flows, In: van der Aalst WMP et al. (eds.). Analysis of Images, Social Networks and Texts. 1st ed. Springer Nature Switzerland. 2018; 221-33.

[24] Abdrakhmanov IR, Kanin EA, Boronin SA, Burnaev EV, Osiptsov AA. Development of Deep Transformer-Based Models for Long-Term Prediction of Transient Production of Oil Wells. In: Proceedings of the SPE Russian Petroleum Technology Conference. 2021; 1-15.

[25] Production performance of the Volve Oilfield [Internet]. WordPress. ©. 2021 [cited: 2022 Mar 28]. Available from: https://discovervolve.com/2021/04/14/production-performance-of-the-volve-oilfield/.

[26] Pandey YN, Rastogi A, Kainkaryam S, Bhattacharya S, Saputelli L. Machine Learning in the Oil and Gas Industry : Including Geosciences, Reservoir Engineering, and Production Engineering with Python, Apress Standard. 2020.

[27] Bergstra J, Yamins D, Cox D. Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures. In: Proceedings of the 30th International Conference on Machine Learning. PMLR. 2013; 28(1): 115-123.

[28] Hyperparameter tuning for humans [Internet]. GitHub, Inc. ©. 2022 [cited: 2022 Mar 30]. Available from: https://github.com/keras-team/keras-tuner.