Scientific Contributions Oil & Gas, Vol. 48. No. 4, December: 203 - 227

SCIENTIFIC CONTRIBUTIONS OIL AND GAS

Testing Center for Oil and Gas
LEMIGAS

Journal Homepage:http://www.journal.lemigas.esdm.go.id
ISSN: 2089-3361, e-ISSN: 2541-0520

The Integration of Hybrid Capacitance Resistance Model and Machine Learning: A Data-Based Workflow for Optimizing Waterflood Performance and Reservoir Management

Syifa Alviola Muhendra, Novia Rita, Fajril Ambia, and Agus Dahlia

Department of Petroleum Engineering, Faculty of Engineering, Universitas Islam Riau
Kaharuddin Nasution Street No. 113, Simpang Tiga, Pekanbaru, Riau, Indonesia.

Corresponding Author: Novia Rita (noviarita@eng.uir.ac.id)

Manuscript received: October 20th, 2025; Revised: November 06th, 2025
Approved: December 05th, 2025; Available online: December 17th, 2025; Published: December 17th, 2025.

ABSTRACT - This study aims to minimize uncertainty in waterflood performance by employing a data-driven workflow that combines the Capacitance Resistance Model (CRM) with Machine Learning. Two CRM variants, CRM-P (Producer-based) and CRM-IP (Injector-Producer-based), are utilized to evaluate interwell connectivity and time constants on three reservoir models: homogeneous, heterogeneous, and a real field scenario (Volve Field). The model is evaluated using R^2 and Mean Absolute Percentage Error (MAPE) and is compared against the Random Forest and eXtreme Gradient Boosting (XGBoost) techniques. The results indicate that CRM-IP provides more realistic estimates than CRM-P, particularly for response time. XGBoost consistently demonstrates superior prediction accuracy, achieving R^2 values of 0.76–0.98 and MAPE values of 0.5–10%. Three-dimensional (3D) visualizations of interwell connectivity and streamline analysis strengthen the understanding of fluid flow and sweep efficiency. This further demonstrates that integrating CRM and Machine Learning serves as a decision-support tool for Enhanced Oil Recovery optimization, as evidenced by R^2 and MAPE analyses that characterize sweep efficiency and the reservoir's capacity to accommodate additional injection.

Keywords: waterflood, CRM, interwell connectivity, machine learning, streamline.

How to cite this article:

Syifa Alviola Muhendra, Novia Rita, Fajril Ambia, and Agus Dahlia, 2025, The Integration of Hybrid Capacitance Resistance Model and Machine Learning: A Data-Based Workflow for Optimizing Waterflood Performance and Reservoir Management, Scientific Contributions Oil and Gas, 48 (4) pp. 203-227. https://doi.org/10.29017/scog.v48i4.1928.

INTRODUCTION

The continuously increasing global energy demand encourages the petroleum industry to maximize production from existing fields by applying enhanced oil recovery (EOR) methods (Du et al., 2024). One of the most commonly used methods in the secondary recovery stage is waterflooding. This method maintains reservoir pressure and pushes remaining oil toward production wells, thereby increasing sweep efficiency and extending field production life (Malvi´c et al., 2020. However, despite being relatively simple and widely implemented, the waterflood method often faces challenges due to reservoir heterogeneity, non-uniform permeability distribution, and uncertainty in interwell connectivity (Fu et al., 2022). These factors can cause early water breakthrough and significantly reduce oil recovery efficiency (Guo et al., 2019).

Successful waterflooding can increase the oil recovery factor from a typical 5-25% in the primary stage to a typical 45% of original oil in place (OOIP) (Usman & Haans, 2017. Waterflood performance evaluation is typically conducted through full-scale reservoir simulation, which requires detailed geological data and extensive computational time. This approach is less efficient when rapid decision-making is needed for field management (Makhotin et al., 2022). To address this issue, data-driven approaches such as the capacitance resistance model (CRM) have been developed to estimate the relationship between injection and production rates using an electrical system analogy (Fu et al., 2022). CRM can calculate interwell connectivity strength and time constants using historical injection and production data, without requiring complex fluid-flow simulation. The two main CRM formulations commonly used are CRM-P (Producer-based) and CRM-IP (Injector-Producer-based), each representing systems with different levels of complexity.

Despite its effectiveness, conventional CRM has limitations in capturing nonlinear behavior in complex and heterogeneous reservoir systems. This model tends to assume a linear relationship between injection and production; thus, it is not fully capable of representing physical phenomena such as channeling, water-saturation variations, and dynamic pressure changes across production zones. To overcome these limitations, CRM is combined with decline curve analysis (DCA), which corrects long-term production rate decline trends (Fu et al., 2022). The integration of CRM+DCA enhances the model's predictive capability during the production decline phase, resulting in smoother history matching and more realistic estimations of reservoir behavior.

In this research, Machine Learning techniques such as Random Forest and eXtreme Gradient Boosting (XGBoost) are not used as primary models but rather as benchmarks to validate the DCA model's prediction accuracy. This approach aims to test the extent to which the CRM-DCA model can achieve the accuracy of advanced data-based algorithms while ensuring model reliability in representing physical reservoir conditions.

This study focuses on implementing a data-driven workflow that combines CRM and DCA to evaluate waterflood performance (secondary recovery water injection) across three different reservoir models: homogeneous, heterogeneous, and real field (volve field). Through analysis of interwell connectivity parameters and time constants, and through result validation using Machine Learning, this research aims to demonstrate the effectiveness of the CRM-DCA model in reducing uncertainty, understanding the dynamic response of waterflood systems, and supporting strategic decision-making in reservoir management and performance optimization.

METHODOLOGY

This study uses a Tnavigator simulator to present historical data on injection and fluid production rates, which are derived from both synthetic models and actual field data. Additionally, CRM-P and CRM-IP are developed in Python, and the model is integrated with DCA.

These models are utilized to forecast historical matching and to assess their efficacy using statistical metrics, including R-squared (R^2) and Mean Absolute Percentage Error (MAPE) (Yousef

et al., 2006). If the R^2 value is close to 1 and the MAPE is below 10%, the model is considered optimal. To strengthen the analysis, machine learning techniques such as Random Forest and XGBoost are employed as benchmarks for the historical matching results generated by the CRM model.

The data utilized include the SPE 5 and SPE 10 synthetic datasets, as well as actual field data from the Volve Field. The selection of these models aims to evaluate how effectively CRM and machine learning can be used to characterize fluid movement in reservoirs.

Capacitance resistance model (CRM)

The CRM is a valuable tool for improving real-time flood management and reservoir analysis, as it accurately simulates gas- and waterflood recovery processes. The CRM is a material–balance–based model that requires only injection and production histories, which are the most accessible data obtained during a reservoir's production life (De Holanda et al., 2018). This study introduces two formulations by examining distinct control volumes to represent varying levels of modeling complexity.

CRM-P: producer-based model

Represents in-situ volumetric balance over the effective pore volume of a producer. According to Nguyen (2012), CRM-P is formulated using Equation 1.

$$q_t(t) = q_j(t-1)e^{-\left(\frac{\Delta t}{\tau_j}\right)} + \left(1 - e^{-\left(\frac{\Delta t}{\tau_j}\right)}\right) \left[\sum_{i=1}^{N_i} f_{ij}\, i_i(t) - J_j \tau_j \frac{P_{wf,j}(t) - P_{wf,j}(t-1)}{\Delta t}\right] \quad (1)$$

where the CRMP assigns one time constant for the draining volume of each producer and one connectivity for each injector – producer. Also Sayarpour et al., (2009, CRMP is a model in which the producer is the center or focus of the control system. For this reason, the CRMP is not recommended for very heterogeneous reservoirs; it performs better when near-homogeneity is present near the producers and when all injectors are at similar distances from the producers, such as in a patterned waterflood.

CRM-IP: injection-producer based model

The volumetric balance in the reservoir is evaluated over the affected pore volume of each injector–producer pair, using Equation 2 from (Nguyen 2012).

$$q_{ij}(t) = q_{ij}(t-1)e^{-\left(\frac{\Delta t}{\tau_{ij}}\right)} + \left(1 - e^{-\left(\frac{\Delta t}{\tau_{ij}}\right)}\right) \left[f_{ij}\, i_i - J_j \tau_{ij} \frac{P_{wf,j}(t) - P_{wf,j}(t-1)}{\Delta t}\right] \quad (3)$$

According to De Holanda et al. (2018), where (q_{ij}) j is the production rate in producer *j* from the injector (*i*)-producer (*j*), as well as $J_j \tau_{ij}$ is the productivity index associated with such a control volume.

Decline curve analysis

According to Maurenza et al. (2023), the Arps method is widely used in production forecasting to predict production performance and estimate remaining reserves. In this study, decline curve analysis (DCA) is applied as a complementary approach to enhance CRM predictions. This integration is necessary because the CRM alone struggles to capture production behavior when a significant decline trend occurs. Therefore, the Exponential Decline equation from DCA is applied as follows:

$$q = q_i e^{(-D_i \Delta t)} \quad (3)$$

where the current production rate q is determined from the initial production rate, which decreases exponentially with the initial decline rate *Di* over the time interval Δt . The exponential indicates that production declines gradually and continuously over time, with the rate of decline determined by the magnitude of *Di.*

Machine learning with random forest and extreme gradient boosting (XGBoost)

The Random forest method operates by constructing numerous decision trees and obtaining the outcome through voting (for classification) or averaging (for regression) across all trees (B. Liu et al., 2025). The primary advantages include its

resistance to overfitting and the absence of data scaling requirements. Meanwhile, XGBoost provides superior accuracy, efficient processing of large datasets, and improved handling of imbalanced datasets compared to Random Forest. The application of machine learning without careful consideration results in inaccurate predictions, which ultimately leads to unusable algorithms. Such inaccuracies arise from common errors and limitations frequently encountered in the application of machine learning, particularly in scientific fields (Zainuri et al., 2023.

In this case, for fields with more than two wells, the data is split using a well-based approach to ensure that the model performs consistently across all wells. However, this split is not applied to the two-dimensional SPE 5 model, which includes only one representative well. Thus, the combination of RF and XGBoost not only provides robust historical-matching results but also facilitates local (well-by-well) data interpretation, thereby supporting faster, more accurate field development decisions.

RESULT AND DISCUSSION

Homogeneous reservoir model

In this reservoir, gas and water are injected into a single injection well, but gas injection has been shut off, and continuous water injection is currently in use. In the base scenario, the permeability value in this model is 50 mDarcy across the entire board.

The simulation system delivers excellent history -matching results in Figure 1, highlighting the two main phases of field production. The initial phase, concluding in 1994, was characterized by primary depletion, during which oil was the only fluid produced. The second phase shows a classic response to water injection: the total fluid rate

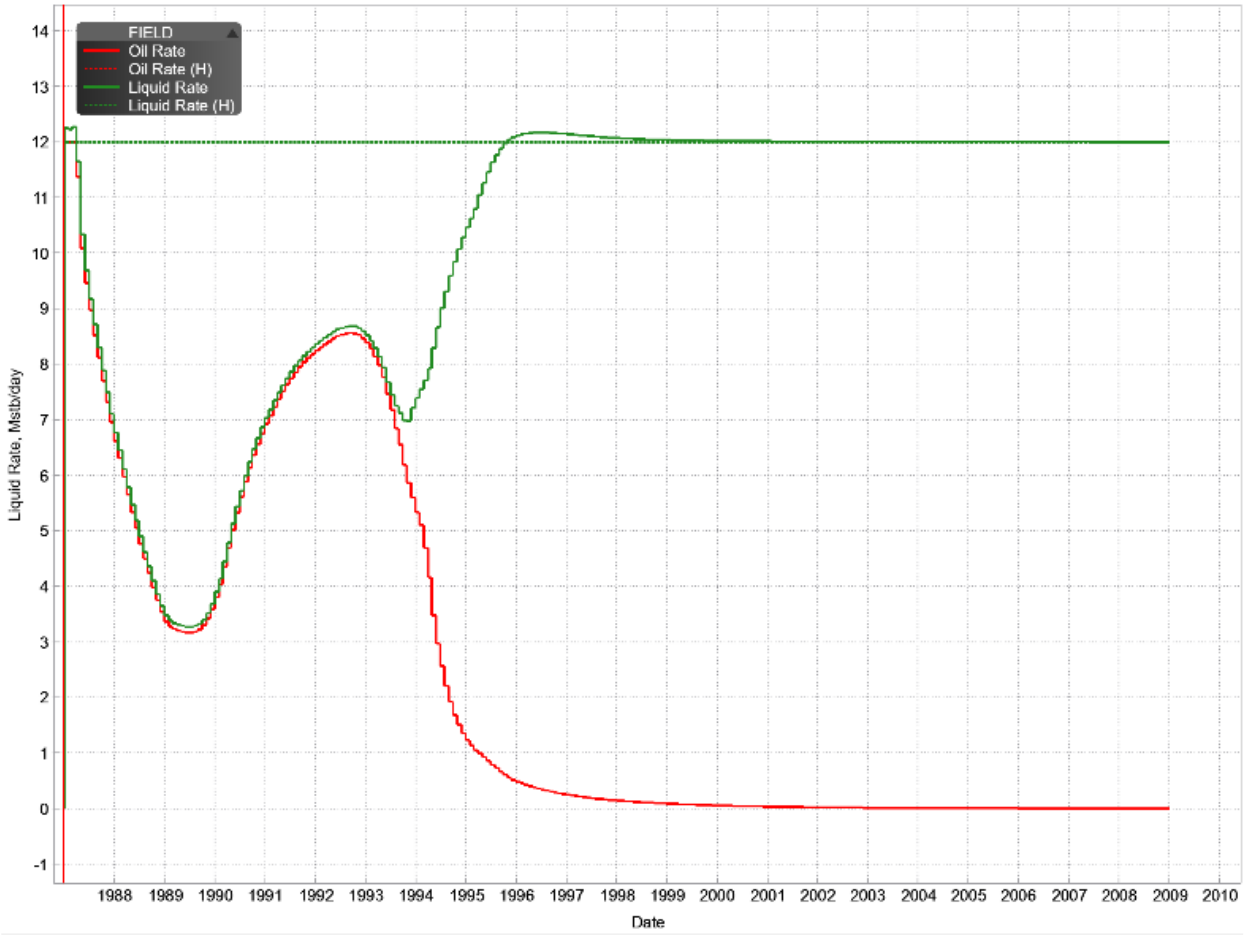


Figure 1. The plot of liquid rate vs oil rate in the homogeneous model

increases and stabilizes, while the oil rate decreases dramatically due to water breakthrough. At the end of the period, the field produces almost entirely water, indicating a high water-sweep efficiency.

History matching analysis and validation of CRM-P and CRM-IP in homogeneous model

A comprehensive evaluation of the hybrid model on well P1 is presented in Figure 2. Although the time-comparison plots and statistical metrics show very high accuracy, they also highlight limitations in fully capturing the complex model of reservoir behaviour. This finding is consistent with research that indicates that hybrid models, despite providing accurate predictions based on conventional statistical metrics, still face challenges in representing complex physical phenomena in reservoir systems (Fan et al., 2025).

Interwell connectivity and time constant analysis in homogeneous model

Table 1 presents a comparison of dynamic reservoir parameters using the CRM-P and CRM-IP models for production well P1. The results show very significant differences in interpretation between the two models. The CRM-P model estimates very weak interwell connectivity (0.000786) and an extreme response time constant of 1,314.20 days. In contrast, the CRM-IP model identifies a much stronger connectivity (0.0102) and a much faster time constant (98.03 days). This drastic difference strongly supports the CRM-IP model, as its response time is much more physically realistic for active waterflood dynamics than the delay time indicated by the CRM-P model.

Table 1. Interwell connectivity and time constant CRM-P and CRM-IP homogeneous model

CRM Type	Well	Injection Well	Time Constant (Day)
CRM-P	P1	0.000786	1314.20
CRM-IP	P1	0.0102	98.03

The perfect single-flow path in Figure 3 shows that there is no geological complexity that could interfere with fluid flow. The high efficiency of this homogeneous base-case scenario has proven to be an important benchmark for reservoir performance evaluation.

Homogeneous conditions allow for optimal sweep efficiency because there are no permeability variations that inhibit flow. Direct pressure from the injector to the producer reflects ideal waterflooding conditions. (Ogbeiwi et al., 2018) Confirm that effective pressure maintenance occurs when there is direct connectivity between the injection and production wells. Meanwhile, (Sidiq et al., 2019) explain that under homogeneous conditions, the displacing phase can move optimally without interference from an adverse mobility ratio.

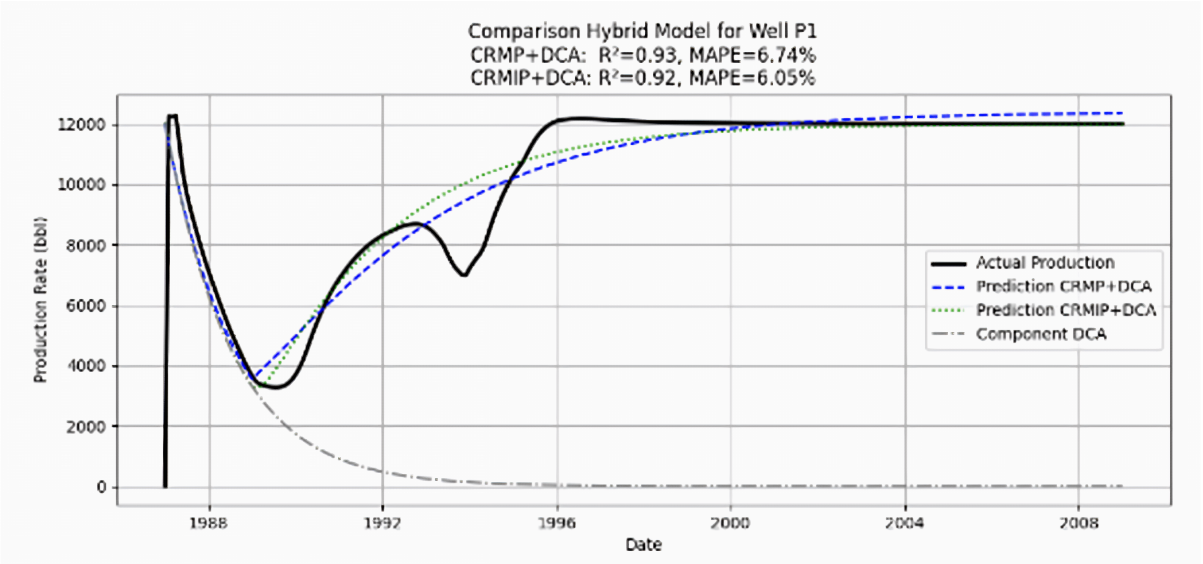


Figure 2. The Plot of production rate vs date for CRM-P and CRM-IP in homogeneous model

Compared to other models, this model enables a systematic evaluation of the impact of reservoir complexity. Ismailova et al. (2021) Emphasize that understanding homogeneous reservoirs as a basic reference is essential for analyzing how formation heterogeneity can cause irregular flow profiles and uneven fluid distribution under more realistic conditions.

Streamline analysis for homogeneous reservoir model

In Figure 4, a streamline plot shows what would happen if an ideal fluid were to flow through a homogeneous reservoir. The white lines/curves show the optimal flow path for fluids from the injector to the producer. This figure illustrates an ideal scenario of 100% sweep efficiency, a rare occurrence in actual field operations. In comparison, (Ramadhan et al., 2023 reported that homogeneous reservoirs achieved an oil recovery factor of 59.86%, whereas heterogeneous reservoirs ranged from 45.83% to 80.46%, depending on heterogeneity level.

Streamline in this homogeneous reservoir served as an important baseline for understanding fluid flow dynamics (Krogstad et al., 2017. This visualization allows reservoir engineers to analyze sweep efficiency and identify areas that may not be swept in a complex field.

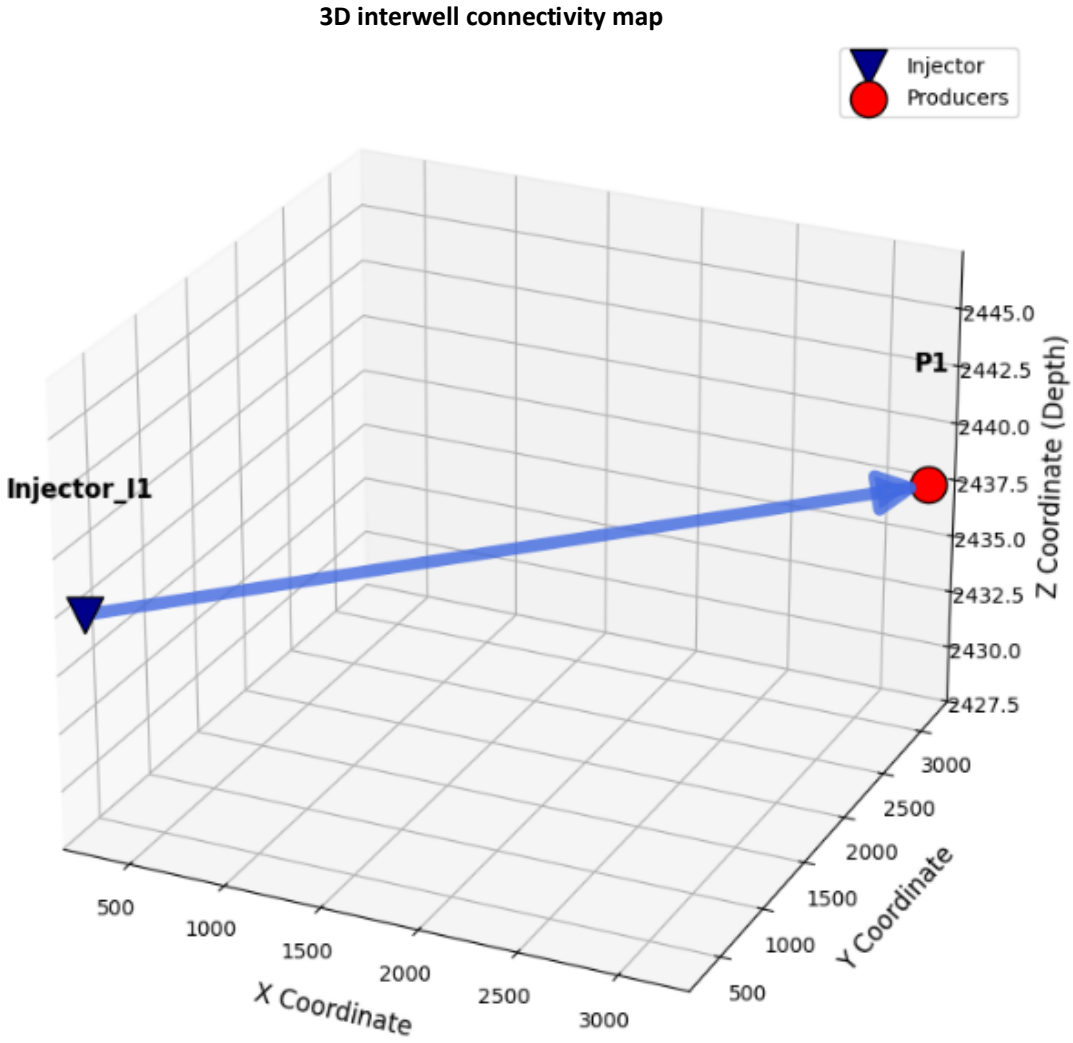


Figure 3. 3D mapwell connectivity in homogeneous reservoir

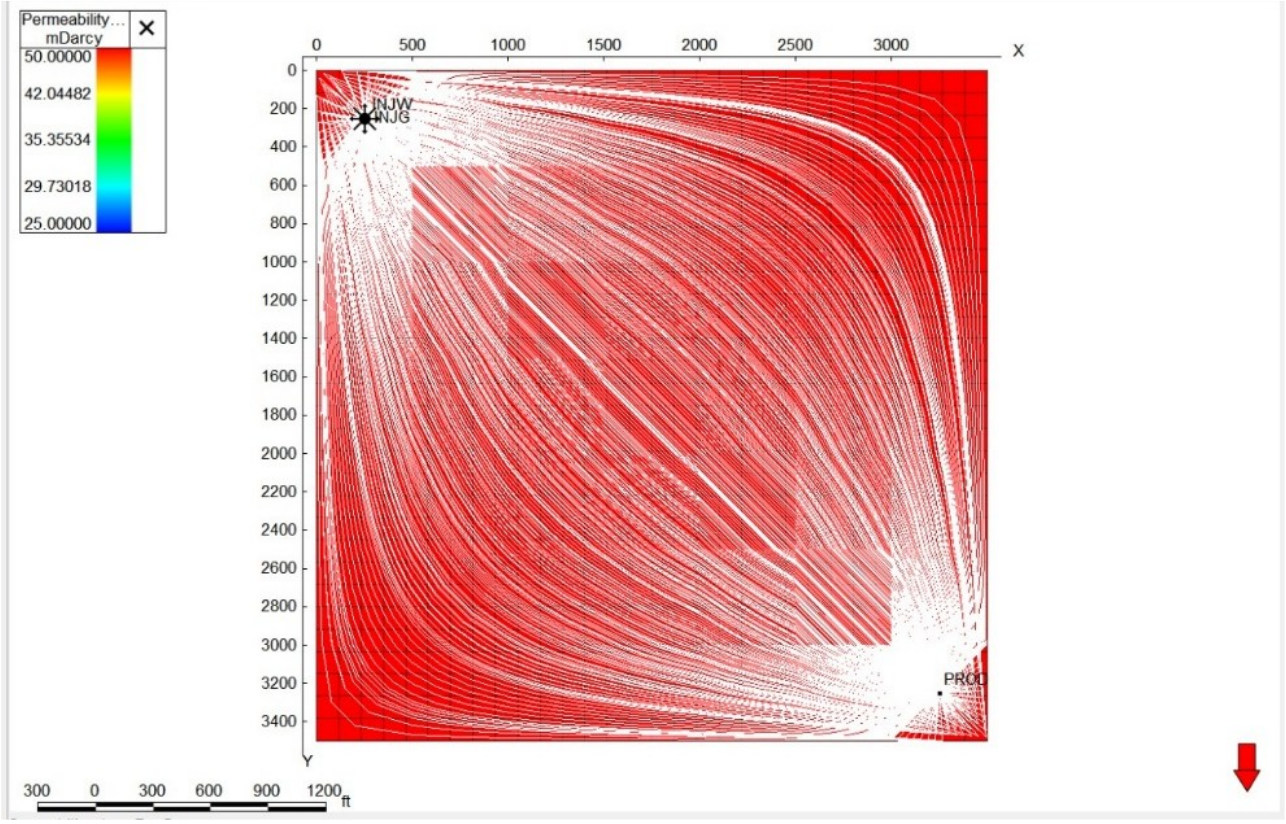


Figure 4. 2D Streamline with permeability analysis in homogeneous model

Machine learning analysis with random forest and eXtreme gradient boosting in homogeneous model

Figures 5 and 6 show Random Forest and XGBoost plots in production prediction evaluation. They demonstrated that both machine learning algorithms, Random Forest and XGBoost, perform exceptionally well during the history-matching phase. The XGBoost model had a slight advantage, with an R^2 of 0.9999 and an MAPE of 0.17%, compared to the Random Forest's R^2 of 0.9992 and MAPE of 0.50%.

(Sri Chandrahas et al., 2022 found that the XGBoost model outperformed other ensemble methods, such as Random Forest, in terms of MAPE, root mean square error (RMSE), and R^2 values. (Fadzil et al., 2021 also showed that XGBoost was the best and most stable model for predicting kinematic viscosity and base oil viscosity index during validation, as evidenced by its accuracy. This indicates that both models are good at predicting production with high confidence and do not exhibit any signs of overfitting.

Heterogeneous reservoir model

The heterogeneous reservoir model with a vertical permeability distribution that varies in Figure 7. The most noticeable feature is the presence of high-permeability pores (red/yellow), which will control fluid flow. The existence of these preferential flow paths creates the risk of early breakthrough and will result in uneven sweep efficiency throughout the reservoir.

Figure 8 shows field production under strong waterflood, characterized by a constant total fluid rate of 4.9 mmstb/day. A significant decrease in oil rate, in contrast to a stable fluid rate, indicates rapid water breakthrough and a continuous increase in water content. This profile is characteristic of a pressure maintenance strategy in which the oil produced is continuously replaced by water.

History matching analysis and validation of CRM-P and CRM-IP in heterogeneous model

The results show a very high level of prediction accuracy. Visually, in Figure 9, the time comparison graph shows that the model predictions

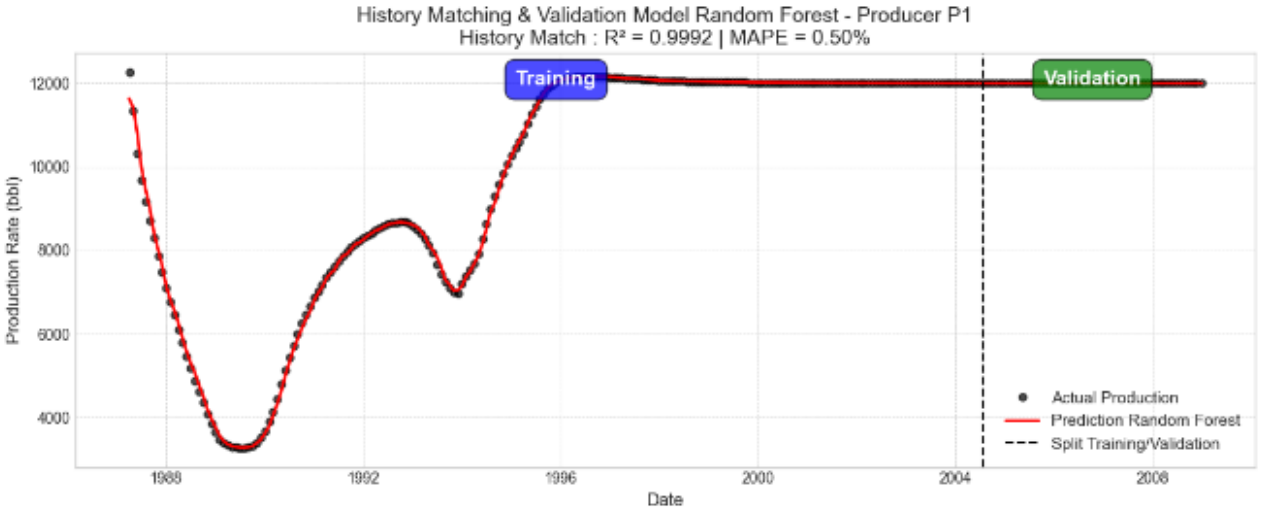


Figure 5. History matching of random forest model for prediction rate in reservoir

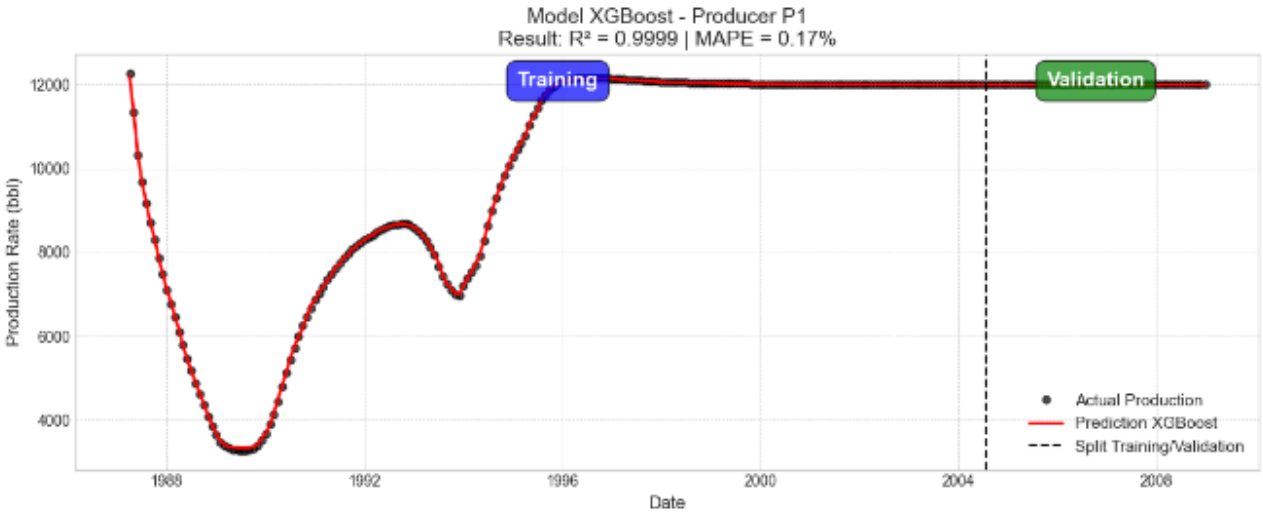


Figure 6. History matching of eXtreme gradient boost model for prediction rate in reservoir

(both CRMP and CRMIP) almost perfectly overlap with the actual production data, hence successfully replicating the initial surge and long plateau phase. From the regression graphs in Figure 10, this model confirms superior performance, with the CRMIP+DCA model achieving an R^2 of 0.88 and a MAPE of only 1.46%. The distribution of errors is concentrated around zero, without any clear pattern, indicating that this model has no systematic bias and is highly reliable.

Overall, the combination of these two analyses provides very strong evidence that the developed hybrid model is a highly accurate and robust representation of field performance.

Interwell connectivity and time constant analysis in heterogeneous model

Tables 2 and 3 present a comparison of parameters estimated by the CRM-P and CRM-IP models for the heterogeneous reservoir case study. These results surprisingly show a very high level of convergence between the two formulations. The interwell connectivity values for each producer are essentially identical between the two models, with P3 consistently identified as the strongest and P1 as the weakest. The time constant parameter (ô) also shows striking similarities, with both models uniformly estimated at around 30 days. This convergence indicates that the heterogeneous

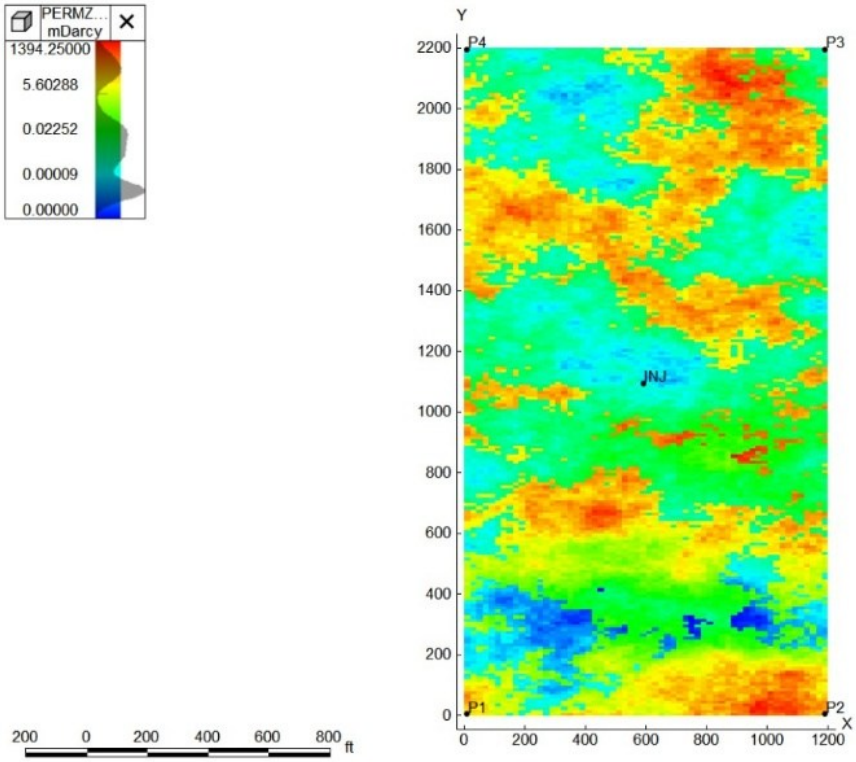


Figure 7. 2D permeability area in heterogeneous reservoir model

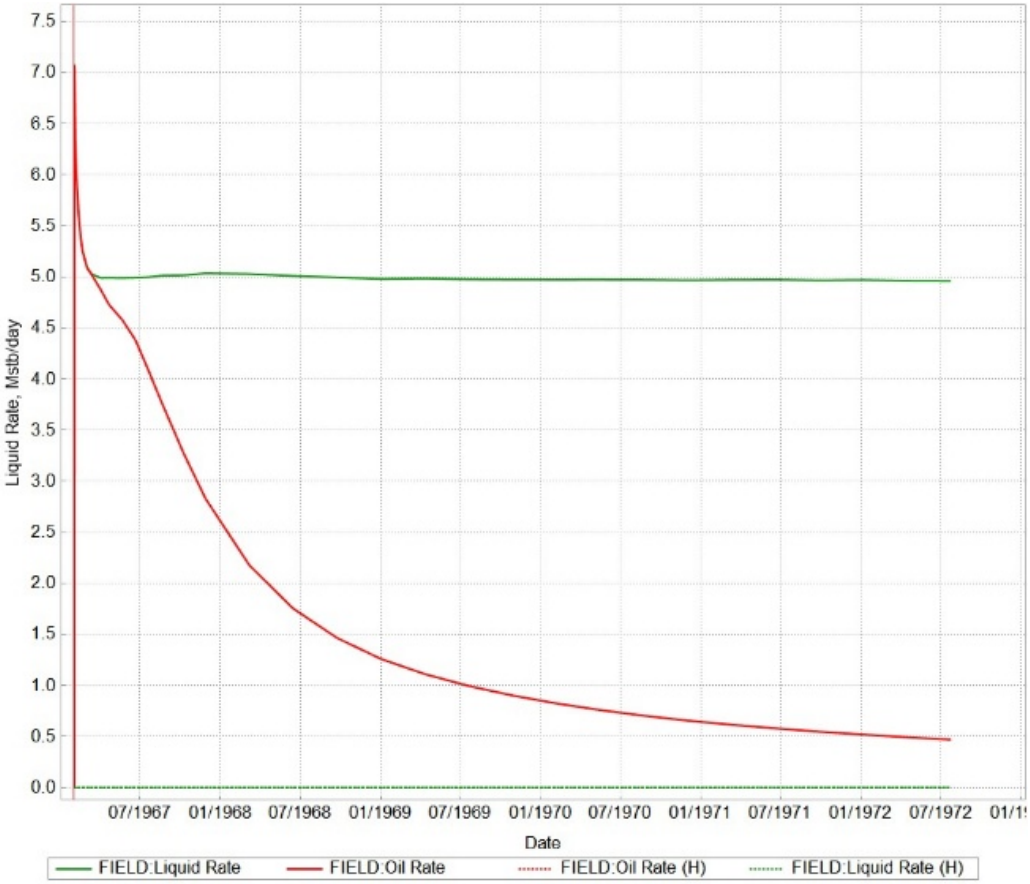


Figure 8. The plot of liquid rate vs oil rate in the heterogeneous model

reservoir system modeled by the CRM-P is adequate and yields results comparable to those of the more complex CRM-IP formulation.

Table 2. Interwell connectivity and time constant CRM-P heterogeneous model

Production well	Connectivity Injection Well	Time Constant (Day)
P1	0.003018	30.38
P2	0.006232	30
P3	0.015958	30
P4	0.00798	30

Table 3. Interwell connectivity and time constant CRM-IP heterogeneous model

Production well	Connectivity Injection Well	Time Constant (Day)
P1	0.0031	30
P2	0.0062	30
P3	0.0160	30
P4	0.0080	30

The 3D visualization in Figure 11 clearly identifies the preferential flow paths within the reservoir. Injector_I1 shows the most dominant connectivity to the P3 production well, as indicated by the thickest arrow. This shows the risk of early breakthrough at P3 and forms the basis for optimizing the injection strategy to balance sweep efficiency across the field.

The importance of optimizing injection strategies to mitigate the risk of early breakthrough is supported by (Pratama & Saptadji 2021), who emphasize that the right production-injection strategy can lead to low reservoir pressure and minimal decline rates. Their research shows that a combination of distributed and deep injection strategies provides the best results in maintaining long-term reservoir performance.

In the context of injection optimization for sweep efficiency, it demonstrates that modeling and optimizing the injection process are crucial for determining the optimal injection scenario, enabling optimization results to identify injection parameters that maximize oil recovery while minimizing operational risks.

Streamline analysis for heterogeneous reservoir model

Figure 12 shows the impact of reservoir heterogeneity on flow patterns. Flow lines are not evenly distributed; rather, they are concentrated in areas of high permeability. This pattern visually confirms the high risk of early breakthrough and uneven sweep efficiency, with most of the reservoir area likely to be missed. This phenomenon is supported by various studies, which show that reservoir heterogeneity has a significant impact on macroscopic connectivity and sweep efficiency

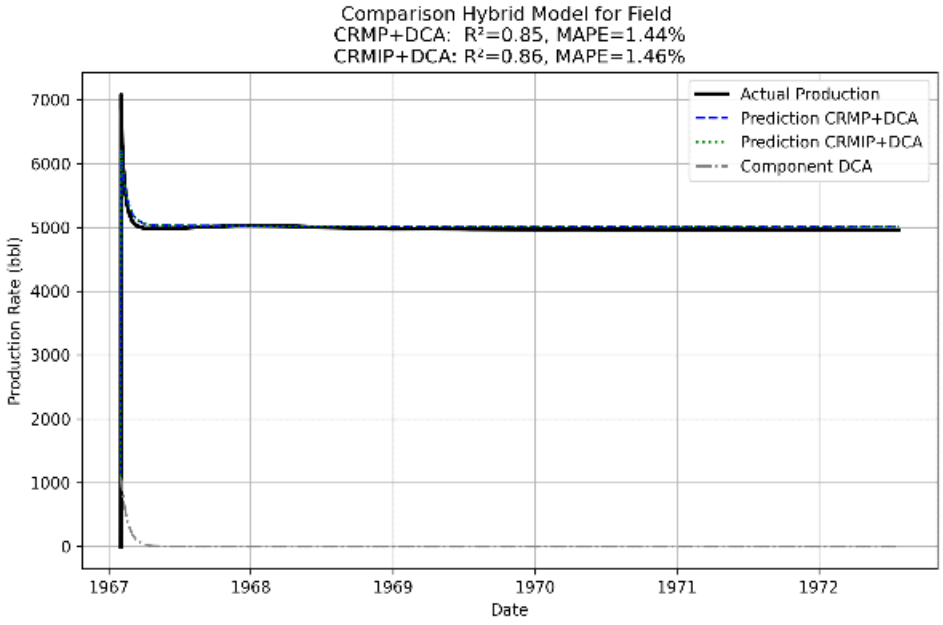


Figure 9. The plot of production rate vs date for CRM-P and CRM-IP in the heterogeneous model

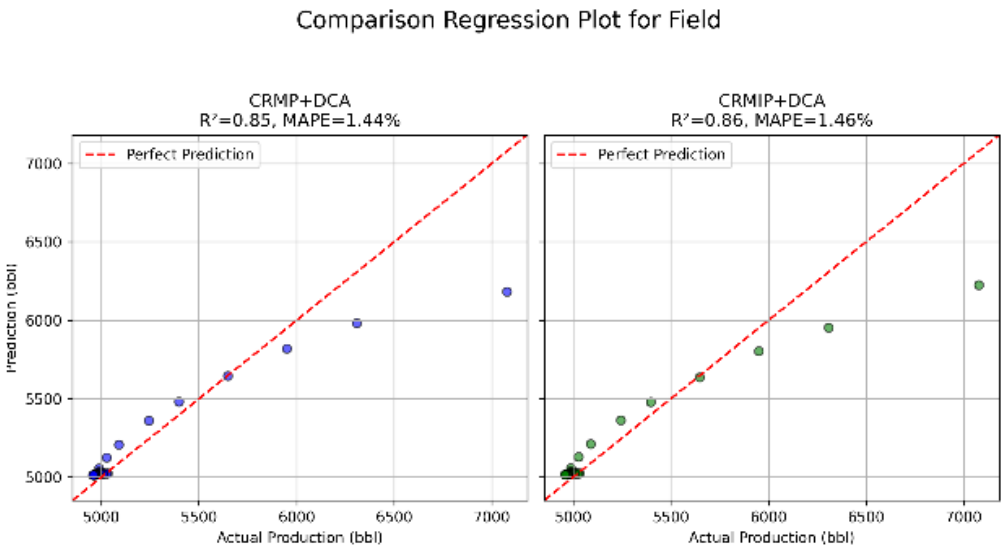


Figure 10. The Regression plot producer well heterogeneous model (a) the result with CRMP (b) the result with CRMIP

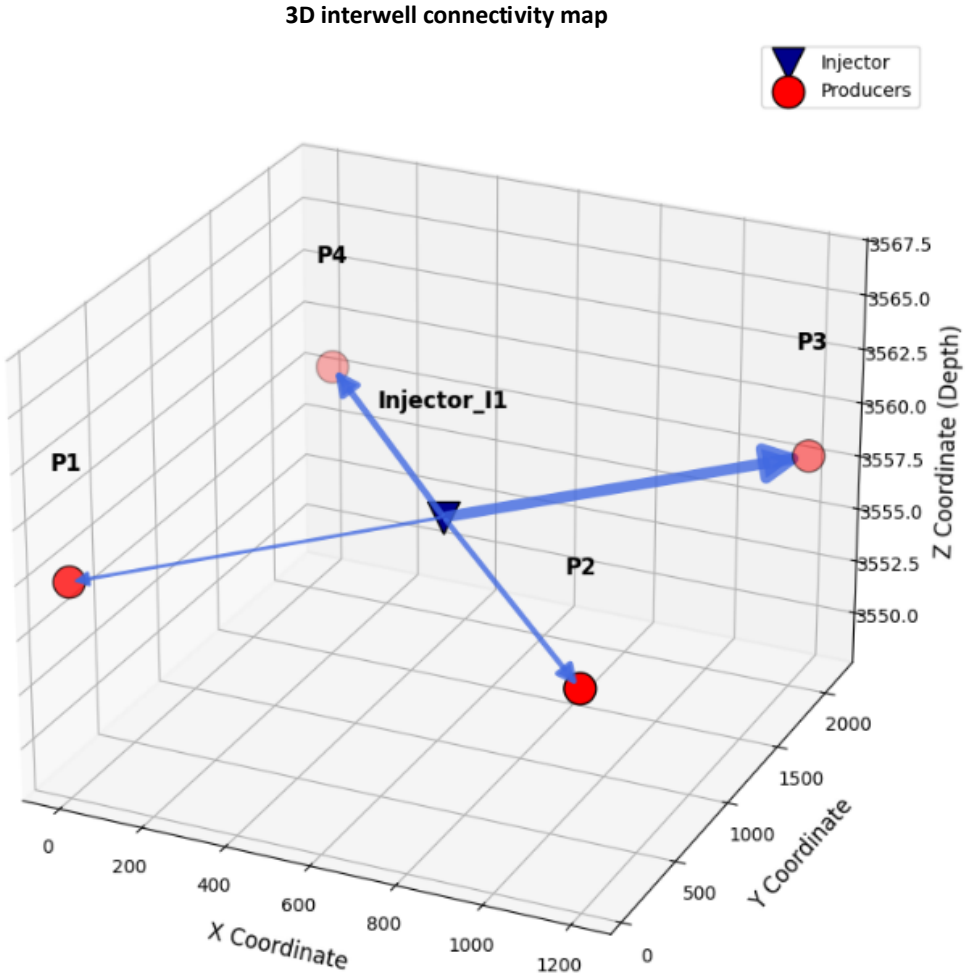


Figure 11. 3D map well connectivity in heterogeneous reservoir

(Unless et al., 2017). Studies show that high heterogeneity in reservoirs leads to early breakthrough and production well blockage, while some oil-bearing areas remain difficult to access (Ismailova et al., 2021).

Analysis of flow in heterogeneous media shows that the injected fluid exhibits channeling, leading to rapid penetration and low area-sweep efficiency. (Soltanmohammadi et al., 2024).

Machine learning analysis with random forest and eXtreme gradient boosting in heterogeneous model

The performance evaluation of the Random Forest and XGBoost models was conducted by dividing them into training phases (P1, P2, P3) and validation (P4). Both models show very high accuracy on the training data. However, on the validation data, XGBoost proved to be much more accurate than Random Forest. These results confirm that the model did not experience overfitting and that XGBoost is the superior model for this forecasting application, as shown in Figure 13.

Real field model reservoir

Figure 14 shows the real field reservoir model of the Volve Field. Its main features are a complex fault system that serves as a layer separator and low-permeability zones (blue) that act as flow barriers. Although it is characterized by high permeability (red/orange), this complex structure will lead to uneven sweep efficiency and pose a major challenge for waterflood management. The results of reservoir model validation through history, which match the process against actual

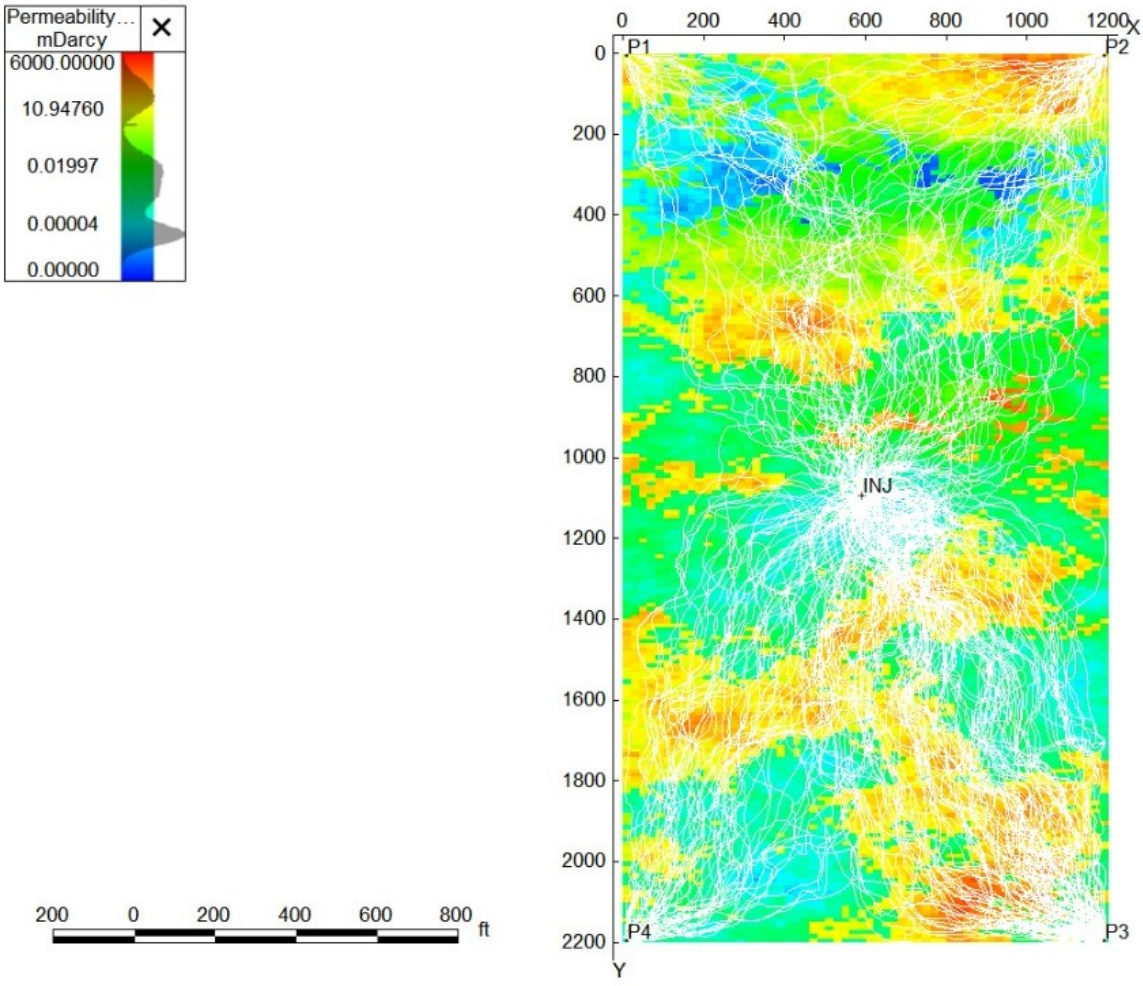


Figure 12. 2D streamline with permeability analysis in heterogeneous model

field production data, are shown in Figure 15. Visually, the simulation results (solid line) closely match the historical data (markers), indicating that this model accurately represents reservoir behavior. The production profile shows the field peaked in mid-2016 and then declined steadily. Notably, water production increased substantially over time, as evidenced by the widening gap between the total fluid rate and oil rate curves. This phenomenon indicates that the field is in a mature production stage with a strong water drive mechanism. The success of this validation is very important, as it provides confidence that the model is a reliable representation of the reservoir and can be used for future scenario forecasting.

History matching analysis and validation of CRM-P and CRM-IP in a real field model

Based on the analysis presented, Figure 16 provides a comprehensive performance analysis of hybrid models at the field scale. Both models (CRMP & CRMIP) capture actual production trends effectively despite their high volatility. Although they do not perfectly replicate every peak and trough, the model predictions consistently follow the general behavior of the historical data. Both models achieved an R^2 of 0.76, indicating that they explained 76% of the variation in the production data. The MAPE of around 20% suggests an acceptable level of prediction error for complex field-scale models, with the CRMIP+DCA model performing slightly better. The use of MAPE as a model evaluation metric is consistent with practices in time series modeling. (Ruhiat & Effendi 2018).

Overall, the combination of these two analyses provides strong evidence that the developed hybrid model is a valid and reliable representation of field performance. The level of accuracy achieved is comparable to that of other predictive model studies, which reported 95.93% accuracy (Widodo et al., 2017), indicating that this hybrid approach is effective for field-scale applications.

Interwell connectivity and time constant analysis in a real field model

Table 4-6 presents the quantitative results of CRM modeling, which summarize the interwell connectivity parameters (f_{ij}) and time constants (ô). Evaluating interwell connectivity using modeling methods has proven effective in analyzing reservoir parameters (J. Liu 2020). Analysis of the CRMP model Table 4, which assumes one time constant per production well, clearly identifies that Injector I_1 has the most dominant influence on Producer P2, with the highest connectivity value (f = 0.019).

The CRMIP model tables 5-6, which provide a more detailed analysis with unique parameters for each well pair, confirm these findings. Table 5 shows that the strongest connectivity remains between I_1 and P2, demonstrating the robustness of the analysis results. According to Poplygin et al. (2022), understanding factors influencing flow velocity between wells is crucial. Notably, Table 6 reveals that the CRMIP model's time constants are relatively uniform, ranging from 30 to 32 days across different well pairs. Overall, the consistency of connectivity results between the two models increases confidence in the identification of the main flow path. In addition, the uniformity of the time constant calculated by the CRMIP model is an important finding that indicates the hydraulic properties of the reservoir in the modeled area are relatively homogeneous, given that reservoir heterogeneity has a significant impact on fluid flow and recovery factors. (Ramadhan et al., 2023).

Table 4. Interwell connectivity and time constant CRM-P real field

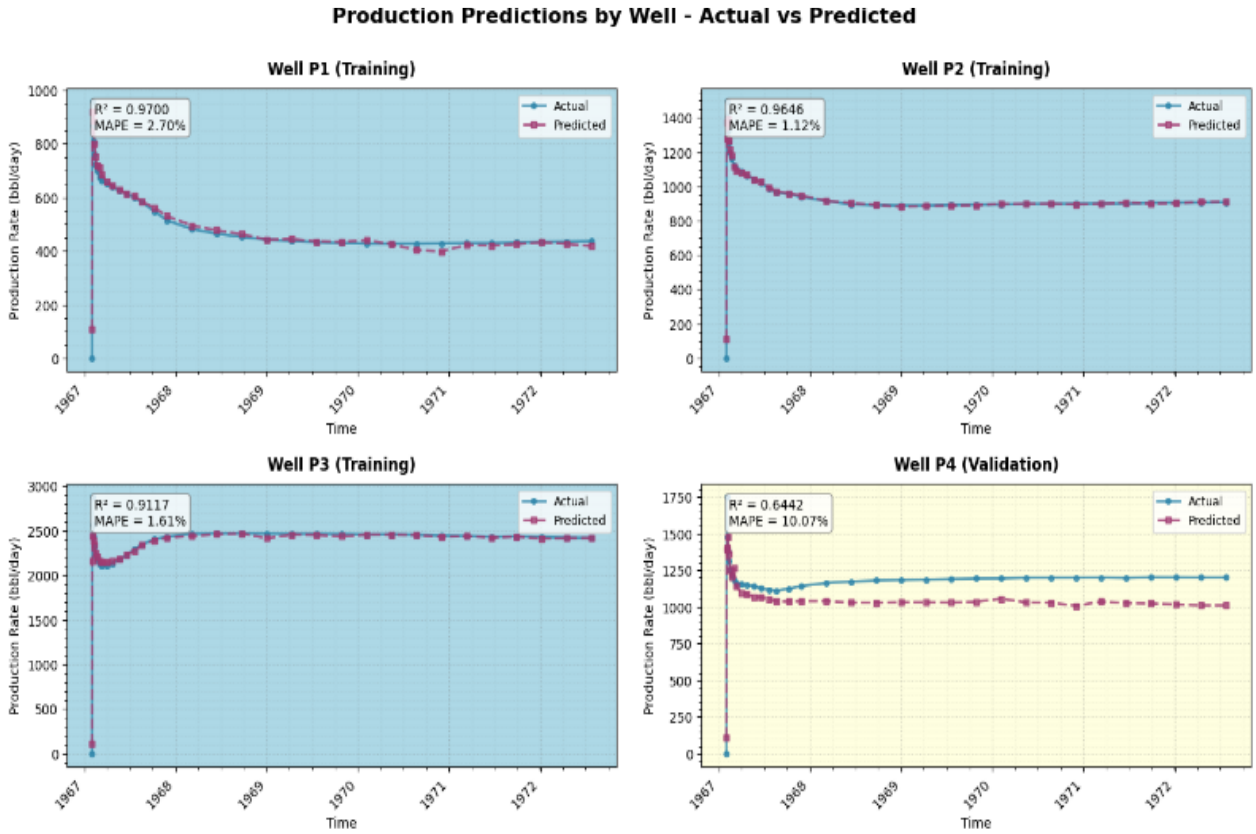
Production well	Connectivity Injection Well		Time Constant
	I_1	I_2	
P1	0.0066	0.0098	30
P2	0.019	0.011	30
P3	0.0066	0.0098	30

Table 5. Interwell connectivity CRM-IP real field

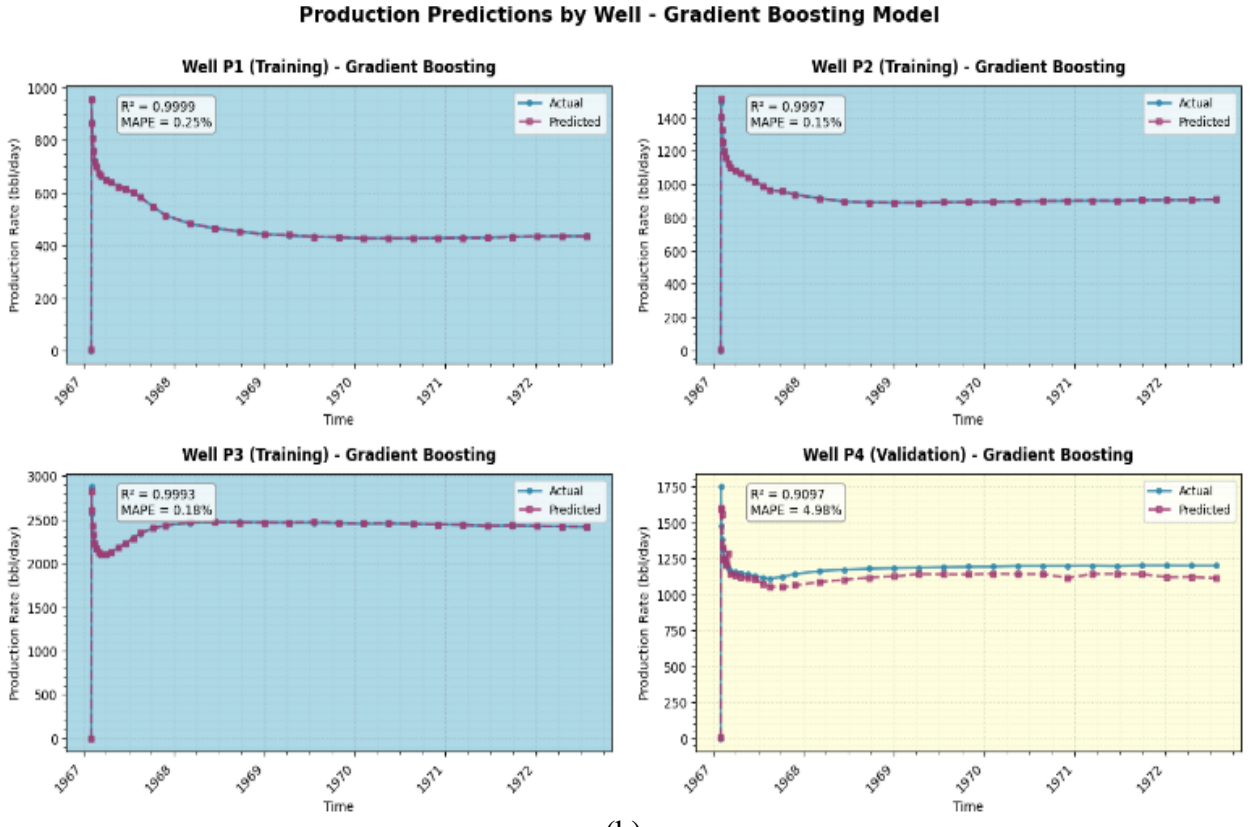
Production well	Connectivity Injection Well	
	I_1	I_2
P1	0.0094	0.0064
P2	0.0191	0.0109
P3	0.0067	0.0099

Table 6. Time constant CRM-IP real field

Production well	Time Constant (τ) Injection Well	
	I_1	I_2
P1	32	30.55
P2	30.86	30.54
P3	30	30



(a)



(b)

Figure 13. The production prediction result with training and validation data (a) random forest model (b) eXtreme gradient boost (eXBoost) model

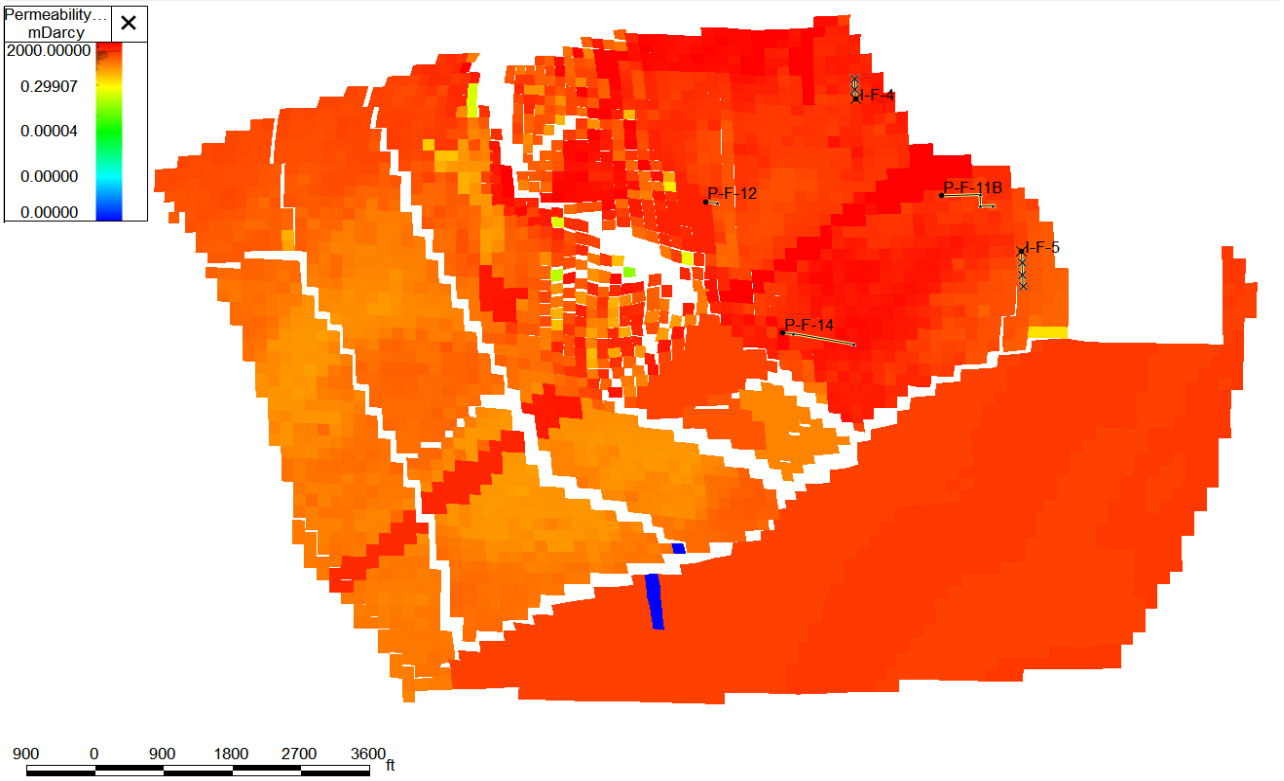


Figure 14. 2D permeability area in real field model

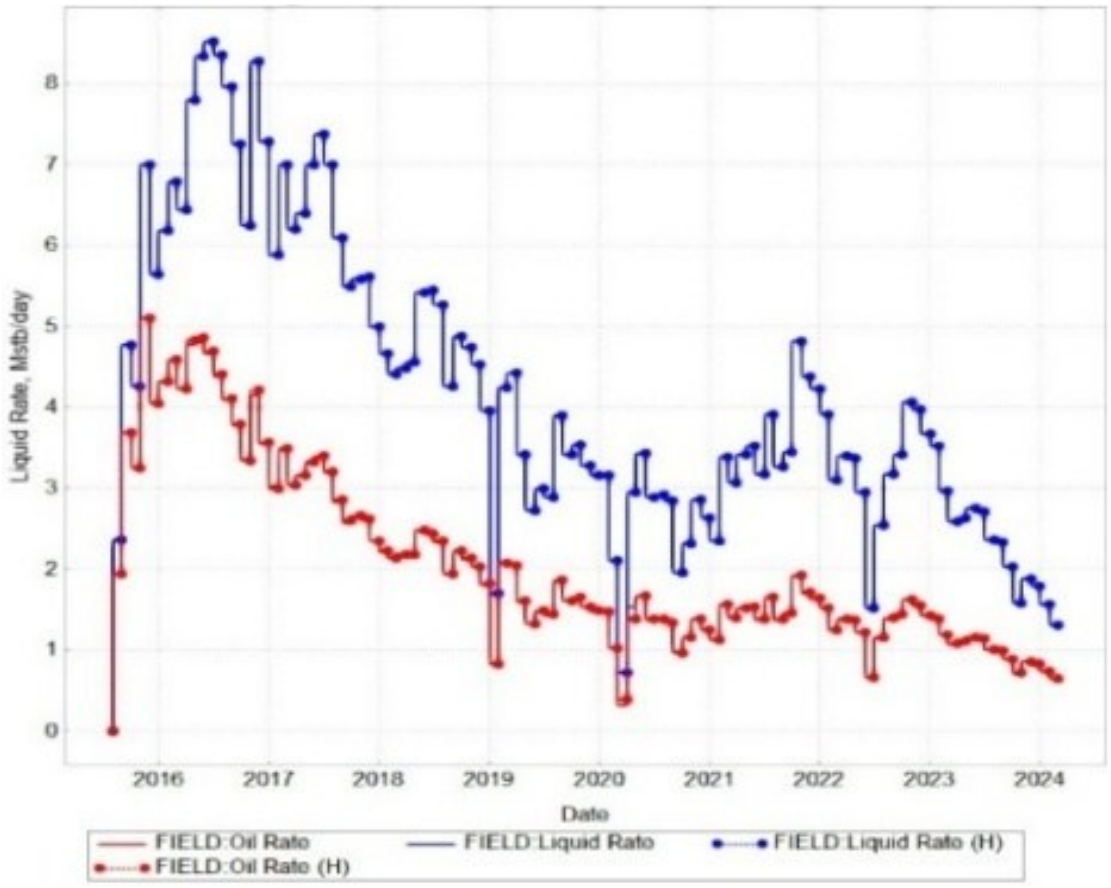


Figure 15. The plot of liquid rate vs oil rate in the real field model

Figure 16. The plot of production rate vs date for CRM-P and CRM-IP in the real field model

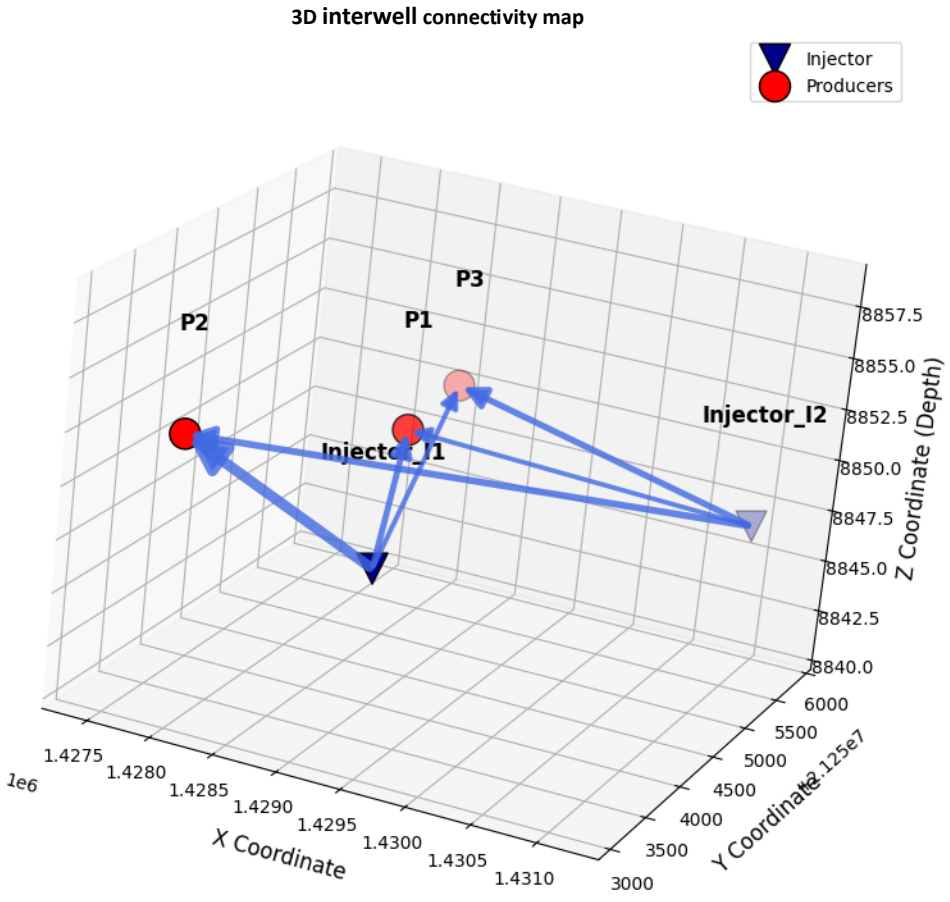


Figure 17. 3D map well connectivity in real field model

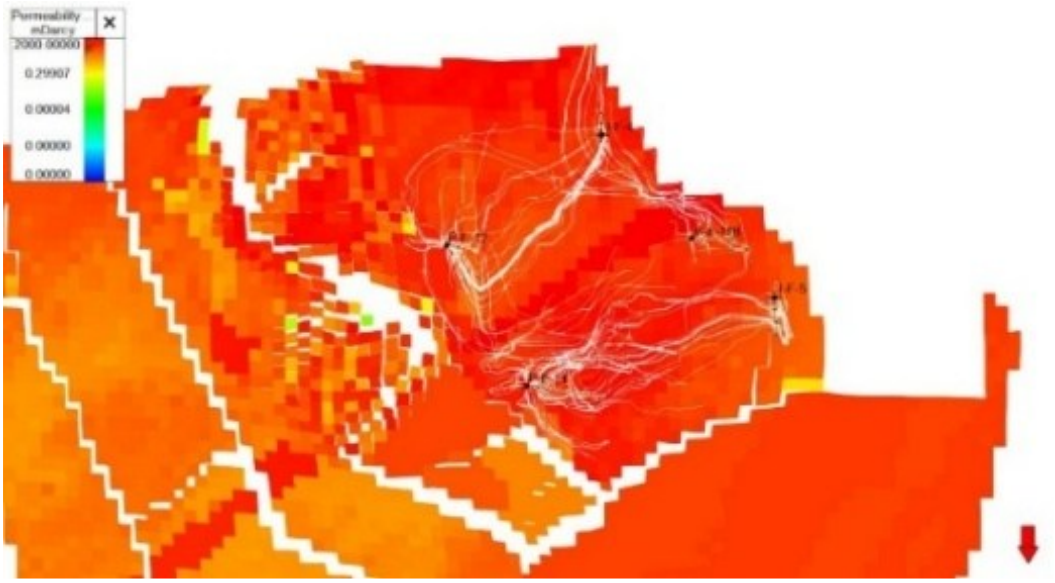


Figure 18. 2D streamline with permeability analysis in a real field model

The 3D visualization in Figure 17 clearly identifies Injector I_1 as the dominant injector with strong connectivity to all producer wells. In contrast, Injector I_2 has a more focusd impact, particularly on well P1. This map forms the basis for optimizing the injection strategy to improve sweep efficiency. This aligns with the principles of waterflood optimization, which state that understanding the connectivity between wells is crucial for designing effective injection strategies (Ogbeiwi et al., 2018). Research shows that optimizing injection rates and injection well placement are key factors in improving oil recovery (Izadmehr et al., 2018). The right injection strategy can improve reservoir pressure maintenance and sweep efficiency, especially when waterflood is carried out in the same zone as production.

Streamline analysis for real field model

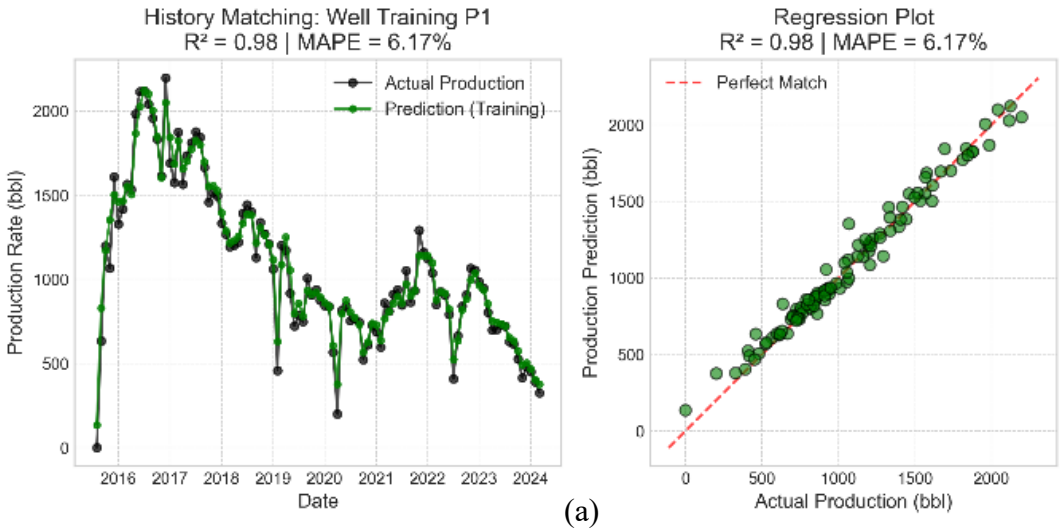
Figure 18, a visualization of fluid flow in this Real Volve Field model, effectively demonstrates the critical impact of geological complexity on fluid injection performance. The uneven flow pattern with concentration in the high-permeability zone in the northeast is consistent with existing research on heterogeneous reservoir behavior. The observed channeling effect has been well documented in the literature. (Weijermars & van Harmelen 2017) Shows how fault barriers and permeability variations cause the loss of planned drainage symmetry and divert the waterflood pattern. At (Pandey et al., 2023), models with high horizontal permeability showed reduced area-sweep efficiency and early breakthrough, which align with the observed flow concentration pattern. (Ramadhan et al., 2023) show that heterogeneous reservoirs can achieve recovery factors ranging from 45.83% to 80.46% depending on permeability correlation, with the lowest values occurring when channeling dominates.

The risk of early breakthrough is a fundamental concern in heterogeneous systems. (Wang et al., 2019) Explains that unstable phase flow, which is determined by underlying heterogeneity, slows down flow in unswept-away areas. This visualization effectively demonstrates how heterogeneity and fractures fundamentally reduce the effectiveness of injection projects, compared with ideal scenarios.

Machine learning analysis with random forest and eXtreme gradient boosting in a real field model

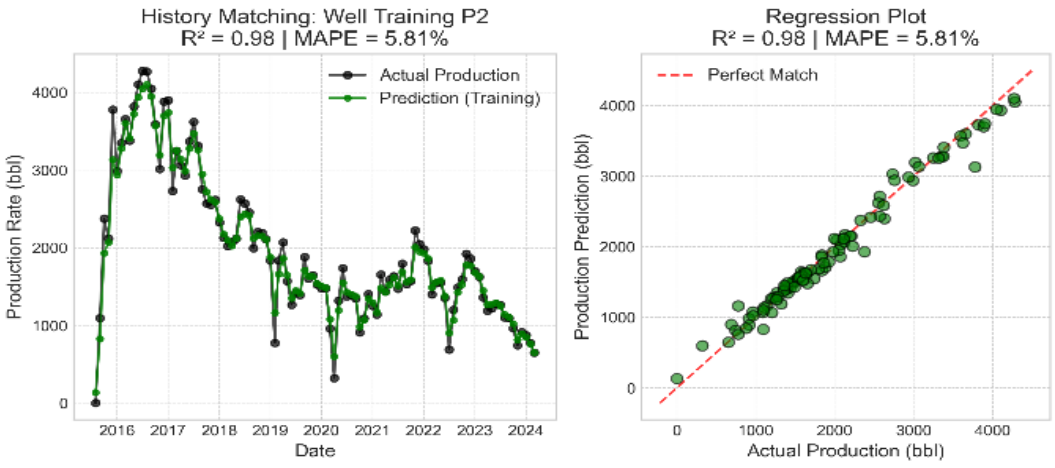
Performance evaluation of the Random Forest machine learning model for the same period, Figure 19 shows very promising results. Based on the historical matching and validation results presented, the Random Forest model shows outstanding performance, with very high training accuracy, achieving $R^2 = 0.98$ for wells P1 and P2.

Analysis performance for well P1



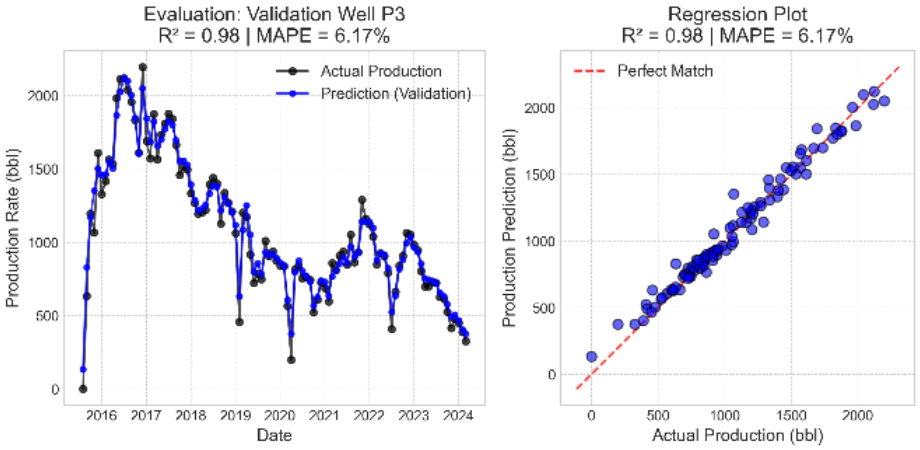
(a)

Analysis performance for well P2



(b)

Analysis performance for well P3



(c)

Figure 19. History matching & validation random forest with regression plot well production (a) well 1 (b) well 2 (c) well 3

More importantly, the model maintained consistent, strong performance during the validation stage, with $R^2 = 0.95$ for well P3. These results indicate that the model does not overfit and exhibits excellent generalization. This performance is consistent with other research findings demonstrating the effectiveness of Random Forests in similar applications. (Rahmanifard & Gates 2024).

The model's consistent accuracy between training and validation data indicates that Random Forest effectively captures complex production data patterns while retaining the ability to accurately predict new data (Ng et al., 2022. This superior performance underscores the great potential of machine learning approaches, particularly Random Forest, for developing highly accurate and reliable production forecasting systems in the oil and gas industry (Moradi et al., 2023).

As a further comparison (Figure 20), the performance of a more advanced machine learning model, XGBoost, was also evaluated using the same methodology, with wells P1 and P2 as training data and well P3 as validation data (Chakraborty & Elzarka 2019). In the training phase using data from wells P1 and P2, the XGBoost model achieved near-perfect accuracy and a strong fit to the historical data. This is evidenced by the R^2, which reached 1.00, and by the MAPE, which was very low at below 0.6%. It demonstrated the model's ability to learn complex production data patterns with very high precision (Alshboul et al., 2022). The true validation lies in the model's performance on previously unseen data from the P3 well.

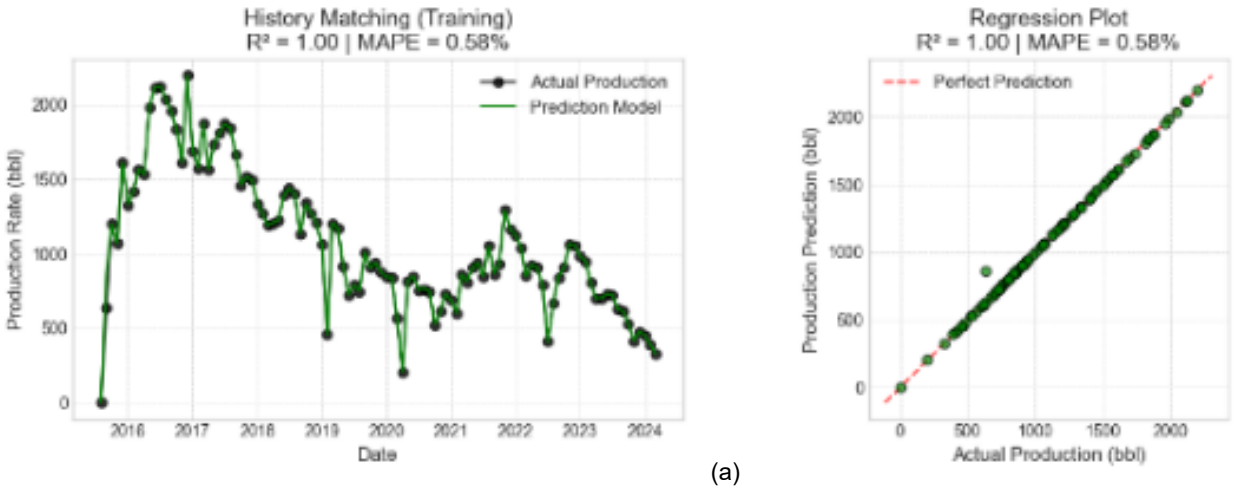
At this validation stage, the model maintained its outstanding performance, with $R^2 = 0.99$ and MAPE = 0.86%. The very high consistency in performance between the training and validation data convincingly demonstrates that the model did not overfit and has strong generalization capabilities. (Shahani et al., 2021). The near-linear regression plot validates the XGBoost model's accuracy and reliability. It demonstrated its superiority for production forecasting in manufacturing (De-Prado-gil et al., 2022).

Comparative analysis of CRM-DCA & machine learning to prediction results for water injection in reservoir model variations

To strengthen the scientific validity of this study, the hybrid CRM–machine learning approach is conceptually compared with other commonly used reservoir analysis methods. Classical CRM provides a clear physical basis but exhibits reduced sensitivity in highly heterogeneous reservoirs. Decline curve analysis is effective for identifying production trends, yet it does not represent the dynamic interaction between injection and production (Li et al., 2022). Pure machine-learning models can capture complex nonlinear relationships but lack a physical interpretive framework. The hybrid CRM–ML approach integrates the strengths of both domains, resulting in a more comprehensive and accurate representation of waterflood dynamics (Jiang et al., 2022; Reginato et al., 2023).

Analysis of results from three reservoir models, homogeneous, heterogeneous, and real field, demonstrates that the level of geological complexity directly influences interwell connectivity, injection response time, and fluid sweep efficiency. In the homogeneous model with uniform permeability, the reservoir system exhibits ideal flow behavior with sweep efficiency approaching 100%. History-matching results show high accuracy ($R^2 > 0.92$; MAPE < 7%), indicating that this model effectively captures the water-injection response. Comparison between the two CRM variants reveals significant differences, where CRM-IP produces stronger connectivity and more realistic time constants. Streamline visualization reinforces these findings by showing direct, unobstructed fluid flow from injector to producer, reflecting efficient, uniform reservoir conditions. In contrast to homogeneous conditions, the heterogeneous model demonstrates flow variations due to non-uniform permeability distribution. The presence of high-permeability zones causes early breakthrough and uneven fluid sweep. Nevertheless, the hybrid CRM+DCA model maintains accurate results with $R^2 = 0.88$ and MAPE = 1.46%. Connectivity and time constant values between CRM-P and CRM-IP are relatively

Analysis performance for well P1

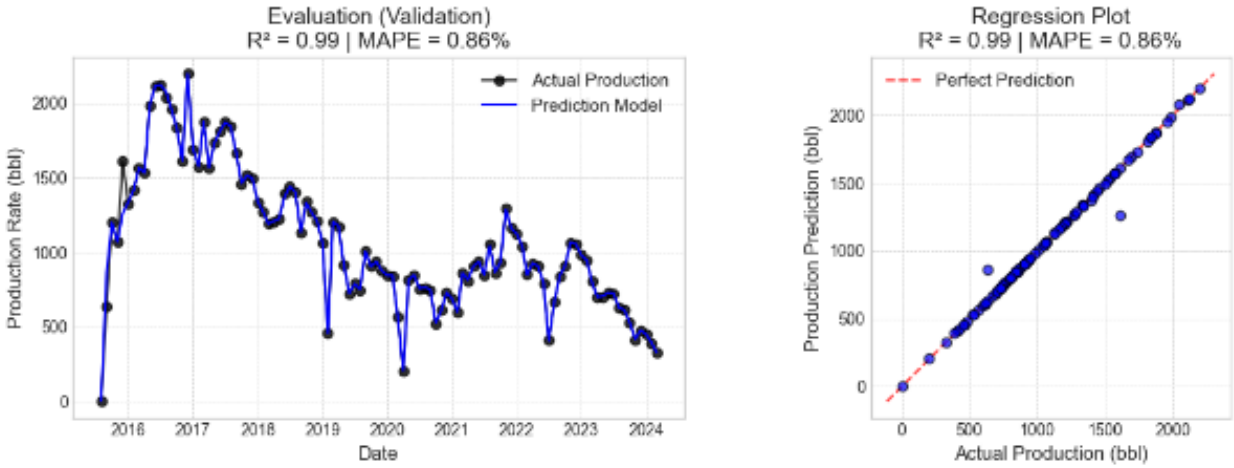


(a)

Analysis performance for well P2

(b)

Analysis performance for well P3



(c)

Figure 20. History matching & validation XGboost with regression plot well production (a) well 1 (b) well 2 (c) well 3

similar ($\tau \approx 30$ days), indicating that both models provide stable interpretation under heterogeneous conditions. Well P3 exhibits the highest connectivity, signaling a dominant flow path with potential for earlier water breakthrough. Three-dimensional visualization shows that injector I1 has the greatest influence on P3, consistent with streamline results where fluid flow concentrates in high-permeability zones. In Machine Learning analysis, the XGBoost model continues to demonstrate superior performance, with an MAPE of 4.98% compared to Random Forest's 10.07%, confirming its ability to handle highly variable data.

Meanwhile, the real field models the most complex reservoir conditions with fault structures, multi-zone layers, and extreme permeability variations. History-matching results show good agreement with the actual data ($R^2 = 0.76$; MAPE $\approx$ 20%), which remains acceptable for field-scale models of high complexity. CRM analysis demonstrates consistency between CRM-P and CRM-IP, with injector I1 having a dominant influence on producer P2 ($f = 0.0191$) and relatively uniform time constant values. This indicates local hydraulic stability despite non-homogeneous reservoir geometry. Streamline visualization confirms the presence of channeling effects in the northeast zone, driven by permeability differences and faulting that impede fluid flow. In production prediction, the XGBoost algorithm again demonstrates the best performance with $R^2 = 0.99$ and MAPE = 0.86%, compared to Random Forest, which achieves $R^2 = 0.95$.

Compared with the homogeneous model, the heterogeneous model provides superior performance, with rapid injection response and high sweep efficiency, because it is unaffected by formation heterogeneity. The heterogeneous model shows decreased sweep efficiency due to varying permeability distribution, but maintains stability in interwell connectivity parameters. Meanwhile, the real field Volve Field model depicts the most realistic conditions with combined influences of geology, inter-zone pressure, and complex permeability variations, yielding consistent connectivity and time constant values but lower sweep efficiency. Overall, comparative results indicate that CRM-IP is more representative and accurate in describing water injection dynamics across varying levels of reservoir complexity. At the same time, XGBoost is the most reliable algorithm for predicting production performance, with the lowest error rates across all scenarios.

CONCLUSION

This research demonstrates that integrating the Capacitance Resistance Model with Decline Curve Analysis constitutes an effective, data-driven approach for evaluating and predicting secondary recovery waterflood performance across various reservoir conditions. The CRM-DCA combination effectively addresses the limitations of conventional CRM in capturing nonlinear behavior in complex, heterogeneous reservoir systems. Analysis results indicate that the CRM-IP model provides more realistic estimates of interwell connectivity and time constants than CRM-P, particularly in homogeneous models with a response time of 98 days; in heterogeneous and real-field models, CRM-DCA delivers stable results with consistent connectivity values and response times across wells. Three-dimensional visualization and streamline analysis also strengthen the physical interpretation of fluid flow paths and sweep efficiency under each reservoir condition.

Furthermore, validation results using Machine Learning (Random Forest and XGBoost) demonstrate that CRM-DCA prediction accuracy falls within acceptable ranges and shows consistent history-matching patterns against actual data. The XGBoost model achieved the best benchmark results, with an R^2 of 0.99 and an MAPE below 1%, reinforcing the validity of the CRM-DCA approach for modeling waterflood performance. Overall, CRM-DCA integration proves capable of reducing waterflood analysis uncertainty, enhancing understanding of interwell connectivity, and providing a solid foundation for strategic decision-making in reservoir optimization and management during the secondary recovery stage.

GLOSSARY OF TERMS

Symbol	Definition	Unit
CRM	Capacitance Resistance Model	
DCA	Decline Curve Analysis	
f_{ij}	Coeficient Connectivity injection, producer	
MAPE	Mean Absolute Percentage Error	%
P1, P2, etc	Production Well 1,2, etc	
R^2	Coeficient Determination	

ACKNOWLEDGEMENT

The researchers would like to thank all members of my family, my paper team, the Petroleum Engineering Department, the Faculty of Engineering and Universitas Islam Riau, and PT. Pertamina Hulu Rokan.

REFERENCES

Alshboul, O., Shehadeh, A., Almasabha, G., & Almuflih, A. S. (2022). Extreme Gradient Boosting-Based Machine Learning Approach for Green Building Cost Prediction. Sustainability (Switzerland), 14(11). https://doi.org/10.3390/su14116651.

Chakraborty, D., & Elzarka, H. (2019). Advanced machine learning techniques for building performance simulation: a comparative analysis. Journal of Building Performance Simulation, 12 (2),193–207.

De-Prado-gil, J., Palencia, C., Jagadesh, P., & Martínez-García, R. (2022). A Comparison of Machine Learning Tools that Model the Splitting Tensile Strength of Self-Compacting Recycled Aggregate Concrete. Materials, 15 (12). https://doi.org/10.3390/ma15124164

De Holanda, R. W., Gildin, E., Jensen, J. L., Lake, L. W., & Shah Kabir, C. (2018). A state-of-the-art literature review on capacitance resistance models for reservoir characterization and performance forecasting. In Energies (Vol. 11, Issue 12). MDPI AG. https://doi.org/10.3390/en11123368

Du, X., Salasakar, S., & Thakur, G. (2024). A Comprehensive Summary of the Application of Machine Learning Techniques for CO2-Enhanced Oil Recovery Projects. Machine Learning and Knowledge Extraction, 6(2), 917–943. https://doi.org/10.3390/make6020043

Fadzil, M. A. M., Zabiri, H., Razali, A. A., Basar, J., & Syamzari Rafeen, M. (2021). Base oil process modelling using machine learning. Energies, 14(20). https://doi.org/10.3390/en14206527.

Fan, D., Lai, S., Sun, H., Yang, Y., Yang, C., Fan, N., & Wang, M. (2025). Review of Machine Learning Methods for Steady State Capacity and Transient Production Forecasting in Oil and Gas Reservoir. Energies, 18(4), 1–25. https://doi.org/10.3390/en18040842

Fu, L., Zhao, L., Chen, S., Xu, A., Ni, J., & Li, X. (2022). A Prediction Method for Development Indexes of Waterflooding Reservoirs Based on Modified Capacitance–Resistance Models. Energies, 15(18). https://doi.org/10.3390/en15186768.

Guo, Y., Zhang, L., Zhu, G., Yao, J., Sun, H., Song, W., Yang, Y., & Zhao, J. (2019). A pore-scale investigation of residual oil distributions and enhanced oil recovery methods. Energies, 12(19). https://doi.org/10.3390/en12193732

Ismailova, J. A., Delikesheva, D. N., Akhymbayeva, B. S., Logvinenko, A., & Narikov, K. A. (2021). Improvement of Sweep Efficiency in a Heterogeneous Reservoir. Smart Science, 9(1), 51–59. https://doi.org/10.1080/23080477.2021.1889259

Izadmehr, M., Daryasafar, A., Bakhshi, P., Tavakoli, R., & Ghayyem, M. A. (2018). Determining influence of different factors on production optimization by developing production scenarios. Journal of Petroleum

Exploration and Production Technology, 8(2), 505–520. https://doi.org/10.1007/s13202-017-0351-1.

Jiang, Y., Zhang, H., Zhang, K., Wang, J., Cui, S., Han, J., Zhang, L., & Yao, J. (2022). Reservoir Characterization and Productivity Forecast Based on Knowledge Interaction Neural Network. Mathematics, 10(9), 1–22. https://doi.org/10.3390/math10091614.

Krogstad, S., Lie, K. A., Nilsen, H. M., Berg, C. F., & Kippe, V. (2017). Efficient flow diagnostics proxies for polymer flooding. Computational Geosciences, 21(5–6), 1203–1218. https://doi.org/10.1007/s10596-017-9681-9.

Li, W., Dong, Z., Lee, J. W., Ma, X., & Qian, S. (2022). Development of Decline Curve Analysis Parameters for Tight Oil Wells Using a Machine Learning Algorithm. Geofluids, 2022. https://doi.org/10.1155/2022/8441075.

Liu, B., Xu, T., Xu, Y., Zhao, H., & Li, B. (2025). Automated Reservoir History Matching Framework: Integrating Graph Neural Networks, Transformer, and Optimization for Enhanced Interwell Connectivity Inversion. Processes, 13(5). https://doi.org/10.3390/pr13051386.

Liu, J. (2020). Potential for Evaluation of Interwell Connectivity under the Effect of Intraformational Bed in Reservoirs Utilizing Machine Learning Methods. Geofluids, 2020 (Cm). https://doi.org/10.1155/2020/1651549.

Makhotin, I., Orlov, D., & Koroteev, D. (2022). Machine Learning to Rate and Predict the Efficiency of Waterflooding for Oil Production. Energies, 15(3). https://doi.org/10.3390/en15031199.

Malvi´c, T., Ivšinovi´c, J., Veli´c, J., Sremac, J., & Barudžija, U. (2020). Increasing Efficiency of Field Water Re-Injection during Water-Flooding in Mature Hydrocarbon Reservoirs: A Case Study from the Sava Depression, Northern Croatia. Sustainability, c. https://doi.org/https://doi.org/10.3390/su12030786.

Maurenza, F., Yasutra, A., & Tungkup, I. L. (2023). Production Forecasting Using Arps Decline Curve Model with The Effect of Artificial Lift Installation. Scientific Contributions Oil and Gas, 46(1), 17–26. https://doi.org/10.29017/SCOG.46.1.1310

Moradi, S., Omar, A., Zhou, Z., Agostino, A., Gandomkar, Z., Bustamante, H., Power, K., Henderson, R., & Leslie, G. (2023). Forecasting and Optimizing Dual Media Filter Performance via Machine Learning. Water Research, 235 (March), 119874. https://doi.org/10.1016/j.watres.2023.119874.

Ng, C. S. W., Jahanbani Ghahfarokhi, A., & Nait Amar, M. (2022). Well production forecast in Volve field: Application of rigorous machine learning techniques and metaheuristic algorithm. Journal of Petroleum Science and Engineering, 208(PB), 109468. https://doi.org/10.1016/j.petrol.2021.109468

Nguyen, A. P. (2012). Capacitance Resistance Modeling for Primary.

Ogbeiwi, P., Aladeitan, Y., & Udebhulu, D. (2018). An approach to waterflood optimization: case study of the reservoir X. Journal of Petroleum Exploration and Production Technology, 8(1), 271–289. https://doi.org/10.1007/s13202-017-0368-5.

Pandey, A., Kesarwani, H., Saxena, A., Azin, R., & Sharma, S. (2023). Effect of heterogeneity and injection rates on the recovery of oil from conventional sand packs: A simulation approach. Petroleum Research, 8(1), 96–102. https://doi.org/10.1016/j.ptlrs.2022.05.005.

Pratama, H. B., & Saptadji, N. M. (2021). Study of Production-Injection Strategies for Sustainable Production in Geothermal Reservoir Two-Phase by Numerical Simulation. Indonesian Journal on Geoscience, 18(1), 25–38. https://doi.org/10.17014/ijog.8.1.25-38.

Rahmanifard, H., & Gates, I. (2024). A Comprehensive review of data-driven approaches for forecasting production from unconventional reservoirs: best practices and future directions. Artificial Intelligence Review, 57(8). https://doi.org/10.1007/s10462-024-10865-5.

Ramadhan, R., Novriansyah, A., Erfando, T., Tangparitkul, S., Daniati, A., Permadi, A. K., & Abdurrahman, M. (2023). Heterogeneity Effect on Polymer Injection: a Study of Sumatra Light Oil. Scientific Contributions Oil and Gas, 46(1), 39–52.https://doi.org/10.29017/SCOG.46.1.1334

Reginato, L. F., Gioria, R. dos S., & Sampaio, M. A. (2023). Hybrid Machine Learning for Modeling the Relative Permeability Changes in Carbonate Reservoirs under Engineered Water Injection. Energies, 16(13). https:// doi.org/10.3390/en16134849.

Ruhiat, D., & Effendi, A. (2018). Pengaruh Faktor Musiman Pada Pemodelan Deret Waktu Untuk Peramalan Debit Sungai Dengan Metode Sarima. TEOREMA : Teori Dan Riset Matematika, 2(2), 117. https://doi.org/10.25157/ teorema.v2i2.1075.

Sayarpour, M., Zuluaga, E., Kabir, C. S., & Lake, L. W. (2009). The use of capacitance-resistance models for rapid estimation of waterflood performance and optimization. Journal of Petroleum Science and Engineering, 69(3–4), 227–238. https://doi.org/10.1016/ j.petrol.2009.09.006.

Shahani, N. M., Zheng, X., Liu, C., Hassan, F. U., & Li, P. (2021). Developing an XGBoost Regression Model for Predicting Young's Modulus of Intact Sedimentary Rocks for the Stability of Surface and Subsurface Structures. Frontiers in Earth Science, 9(October), 1–13. https://doi.org/10.3389/feart.2021.761990.

Sidiq, H., Abdulsalam, V., & Nabaz, Z. (2019). Reservoir simulation study of enhanced oil recovery by sequential polymer flooding method. Advances in Geo-Energy Research, 3 (2), 115–121. https://doi.org/10.26804/ ager.2019.02.01.

Soltanmohammadi, R., Iraji, S., Rodrigues de Almeida, T., Basso, M., Ruidiaz Munoz, E., & Campane Vidal, A. (2024). Investigation of pore geometry influence on fluid flow in heterogeneous porous media: A pore-scale study. Energy Geoscience, 5(1), 100222. https:// doi.org/10.1016/j.engeos.2023.100222.

Sri Chandrahas, N., Choudhary, B. S., Vishnu Teja, M., Venkataramayya, M. S., & Krishna Prasad, N. S. R. (2022). XG Boost Algorithm to Simultaneous Prediction of Rock Fragmentation and Induced Ground Vibration Using Unique Blast Data. Applied Sciences (Switzerland), 12 (10). https://doi.org/10.3390/app12105269.

Unless, R., Act, P., Rose, W., If, T., & Rose, W. (2017). This is a repository copy of The Impact of Fine-scale Reservoir Geometries on Streamline Flow Patterns in Submarine Lobe Deposits Using Outcrop Analogues from the Karoo Version : Accepted Version Article : The Impact of Fine-scale Reservoir Geometries on.

Usman, & Haans, A. (2017). Mengoptimalkan Perolehan Minyak Pada Lahan Terbatas Menggunakan Sumur Berarah Dan Pendesakan Air. Lembaran Publikasi Minyak Dan Gas Bumi, 51(1), 1–12. https://journal.lemigas.esdm.go.id/ index.php/LPMGB/article/view/9.

Wang, G., Pickup, G. E., Sorbie, K. S., & Mackay, E. J. (2019). Analysis of Compositional Effects on Global Flow Regimes in CO2 Near-Miscible Displacements in Heterogeneous Systems. Transport in Porous Media, 129(3), 743–759. https://doi.org/10.1007/s11242-019-01304-z.

Weijermars, R., & van Harmelen, A. (2017). Advancement of sweep zones in waterflooding: conceptual insight based on flow visualizations of oil-withdrawal contours and waterflood time-of-flight contours using complex potentials. Journal of Petroleum Exploration and Production Technology, 7(3), 785–812. https:// doi.org/10.1007/s13202-016-0294-y.

Widodo, A. P., Sarwoko, E. A., & Firdaus, Z. (2017). Akurasi Model Prediksi Metode Backpropagation Menggunakan Kombinasi Hidden Neuron Dengan Alpha. Matematika, 20(2), 79–84.

Yousef, A. A., Gentil, P., Jensen, J. L., & Lake, L. W. (2006). A capacitance model to infer

interwell connectivity from production- and injection-rate fluctuations. SPE Reservoir Evaluation and Engineering, 9(6), 630–646. https://doi.org/10.2118/95322-PA

Zainuri, A. P. P., Sinurat, P. D., Irawan, D., & Sasongko, H. (2023). Trap Prevention in Machine Learning in Prediction of Petrophysical Parameters: A Case Study in The Field X. Scientific Contributions Oil and Gas, 46(3), 115–127. https://doi.org/10.29017/SCOG.46.3.1586.