

Optimizing CO₂-Enhanced Oil Recovery Through Machine Learning

Abstract

Producing more oil from an underground reservoir is more than a matter of increasing injection rates. Changes in permeability, pressure, and production conditions can significantly all have an impact on how the reservoir responds to CO₂ and water injection. This project uses machine learning and data science to look at these relationships and to better understand the impact of CO₂-EOR decisions on both oil production and economic profits.

I built a physics-based digital twin based on the SPE10 Model 2 reservoir benchmark and used it to generate 1,000 simulated reservoir scenarios. In each of the scenarios, I varied the permeability, CO₂ injection rate, water injection rate, production rate, and initial pressure. I then looked at the resulting change in oil production, reservoir pressure, and calculated profit, then trained a Random Forest model to see which conditions were most strongly associated with these outcomes.

The simulations show how computational modeling can be used to test a number of possible reservoir conditions without being bound to a single operating scenario. More generally, the project explores how machine learning can add onto reservoir modeling by taking the myriad of interactions between the variables and turning them into patterns. Future work will be used to include even a more heterogenous version and testing the model on even more datasets, other than just the SPE10 benchmark.

1. Background

1.1 Introduction

CO₂-enhanced oil recovery can serve the following benefits: increasing oil production and storing some of the injected CO₂ underground ([Yan et al., 2021](#); [Zhu et al., 2025](#)). These advantages can be even more efficient if we select injection strategies that balance hydrocarbon recovery with secure CO₂ storage ([Mao & Ghahfarokhi, 2026](#)). This study develops a physics-inspired computational framework for CO₂-EOR optimization under reservoir uncertainty and uses the SPE10 benchmark to evaluate the representativeness of its permeability assumptions.

1.2 Reservoir Uncertainty

Uncertainty in reservoir conditions and operational decisions makes it difficult to predict how different CO₂ injection strategies will perform. Geological heterogeneity can lead to different performance across different reservoirs, while pressure buildup and plume migration pose strong concerns. Therefore, evaluating injection strategies across varying permeability conditions is important for understanding how geological uncertainty can affect predicted production and economic outcomes ([Diplas et al., 2025](#)).

1.3 Machine Learning

Evaluating many different CO₂ injection strategies with high-fidelity simulations requires a lot of time and resources, limiting how extensively the possible options can be explored ([Yan et al., 2021](#); [Zhu et al., 2025](#)). These simulations use complex equations to explain how fluids move through the reservoir. They also require meticulous calculations across different conditions ([Jiang et al., 2023](#); [Nagao & Datta-Gupta, 2026](#)). Machine-learning surrogate models offer a way to accelerate performance by quickly predicting the performance, which allows optimization and uncertainty analysis to be performed with fewer trials of the original simulator ([Yan et al., 2021](#)).

This study uses a Random Forest surrogate model with reservoir dynamics to predict oil production from reservoir and injection variables, allowing a large number of candidate strategies to be evaluated efficiently ([Chen et al., 2025a, 2025b](#)).

1.4 Research Gap

Although surrogate models can reduce the computational burden of reservoir simulation, evaluating injection strategies under geological uncertainty still requires a framework that can efficiently compare outcomes across varying reservoir conditions while accounting for other constraints. Different operational constraints like pressure buildup and CO₂ plume movement should also be noted. Furthermore, it should consider a good balance between increasing oil recovery and storing CO₂ ([Diplas et al., 2025](#); [Mao & Ghahfarokhi, 2026](#)).

1.5 Objective & Research Questions

The primary objective of this study is to develop and evaluate a machine-learning based computational model for analyzing and optimizing CO₂ injection strategies under reservoir uncertainty. The framework combines a physics-inspired reservoir digital twin, Monte Carlo analysis, and permeability analysis.

The research is guided by three questions:

1. How accurately can a Random Forest model predict oil production from the simulation-generated reservoir and injection parameters?
2. How does uncertainty in permeability affect predicted oil production, pressure, and economics?
3. How representative is the range used in the Phase 2 analysis compared with the heterogeneous permeability distribution of the SPE10 benchmark?

2. Methodology

2.1 Research Framework

The model takes into account a variety of permeability, CO₂ injection rate, water injection rate, production rate, and initial pressure. The outcomes/results measured are oil production, pressure, and economic results.

The research follows this general process:

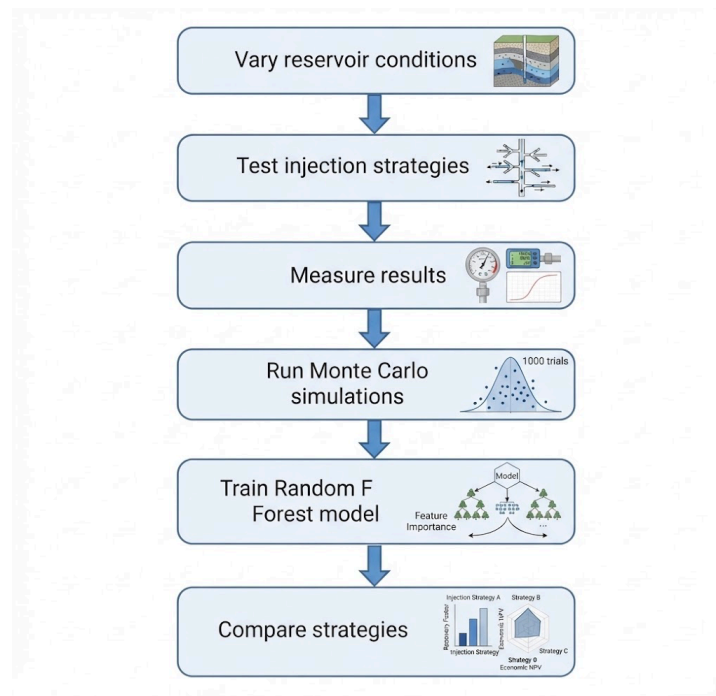


Figure 1. Proposed Architecture

2.2 Monte Carlo Analysis

After the baseline model of phase 1, in Phase 2, I introduced uncertainty by running 1,000 Monte Carlo simulations. Each simulation had different permeability, CO₂ injection rate, water injection rate, production rate, and initial pressure.

This allowed me to study how changes in reservoir conditions affected oil production, final pressure, and profit. Because these values were simulated, the results describe how the model behaves under the synthetic assumptions rather than the actual statistics and behavior in a real reservoir.

After generating the synthetic data, I used a Random Forest model to predict oil production from the input variables. This allowed me to compare different scenarios more efficiently.

2.3 SPE10 Comparison

In Phase 3, I compared the permeability range used in Phase 2 with the SPE10 Model 2 benchmark. SPE10 contains 1,122,000 (1.2 million) Kx permeability values and is a very heterogeneous reservoir.

To see how well our earlier 50–200 mD range captured our permeability, I analyzed the dataset by calculating its overall range, median, and the exact percentage of values that fell within the Phase 2 range and comparing it with the SPE10 permeability values.

3. Results

3.1 Monte Carlo Results

The 1,000 Monte Carlo simulations revealed significant differences in predicted oil production, final pressure, and calculated profit as the reservoir and operating conditions changed.

By looking closer at key inputs like permeability and production rate drive the results, it's evident that even minor shifts can lead to noticeably different outcomes and show the importance of considering uncertainty when comparing different injection strategies.

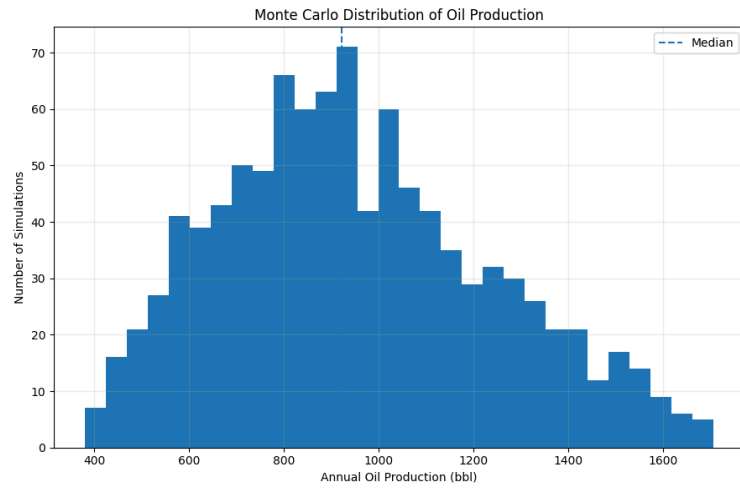


Figure 2. Monte Carlo Distribution of Predicted Oil Production

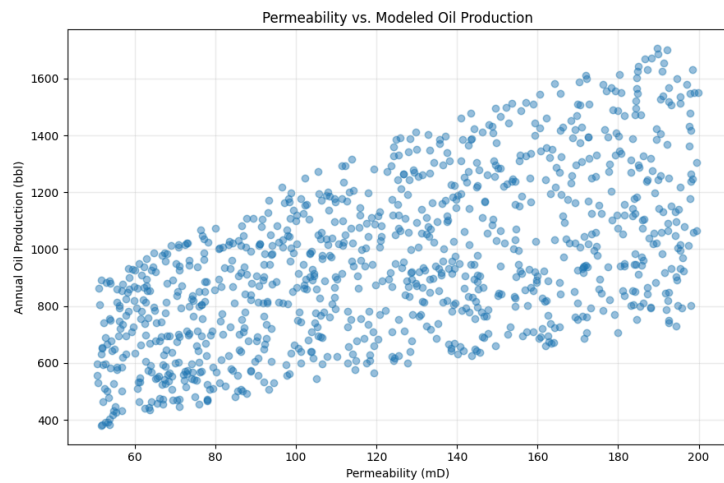


Figure 3. Permeability and Modeled Oil Production

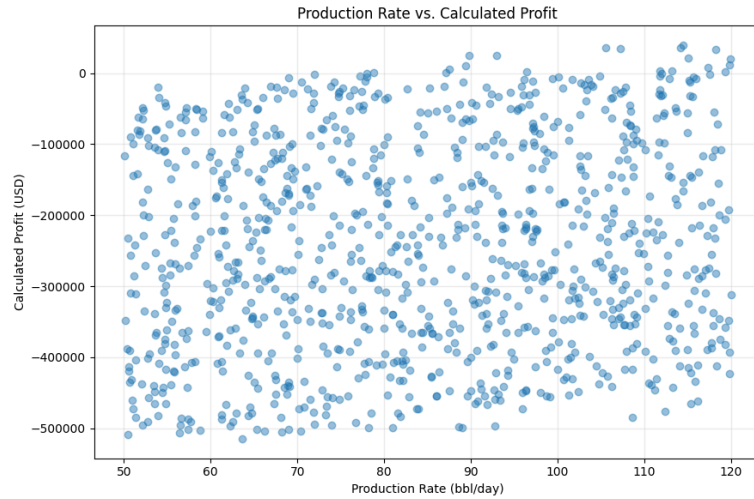


Figure 4. Production Rate and Profit

3.2 SPE10 Permeability Results

The SPE10 dataset contained 1,122,000 (1.2 million) Kx permeability values. The values ranged from approximately 0.001 to 20,000 mD with a median of 20.445 mD.

The Phase 2 model used a permeability range of 50–200 mD. When the SPE10 values were compared with this range, only 17.33% of the Kx values fell between 50 and 200 mD.

Therefore, approximately 82.67% of the SPE10 Kx values were outside the permeability range used in Phase 2.

This shows that the permeability range used in Phase 2 represents only a part of the broader heterogeneity found in the SPE10 benchmark.

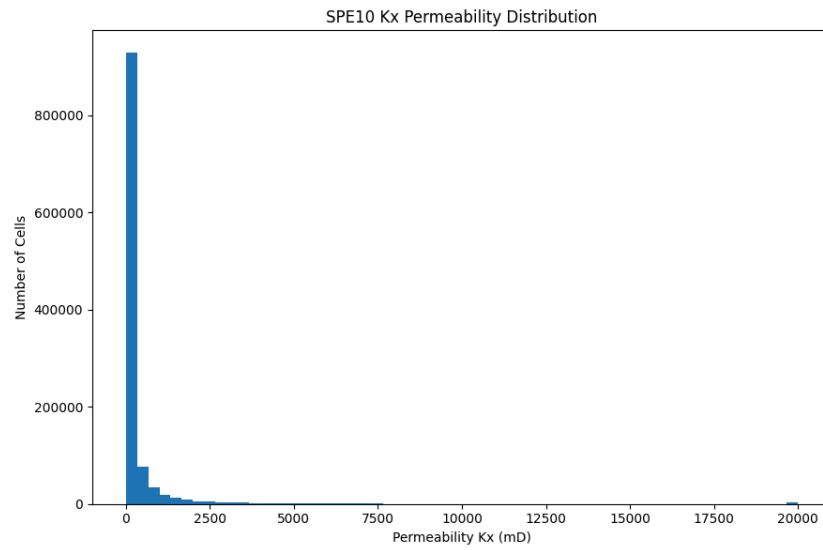


Figure 5. SPE10 K_x Permeability Distribution

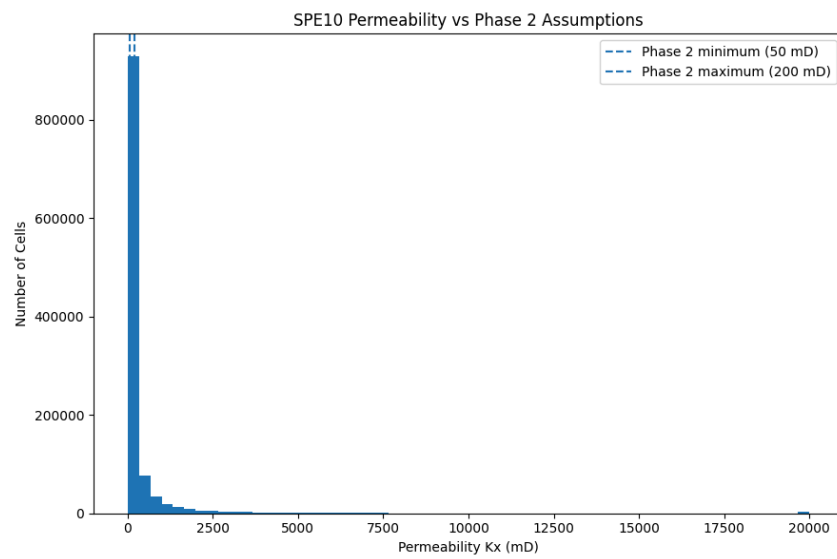


Figure 6. SPE10 Permeability vs. Phase 2 Assumptions

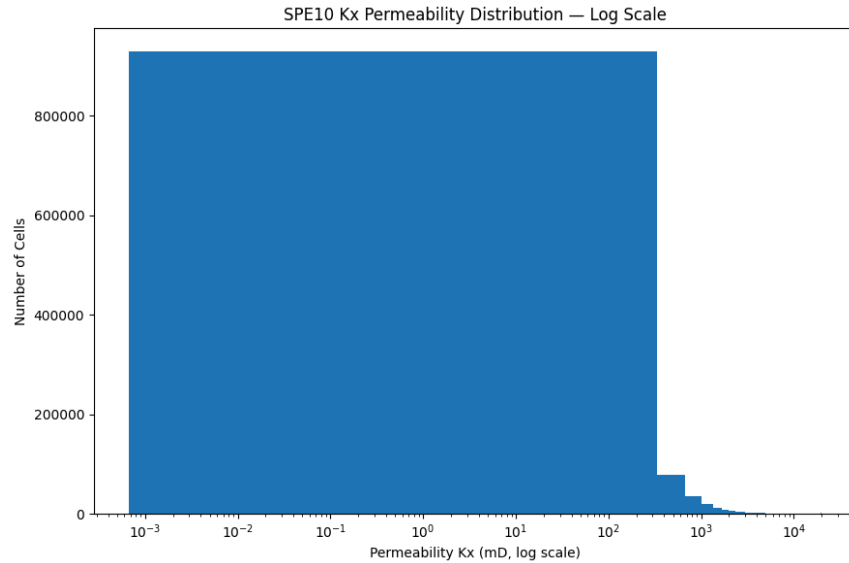


Figure 7. SPE10 Kx Probability Distribution

4. Discussion

The results show that reservoir uncertainty can have a substantial impact on the predicted performance of CO₂-EOR strategies. The 1,000 Monte Carlo simulations produced a range of outcomes in terms of oil production, pressure, and profit when the input conditions were changed. This suggests that it would be misleading to assess a strategy on the basis of a single set of reservoir conditions. ([Chen et al., 2022](#); [Wu et al., 2021](#)).

The Random Forest approach provides a means of evaluating the various conditions more efficiently. It can be trained on the simulation-generated data and used to predict the oil production that would result from different combinations of reservoir and injection parameters ([Wen et al., 2020](#)). This could help reduce the number of simulations needed when comparing many possible strategies.

The SPE10 comparison also highlighted a significant limitation of the Phase 2 model. Just 17.33% of the SPE10 Kx values fell within the 50–200 mD permeability range used in Phase 2. This means that the Phase 2 range represents only a portion of the heterogeneity found in the SPE10 benchmark.

This does not mean that the Phase 2 model is invalid. Instead, it shows that its results should be interpreted within the permeability range that was used in the study. The SPE10 comparison provides a reference point for understanding how the simplified assumptions differ from a more heterogeneous reservoir.

Overall, the results support the use of uncertainty analysis and machine learning as screening tools for CO₂-EOR strategies. However, future work would be to test the model out under a broader range of heterogeneous permeability conditions to see if its predictions remain reliable under more realistic reservoir conditions.

5. Limitations and Future Work

5.1 Limitations

One limitation of this study is that the Phase 2 analysis uses a very simplistic permeability range of 50-200 mD. Because of this, the SPE10 comparison showed that this range represents only a small portion of the permeability variation present in the benchmark.

Another limitation is that the Monte Carlo results depend on the simulated input ranges and assumptions. Therefore, the results reflect the behavior of the reservoir under those assumptions and will not in any way be a true representation of a real oil reservoir.

The Random Forest model is also reliant on the data it was trained on. If certain reservoir conditions are not included in the training data, its predictions and outputs under those situations are likely to be less reliable.

5.2 Future Work

This model could be improved by incorporating the permeability range that was used in the simulations to better reflect the heterogeneity found in SPE10. The model could also be tested using the distribution of SPE10 permeability rather than merely comparing its permeability values ([Rios, 2021](#)).

A further course of action would be to assess additional machine-learning models and compare their prediction accuracy with that of the Random Forest model.

Finally, future work could include CO₂ storage as an extra result so that injection strategies can be compared based on both oil recovery and the amount of CO₂ left underground ([Amporah, 2017](#)).

6. Conclusion

The objective of this research was to develop a computational model for evaluating CO₂-EOR strategies in the presence of reservoir uncertainty. The 1,000 Monte Carlo simulations showed

that changes in reservoir and operating conditions can affect predicted oil production, pressure, and profit.

The Random Forest model enabled the prediction of oil production from the simulation-generated data, and thus made it possible to evaluate different conditions more efficiently. Moreover, the comparison with the SPE10 benchmark showed that the 50-200 mD permeability range used in Phase 2 represents only a portion of the heterogeneity found in the benchmark.

Overall, the results show the importance of taking reservoir uncertainty into account when assessing different CO₂-EOR strategies. Future work using broader and more heterogeneous reservoir conditions could improve the model and lead to a more realistic evaluation of CO₂-EOR performance.

7. Works Cited

Ampomah, W., Balch, R.S, Cather, M., Will, R., Gunda, D., Dai, Z., & Soltanian, M.R. (2017).

Optimum design of CO₂ storage and oil recovery under geological uncertainty. *Applied Energy*, 195, 80–92. <https://doi.org/10.1016/j.apenergy.2017.03.017>

Chen, B., Mehana, M. Z., & Pawar, R. J. (2022). Quantitatively evaluating greenhouse gas

leakage from CO₂ enhanced oil recovery fields. *Advances in Geo-Energy Research*, 7(1), 20–27. <https://doi.org/10.46690/ager.2023.01.03>

Chen, Y., Quang Minh, N., Usman, A., Mustafa, O., & Omer, A. F. (2026). Application of embed-to-control observation deep-learning-based reservoir surrogate in CO₂ storage monitoring and multiobjective optimization. *SPE J.*, 31(02), 1119–1140.

<https://doi.org/10.2118/223914-PA>

de Souza Rios, V., Schiozer, D. J., dos Santos, Luiz O.S, & Arne, S. (2021). Improving

coarse-scale simulation models with a dual-porosity dual-permeability upscaling

technique and a near-well approach. *Journal of Petroleum Science and Engineering*, 198,

108132. <https://doi.org/10.1016/j.petrol.2020.108132>

- Diplas, D., Fotias, S. P., Ismail, I., Bellas, S., & Gaganis, V. (2026). A Fully Automated Design of Experiments-Based Method for Rapidly Screening Near-Optimal CO₂ Injection Strategies. *Energies*, 19(5), 1361. <https://doi.org/10.3390/en19051361>
- Mao, J., & Jahanbani Ghahfarokhi, A. (2026). Accelerating multi-well placement optimization with machine learning-based proxy models for CO₂ enhanced oil recovery and storage. *Applied Computing and Geosciences*, 29, 100317. <https://doi.org/10.1016/j.acags.2026.100317>
- Nagao, M., & Datta-Gupta, A. (2026). Physics-informed machine learning for identification of preferential flow paths and performance forecasting of subsurface carbon storage and utilization. *Geoenergy Science and Engineering*, 259, 214357. <https://doi.org/10.1016/j.geoen.2025.214357>
- Wen, G., Meng, T., & Benson, S. M. (2019). *Multiphase flow prediction with deep neural networks* (1910.09657V1). arXiv. <https://doi.org/10.1016/j.ijggc.2020.103223>
- Zhu, Y., Li, H., Zheng, Y., Li, C., Guo, C., & Wang, X. (2025). Ensemble Surrogates and NSGA-II with Active Learning for Multi-Objective Optimization of WAG Injection in CO₂-EOR. *Energies*, 18(24), 6575. <https://doi.org/10.3390/en18246575>