geoconvention
Calgary • Canada • May 11-13 2026

From Reservoir Properties to Predictions: An Integrated ML–Geoscience Approach to SAGD Forecasting

Andrew Pike, Jurgen Lehman*, Brent Rodgers*, Vanja Vracar**
**CNOOC International*

Summary

Robust SAGD greenfield forecasting demands more than just a predictive model – it requires a carefully engineered process that accurately links reservoir properties, well configuration, and steam injection to oil production. This work details an integrated ML–geoscience–reservoir engineering workflow founded on a rigorous backend data construction process.

At the core of this system is a structured data build that classifies wells by type, geometry, and drainage context, assigning reservoir conditions specifically at each well's on-production date. By automating the extraction of geoscience (petrophysics) and reservoir-engineering parameters, we create a consistent architecture that links steam input to observed oil response. This forms the foundation for training physically grounded ML models. We also address key integration challenges – including variable data quality, confounding features, and the risk of statistical patterns overshadowing physics – by outlining solutions such as automated spatial-to-well pipelines, recursive feature selection, and SHAP-based model interrogation. The result is a scalable, transparent system that delivers consistent, high-confidence greenfield forecasts.

Theory / Method / Workflow

We utilized a data-first approach to drive meaningful integration between geoscience, reservoir engineering, and machine learning. The workflow begins with a backend construction process that standardizes temporal and spatial consistency – a critical step for establishing reliable steam-oil relationships.

We employed geoscience-guided layouts to define realistic well configurations, automating vertical placement constrained by geological surfaces and reservoir quality indicators. Using planned well trajectories, we generated reservoir properties – including gridded geoscience models and injector-producer relationships – to construct an integrated dataset. Gradient Boosting (ML) models were then developed to predict EUR, peak oil rate, and cSOR. Our feature selection process combines domain knowledge with statistical screening to retain physically meaningful predictors while managing correlated variables. To ensure the models respect SAGD first principles, we interrogated model behaviour using explainable-AI diagnostics. Finally, we implemented regular lookbacks and post-prediction adjustments to track drift and allow for targeted modifications without requiring full model rebuilds.

Results, Observations, Conclusions

Our integrated workflow demonstrates that backend data construction, rather than algorithm complexity, is the primary enabler of successful ML-based SAGD forecasting. By explicitly tying reservoir properties and development design parameters to steam-oil response, we significantly improved model stability and predictive reliability.

Explainable-AI diagnostics confirm that key geological and reservoir-engineering drivers influence predictions in a manner consistent with SAGD physics, increasing confidence in forecast outcomes. This approach enables rapid, scalable greenfield evaluations while maintaining engineering oversight. By clearly defining model applicability and managing extrapolation risk, this workflow provides a robust foundation for development planning and decision-making.

Novel / Additive Information

This work highlights that the effectiveness of machine learning in SAGD forecasting depends less on model complexity and more on the quality, structure, and physical relevance of the underlying data. The explicit classification of wells and assignment of reservoir conditions at the on-production date provide a repeatable method for linking steam injection to oil response, enabling explainable, physically grounded ML predictions in greenfield developments.

Acknowledgements

The authors acknowledge the contributions of CNOOC International's geoscience, reservoir engineering, and operations teams, whose collaboration and domain expertise were essential to the development and validation of this workflow.

References

Reference Style (use Arial 9pt normal)