RESEARCH ARTICLE OPEN ACCESS

A DATA-DRIVEN SURROGATE MODELLING FRAMEWORK FOR FIELD-WIDE PRODUCTION OPTIMIZATION IN MATURE OIL FIELDS

Umar Alhaji Mohammed [1], Abdullahi Usman [2], Ahmad Usman Ahmad[3]
[1](*Department of Computer Science, Usmanu Danfodiyo University, Sokoto, Nigeria.* ballah236@gmail.com)
[2](*Department Computer Science, Federal University Birnin Kebbi, Kebbi, Nigeria.* usman.gulumbe@fubk.edu.ng)
[3](*Department Computer Science, Heriot-watt University Edinburgh, Scotland.* aau4001@hw.ac.uk)

---------------------------*********************** ----------------------

Abstract:

Production optimization in mature oil fields is challenging due to decreasing reservoir performance and the limitations of conventional physics-based simulation tools, which are computationally expensive and inflexible for decision-making. This study proposes a data-driven surrogate modelling framework for field-wide production optimization using machine learning techniques. A multi-well dataset from the Volve oil field, comprising approximately 8,000 daily production records from five wells, was used. The data were pre-processed and enhanced through physics-informed feature engineering, including lagged and cumulative production variables to capture temporal dynamics and reservoir depletion effects. Several regression models were evaluated, including Linear Regression, Random Forest, XGBoost, and Gradient Boosting methods. The Histogram Gradient Boosting model achieved the best performance, with a coefficient of determination (R^2) of 0.9995 and low prediction errors. The trained surrogate model was then applied to evaluate multiple operating scenarios involving choke settings and pressure conditions. The results show that the proposed approach achieved a field-wide production improvement of 0.42% compared to baseline operations, outperforming conventional manual optimization. Although the improvement is modest, it is significant for mature fields operating near optimal conditions. The findings demonstrate that data-driven surrogate models can provide efficient and flexible decision-support tools for real-time production optimization while reducing dependence on computationally intensive simulation workflows.

Keywords: *Surrogate modelling, production optimization, mature oil fields, machine learning, data-driven modelling, digital oilfield, field-wide optimization*

---------------------------***********************----------------------

I. INTRODUCTION

The demand for oil and gas remains high, but many producing fields are now mature. These fields often face challenges such as declining reservoir pressure, increasing water cut, and aging infrastructure. Because of these issues, improving production is becoming more difficult. At the same time, there is a need to increase recovery while keeping operational cost (OPEX) and environmental impact as low as possible [18].

One common approach used to handle this problem is Integrated Production System Modelling (IPSM). IPSM combines the reservoir, wellbore, and surface facilities into a single system for analysis [12]. By looking at the full system, it becomes easier to identify issues such as pressure losses and equipment limitations [18]. However, most IPSM workflows depend on physics-based simulation tools such as MBAL, PROSPER, and GAP. These tools require regular calibration through history matching to reflect real field conditions [5]. This process is time-consuming and often outdated due to the dynamic and random nature of fluid flow in brownfields [9]. In practice, production optimization is often done manually by adjusting parameters like choke size and wellhead pressure. While this method works, it is limited because it cannot fully capture the complex relationships between multiple production variables. As field conditions change over time, updating IPSM models becomes more difficult, which can delay decision-making and reduce the effectiveness of optimization efforts.

With the growth of digital oilfield technologies and the availability of high-frequency production data, new approaches have emerged. Machine learning provides a way to learn patterns directly from data without relying entirely on physical models. In this context, Surrogate models provide a data-driven alternative by learning the relationships between operational inputs and production outputs directly from historical data [10]. This makes them useful for scenario analysis and real-time decision support [9], [21].

In this study, a machine learning-based surrogate modelling framework is used to predict field-wide production and support optimization in a mature oil field. The main objectives of the study are:

i. Develop a multi-well data processing and feature engineering framework.
ii. Design and implement a machine learning-based surrogate model.
iii. Evaluate the model performance and compare it with baseline methods.

II. RELATED WORKS

The optimization of hydrocarbon production has changed over time from mainly engineering-based approaches to more data-driven methods. In recent petroleum operations, the goal is not only to increase production but also to improve asset value, often measured using net present value (NPV). This requires understanding how the reservoir, wells, and surface facilities interact with each other. In mature fields such as those in the Niger Delta and the North Sea, challenges like declining pressure and increasing water cut make this task more difficult [18].

Recent developments in digital technologies have improved how production systems are monitored and managed. Machine learning (ML) is now being used to analyse large volumes of production data and identify useful patterns. This has helped to connect traditional petroleum engineering methods with modern data-driven approaches, leading to better decision-making [15]. This section discusses Integrated Production System Modelling (IPSM), its limitations, and how machine learning-based surrogate models are being used as an alternative.

A. Evolution of Integrated Production System Modelling (IPSM)

In earlier studies, petroleum production systems were analyzed separately. Reservoir, wellbore, and surface facilities were treated as independent units. However, later research showed that these components are closely linked. For example, well performance depends on both inflow performance (IPR) and vertical lift performance (VLP), as well as surface constraints [12], [18]. To solve this, Integrated Production System Modelling (IPSM) was introduced. IPSM combines all parts of the production system into one framework. This makes it easier to identify system-wide problems such as pressure losses and facility limitations. Some studies have shown that IPSM can improve recovery efficiency by optimizing factors like gas lift allocation [3]. However, IPSM has some limitations. It depends heavily on physics-based simulation tools such as MBAL, PROSPER, and GAP. These tools require continuous calibration through history matching, which takes time and effort [5]. In mature fields where conditions change frequently, keeping these models updated becomes difficult.

B. Limiting Factor Identification and Choke Performance

Choke control plays an important role in production optimization. It regulates flow rate and helps in managing reservoir energy. Poor choke settings can lead to problems such as pressure drops, early water breakthrough, and sand production [21].

Machine learning has been applied to improve choke performance and predict production rates. For example, some studies used models like Random Forest and Gradient Boosting to predict oil production in the Niger Delta fields and achieved high accuracy [1]. Other work has shown that variables such as choke size, pressure, and fluid properties strongly affect production, which makes feature selection very important [17].

There are also studies where machine learning models were used to predict multiphase flow through wellhead chokes. These models performed better than traditional empirical methods [4]. However, most of these studies focus mainly on prediction and do not fully address field-wide optimization across multiple wells.

C. *Surrogate Modelling and the Digital Twin Concept*

With the availability of large production datasets, surrogate models are becoming more common in petroleum engineering. These models act as simplified representations of complex physical systems and can be used to test different production scenarios quickly.

Compared to traditional simulation methods, surrogate models require less computation time while still maintaining good accuracy. Some studies have shown that machine learning models can predict reservoir behavior efficiently and serve as alternatives to physics-based simulators. For instance, models have been used to predict bottom-hole pressure using surface data and engineered features [22]. These developments are part of the broader digital oilfield concept, where data-driven tools are used to support decision-making. Surrogate models are useful, especially for scenario testing and optimization in complex systems [15].

D. *Machine Learning Applications in Petroleum Engineering*

Machine learning has been used in petroleum engineering for tasks such as production forecasting, reservoir characterization, and optimization. Models like Random Forest, Gradient Boosting, and XGBoost are effective because they can handle nonlinear relationships in production data.

Studies using the Volve dataset have shown that machine learning models can achieve high prediction accuracy. In many cases, these models perform better than traditional statistical methods, especially when the data is noisy or complex. Newer approaches, such as Kolmogorov-Arnold Networks (KAN), have also been introduced to improve both accuracy and interpretability. These models are useful for capturing time-based patterns in production data [22]. Despite these advances, many studies still focus mainly on prediction. There is still a need for models that can support real decision-making and optimization in multi-well systems [15].

E. *Feature Engineering and Physics-Informed Modelling*

One challenge in applying machine learning to petroleum systems is how to represent the physical behavior of the reservoir. Early models have treated the system as a black box and did not capture important patterns such as time dependence and reservoir depletion. To solve this, recent studies have introduced physics-informed feature engineering, which involves adding features such as lagged production values, cumulative production, and operational parameters. These features can help the model better understand how the system behaves over time.

Some research has shown that combining domain knowledge with machine learning improves both accuracy and reliability. Physics-informed models can also help ensure that predictions remain consistent with real-world behavior. Despite these improvements, some gaps still exist. Many studies focus only on prediction and do not extend to optimization. Also, traditional IPSM methods are still mostly manual and limited in scope. In addition, most studies focus on single wells rather than full field systems where wells interact with each other.

This study addresses these gaps by developing a multi-well surrogate modelling framework that supports both prediction and optimization. It also introduces a yield gap analysis to compare baseline production with optimized results.

III. MATERIALS AND METHODS

This section describes the method used to develop and evaluate the multi-well machine learning surrogate modelling framework for optimization in a brownfield asset.

The study adopts an experimental design and is based on a surrogate-based optimization. The surrogate model learns relationships between the operational variables and production performance from historical data. The study is executed in four phases:

i. multi-well data acquisition and harmonization
ii. Physics-informed feature engineering

iii. Surrogate model development and benchmarking
iv. Field-wide optimization through scenario interrogation

Figure 1 shows the architectural diagram beginning with the extraction of the multi-well production data, preprocessing, feature engineering, model training, and optimization. The output is the estimation of production improvement (the yield gap).

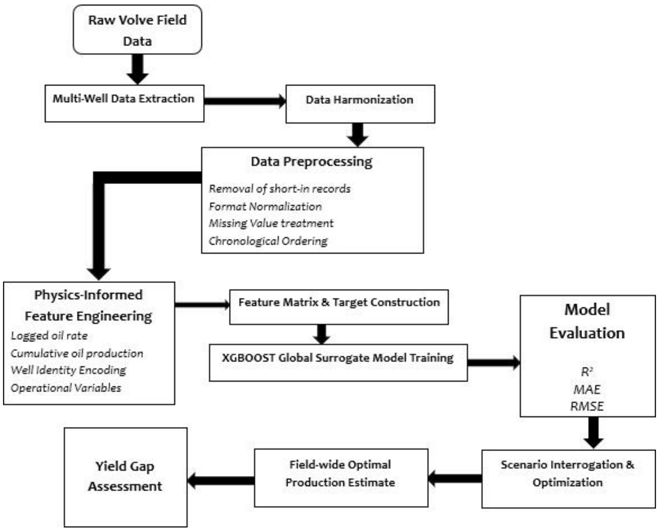

Figure 1: *Research Workflow Architecture*

The dataset used in this study is the Volve Field dataset from the Kaggle repository. The data from five wells (15/9-F-11, 15/9-F-12, 15/9-F-14, 15/9-F-15 D, and 15/9-F-1 C) were extracted; these wells comprise approximately 8,000 daily production records.

A. Data Preprocessing

A preprocessing procedure is applied to improve data quality and suitability for model development. Missing values are handled using forward-fill. The data is sorted chronologically to maintain temporal integrity. Negative production rates and unrealistic values are dropped.

B. Physics-Informed Feature Engineering

A Feature engineering method is performed to capture production system dynamics. The variables that are used as predictors include choke size or choke opening percentage, average wellhead pressure, bottom-hole pressure, gas production rate, water production rate, and On-stream hours. These variables represent the operating conditions that influence fluid production from all the wells. A lagged oil-rate variable is introduced to capture the production memory and temporal dependence. This feature represents oil production at the immediately previous time step and is defined as:

$$Q_{oil}^{lag}(t) = Q_{oil}(t-1) \qquad \text{eq. (1)}$$

A cumulative oil production is calculated for each well. This will represent reservoir depletion effects as:

$$CumOil_t = \sum_{i=1}^{t} Q_{oil}(i) \qquad \text{eq. (2)}$$

where $Q_{oil}(i)$ is the oil production rate at time i [2].

C. Surrogate Model Development

Multiple regression models, such as Linear Regression, Ridge Regression, Random Forest, Extra Trees, Gradient Boosting, Histogram Gradient Boosting, and XGBoost, are then developed. This multiple model approach will enable better evaluation of the model performance, and it will ensure that the selected model provides an accurate representation of the production system.

Linear regression assumes a linear relationship between the input and the target variables. For any given input feature vector, the predicted output is expressed as:

$$\hat{y}_i = \beta_0 + \sum_{j=1}^{P} \beta_j x_{ij} \qquad \text{eq. (3)}$$

where β_0 is the intercept term, and β_j are the regression coefficients.

The Random Forest ensemble learning method constructs multiple decision trees and aggregates their predictions. This is given as:

$$\hat{y}_i = \frac{1}{T} + \sum_{t=1}^{T} f_t(x_i) \qquad \text{eq. (4)}$$

where T is the number of trees. Random forest can improve prediction accuracy by reducing the variance through bootstrap aggregation and random feature selection [13].

The XGBoost model predicts the target value as:

$$\hat{y}_i = \phi(x_i) = \sum_{k=1}^{K} f_k(x_i) \qquad \text{eq. (5)}$$

where $\hat{y}_i$ is the predicted output.

A Histogram Gradient Boosting builds decision trees sequentially while minimizing the loss function by using gradient descent. The prediction is expressed as:

$$\hat{y}_i = \sum_{m=1}^{M} \gamma_m h_m(x_i) \qquad \text{eq. (6)}$$

Where M is the number of boosting iterations, h_m is the weak learner at iteration m, and γ_m is the learning rate or step size. At each iteration, the model is minimizing the loss function.

The histogram gradient boosting differs from the traditional gradient boosting by discretising continuous features into bins. This will improve efficiency while maintaining higher prediction accuracy.

D. Model Training Procedure

After preprocessing and feature engineering, the dataset was divided into predictor variables and the target variable. The predictor set includes operational variables, lagged oil production, cumulative oil production, and well identifier, while the target variable was the log-transformed oil production rate. A chronological data split was done so that earlier observations are used for model training and later observations are used for model testing. Multiple regression algorithms are then trained on the multi-well training dataset. Their predictive performance was compared on the test dataset using standard regression metrics.

E. Model Evaluation Metrics

The predictive performance of the surrogate models is evaluated using standard regression metrics, such as the coefficient of determination (R^2), the Mean Absolute Error (MAE), and the Root Mean Squared Error (RMSE).

1. *Coefficient of Determination* (R^2) *:* The coefficient of determination is used to measure the proportion of variance in the observed production data explained by the model:

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \qquad \text{eq. (8)}$$

A higher R^2 value indicates better predictive performance.

2. *Mean Absolute Error (MAE):* The mean absolute error is used to measure the average magnitude of prediction errors:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \qquad \text{eq. (9)}$$

3. *Root Mean Squared Error (RMSE):* The root mean squared error is also used to assess the prediction error:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad \text{eq. (10)}$$

RMSE is useful because it penalizes larger errors more strongly than MAE [13].

F. Optimization Framework

The best-performing surrogate model is used for field-wide optimization through scenario interrogation.

Unlike traditional manual optimization approaches that adjust one variable at a time, this study adopts a broader interrogation strategy in which the model evaluates multiple choke settings and pressure conditions across the five wells. The optimization problem is defined as:

$$\max \sum_{j=1}^{n} Surrogate(Choke_j, Pressure_j, History_j, Well_j) \qquad \text{eq. (11)}$$

subject to:

$$0 \leq Choke_j \leq 100 \qquad \text{eq. (12)}$$

The use of this optimization is to determine the best field-wide operating state predicted by the surrogate model and to estimate the difference between the baseline production state and the optimized production state. This difference is referred to as the Yield Gap and is expressed as:

$$\text{Yield Gap} = \frac{Q_{optimized} - Q_{baseline}}{Q_{baseline}} \times 100 \qquad \text{eq. (13)}$$

where $Q_{optimized}$ is the field production predicted under the AI-derived optimized scenario, and $Q_{baseline}$ is the field production under the observed baseline scenario [11]. This yield gap will provide a measure of the production opportunity that is identified through the surrogate-based optimization.

IV. RESULTS AND DISCUSSION

This section presents the results obtained from the experiment and explains how the proposed surrogate model performed in production prediction and optimization.

A. Model Comparison Results

Table 1 shows the Performance of the models used in this study.

Table 1: *Model Results*

Model	R^2	MAE	RMSE	MAPE (%)	Training Time(s)
Hist Gradient Boosting	0.9995	5.16	7.28	2.40	5.59
Random Forest	0.9992	6.06	9.33	2.18	12.51
XGBoost	0.9988	6.27	11.77	3.65	6.59
Extra Trees	0.9973	12.04	17.25	4.77	10.12
Gradient Boosting	0.9967	13.82	19.33	4.74	5.27
Naive Lag-1	0.9629	20.99	64.36	36.56	0.0009
Linear Regressi on	0.7225	126.2	175.9	43.78	0.07
Ridge	0.6987	129.6	183.3	42.1	0.05
Dummy Mean	-10.94	1104.5	1153.9	822.6	0.17

From the results, it can be seen that the ensemble tree-based models have performed better than the traditional regression models. In particular, the Histogram Gradient Boosting model has achieved the best performance. It achieved an R^2 value of 0.9995, with low MAE (5.16), RMSE (7.28), and MAPE (2.40%). This means that the model was able to predict production values very close to the actual data. Because of this strong performance, the Histogram Gradient Boosting model was selected for further analysis in this study. Figure 2 shows how all the models compare based on their R^2 values.

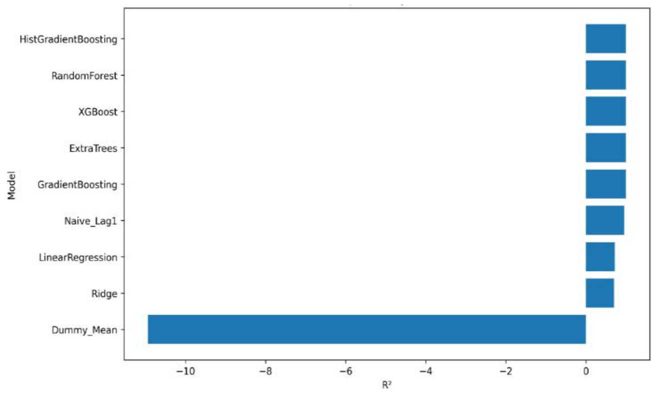

Fig 2: Model comparison by R^2

The figure gives a clear ranking of the models and confirms that the ensemble methods perform better than simpler models.

To check the accuracy of the selected model, the predicted oil production values were compared with the actual values from the test dataset. This is shown in Figure 3.

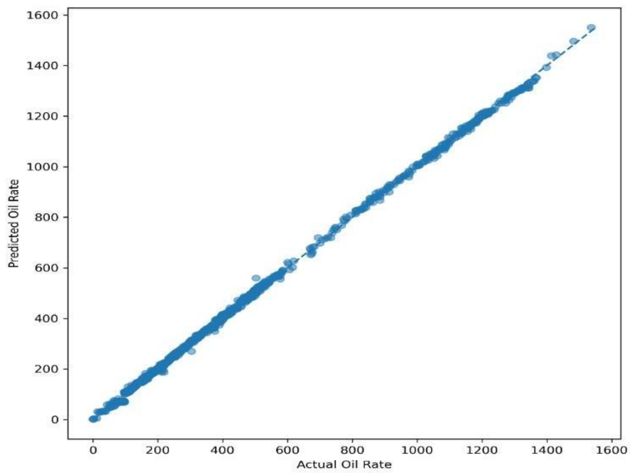

Fig 3: Actual vs Predicted oil rate

The points are closely aligned along the diagonal line, which shows that the predictions are very close to the real values. This suggests that the model is able to capture the nonlinear relationships in the production data. The residual distribution is shown in Figure 4.

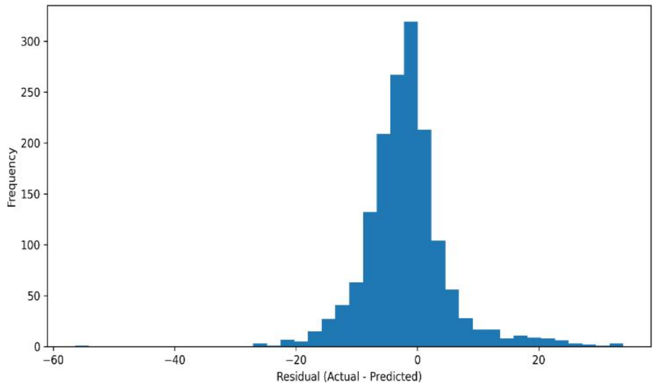

Fig. 4: Residual distribution

The histogram shows that prediction errors are centered around zero. This means there is an absence of strong bias in the model predictions. Figure 5 below shows the oil Production across the five Wells.

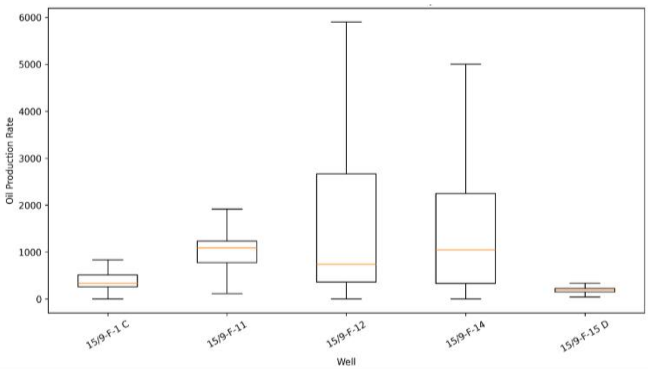

Fig 5: Oil production distribution by well

The boxplot shows that production varies across the wells. This difference between wells supports the need for a field-wide model that can handle multiple well behaviors at the same time.

Feature importance was also analyzed using the selected model. The most important variables include lagged oil production, cumulative oil production, choke size, wellhead pressure, gas production rate, and water production rate. These are expected variables based on petroleum engineering knowledge, which shows that the model is learning meaningful relationships from the data.

B. Field-Wide Optimization Results

The trained model was used to test different production scenarios. Two approaches were compared: the optimization which adjust one variable at a time, and the AI-based optimization that tests multiple variables together. Table 2 below shows the results for each well.

Table 2: *Well-Level Optimization Results*

Well	Baseline Production	Manual Optimized	AI Optimized	AI Uplift	Yield Gap (%)
F-1 C	207.44	207.44	207.44	0	0
F-11	192.58	194.20	194.20	1.62	0.84
F-12	111.77	111.77	113.06	1.29	1.16
F-14	73.18	73.18	73.18	0	0
F-15D	110.43	110.43	110.43	0	0

From the results, it can be seen that most of the production improvement came from wells F-11 and F-12. The other wells showed little or no change, which suggests that they were already operating close to their optimal conditions. This also shows that production systems are not uniform across wells. Figure 6 shows the comparison between manual and AI optimization.

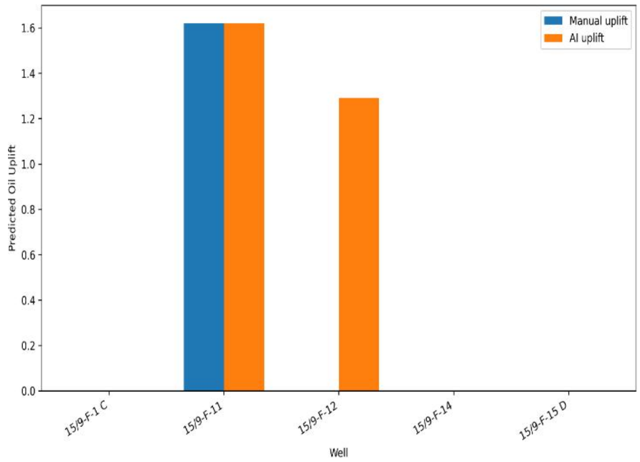

Fig 6: Manual vs AI optimization uplift by well

C. Field-Wide Yield Gap Analysis

To understand the impact, the values produced from all wells were combined, and the summary is shown in Table 3.

Table 3: *Field-wide Optimization summary*

Metric	**Value**
Baseline Field Production	695.40
Manual Optimized Production	697.02
AI Optimized Production	698.31
Manual Yield Gap	0.23%
AI Yield Gap	0.42%
Additional Gain over Manual	1.29

The AI-based method achieved a yield gap of 0.42%, compared to 0.23% from the manual approach. This shows that the surrogate model was able to find additional production opportunities, even if the improvement is small. The additional gain of 1.29 units over the manual method shows that using data-driven models can support better decision-making. At the same time, the small improvement suggests that the field may already be operating close to its best condition, which is common in mature fields. Figure 7 shows how the predicted production compares with actual production over time.

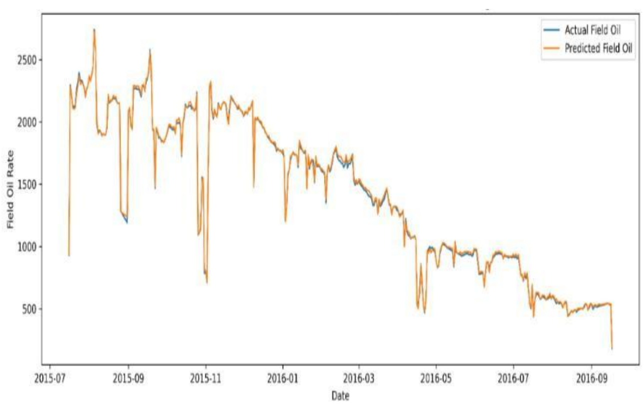

Fig 7: Field-level Actual vs predicted oil over time

The model follows the general trend of the data, which indicates that it can capture changes in production over time.

The results show that ensemble boosting models perform better than linear and simple baseline models. The Histogram Gradient Boosting model gave the best results in this case. The selected features, especially lagged and cumulative production, helped the model capture both operational and time-based behavior in the system.

V. CONCLUSION AND FUTURE WORK

In this study, a machine learning-based surrogate model was used to improve production prediction and optimization in a mature oil field. The results show that the model can learn the relationship between operational variables and production output using historical data. Among the tested models, Histogram Gradient Boosting has performed best and achieved better predictions compared to the linear models.

The study also shows that the surrogate models can be used to test different operating conditions without relying fully on complex physics-based simulators. This makes the process faster and more flexible. The optimization results indicate that the AI-based approach was able to find small additional production gains compared to the manual method.

Another important point is that combining petroleum engineering knowledge with machine learning can improve the model's performance. Features such as lagged production and cumulative production helped the model to capture real system behaviour. This shows that domain knowledge is still important when applying machine learning in engineering problems.

This study contributes by developing a multi-well modelling approach, testing different machine learning models, and applying the best model for production optimization. It also provides a simple workflow that can be reused in similar oil field problems.

For future work, the model can be improved by combining machine learning with physics-based simulation to form a hybrid system. Reinforcement learning can also be explored to allow real-time adjustment of production parameters as new data becomes available. In addition, this approach can be extended to other systems such as gas production, offshore operations, and even renewable energy systems.

ACKNOWLEDGMENT

The authors would like to acknowledge the support of their respective institutions. We also appreciate the availability of the Volve field dataset, which made this research possible.

REFERENCES

[1] O. T. Adewale, S. O. Olatunji, and A. A. Adebayo, "Machine learning approaches for production forecasting and choke optimization in the Niger Delta oil wells," *Journal of Petroleum Science and Engineering*, vol. 215, p. 110764, 2026.

[2] T. Ahmed, Reservoir Engineering Handbook, 5th ed. *Gulf Professional Publishing*, 2019.

[3] S. Ahuja, N. Kumar, and R. Singh, "Integrated production system modelling for optimization of mature oil fields," *Journal of Petroleum Exploration and Production Technology*, vol. 12, no. 3, pp. 873-885, 2022.

[4] S. A. Alarifi, "Machine learning-based prediction of multiphase flow through wellhead chokes," *Journal of Petroleum Science and Engineering*, vol. 208, p. 109312, 2022.

[5] A. Alkindi and S. Linthorst, "Integrated production system modelling for reservoir management and production optimization," *SPE Production & Operations*, vol. 27, no. 2, pp. 128-138, 2012.

[6] A. S. Arief, A. Nugroho, and B. Setiawan, "Feature engineering strategies for oil production prediction using machine learning models," *Energy Reports*, vol. 7, pp. 8423-8432, 2021.

[7] O. Bello, S. Adebayo, and T. Ibraheem, "Digital oilfield technologies and data-driven production optimization in upstream petroleum operations," *Energy Informatics*, vol. 6, no. 1, pp. 1-16, 2023.

[8] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, Time Series Analysis: Forecasting and Control, 5th ed. *Wiley*, 2015.

[9] D. Bukharbayeva, Z. Liu, and Y. Kurniawan, "Surrogate modelling techniques for reservoir and production optimization: A review," *Journal of Petroleum Science and Engineering*, vol. 189, p. 107021, 2020.

[10] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," *in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785-794.

[11] M. J. Economides, A. D. Hill, and C. Ehlig-Economides, Petroleum Production Systems, 2nd ed. *Pearson*, 2013.

[12] J. O. Eli, M. A. Adewumi, and B. Ogunyomi, "Integrated production system modelling for improved field performance," *Journal of Petroleum Technology*, vol. 65, no. 4, pp. 74-82, 2013.

[13] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed. *Springer*, 2009.

[14] International Energy Agency, World Energy Outlook 2023. *International Energy Agency*, 2023.

[15] J. Kim, D. Lee, and S. Park, "Artificial intelligence applications in geoenergy systems: A review of digital oilfield technologies," *Applied Energy*, vol. 355, p. 121876, 2025.

[16] M. H. Kutner, C. J. Nachtsheim, J. Neter, and W. Li, Applied Linear Statistical Models, 5th ed. McGraw-Hill/Irwin, 2005.

[17] X. Ma, Y. Zhang, and H. Li, "Machine learning-based production prediction using operational and reservoir variables," *Journal of Natural Gas Science and Engineering*, vol. 115, p. 104987, 2025.

[18] S. O. Mepaiyeda, M. Adewumi, and M. O. Onyekonwu, "Production optimization in the Niger Delta mature fields using integrated production system modelling," *SPE Production & Operations*, vol. 39, no. 2, pp. 223-237, 2024.

[19] M. K. Ng, R. Hauge, and O. Sævareid, "Oil production forecasting using machine learning: A case study of the Volve field dataset," *Journal of Petroleum Science and Engineering*, vol. 196, p. 107733, 2021.

[20] A. Samad, M. Khan, and S. Ahmed, "Machine learning models for oil production prediction using the Volve dataset," *Energy AI*, vol. 14, p. 100273, 2025.

[21] A. Sircar, K. Yadav, and S. Mishra, "Production optimization in offshore oil fields through choke management and machine learning," *Energy Reports*, vol. 7, pp. 4583-4595, 2021.

[22] A. Adeyinka, O. Oriola, and O. S. Tomomewo, "An innovative approach for oil well bottomhole pressure forecasting using Kolmogorov-Arnold Neural Networks (KANs): A case study in an offshore oilfield," *Deep Resources Engineering*, vol. xx, pp. 100233, 2025.