# Geologic Sweet Spot Identifier: A Machine Learning Approach to Optimal Drilling Locations

Aditya Holla, Ameera Aslam, Beshoy Shaker, Elijah Flores, Jayant Bhaskarun

# CONTENTS

**1.0 ABSTRACT** ..... **4**

**2.0 EXECUTIVE SUMMARY** ..... **5**

**3.0 INTRODUCTION** ..... **6**

3.1 CONTEXT ..... 6

3.2 PROBLEM ..... 7

3.3 RESPONSE ..... 7

**4.0 BACKGROUND** ..... **8**

4.1 PRIOR RESEARCH IN SWEET SPOT IDENTIFICATION ..... 8

4.2 MACHINE LEARNING APPLICATIONS IN RESERVOIR CHARACTERIZATION ..... 9

**5.0 METHODS** ..... **9**

5.1 DATA COLLECTION ..... 10

5.2 DATA CLEANING AND MUNGING ..... 10

*5.2.1 Missing and Zero Values* ..... *10*

*5.2.2 Outlier Detection and Transformation* ..... *10*

*5.2.3 Feature Selection and Redundancy Control* ..... *11*

5.3 FEATURE ENGINEERING ..... 11

*5.3.1 Variable Renaming and Encoding* ..... *11*

*5.3.2 Correlation and Multicollinearity Handling* ..... *11*

*5.3.3 Normalization* ..... *12*

*5.3.4 Exploratory Statistical Analysis* ..... *12*

5.4 ANALYSIS METHODS ..... 12

*5.4.1 Regression Modeling and Validation* ..... *12*

*5.4.2 Sweet Spot and Quality Index Formulation* ..... *13*

5.5 ALGORITHMS ..... 13

5.6 Spatial Modeling and Visualization ..... 13

**6.0 RESULTS ..... 14**

6.1 Sweet Spot Score Distribution ..... 14

6.2 Interpolated Map of Reservoir Productivity ..... 15

**7.0 DISCUSSION/INTERPOLATION ..... 15**

7.1 Geological Implication of High-Score Zones ..... 15

7.2 Reservoir Connectivity and Anisotropy ..... 16

7.3 Compartmentalization and Flow Barriers ..... 16

7.4 Model Limitations and Sources of Uncertainty ..... 17

7.5 Support for the Study Hypothesis ..... 17

7.6 Unexpected Results and Their Interpretation ..... 17

7.7 Unexpected Results and Their Interpretation ..... 18

7.8 Implications for Field Development and Decision-Making ..... 18

**8.0 CONCLUSION ..... 19**

**9.0 REFERENCED ..... 20**

**10.0 FIGURE CAPTIONS ..... 22**

# 1.0 ABSTRACT

Identifying optimal drilling locations (“sweet spots”) remains one of the most critical and uncertain challenges in oil and gas field development. Traditional approaches rely heavily on geological interpretation and seismic analysis, which, while valuable, are often subjective and difficult to scale across large spatial domains. Recent advances in data analytics and machine learning provide an opportunity to augment these traditional methods with objective, data-driven decision frameworks. This study addresses the problem of sweet-spot identification using a well-level dataset provided by ConocoPhillips, consisting of 55 wells and 14 petrophysical, production, and spatial variables.

Building on prior work in spatial reservoir characterization and machine-learning-based production forecasting, this project integrates exploratory data analysis, spatial interpolation, feature engineering, and non-linear predictive modeling. Missing data were handled using median imputation, and extreme water-production outliers were isolated to prevent bias in productivity measures. Spatial kriging was applied to visualize reservoir and production trends while explicitly accounting for interpolation uncertainty. A composite Quality Index was developed using the harmonic mean to generate a conservative, balanced ranking of wells based on porosity, permeability, pressure, and production metrics. Facies quality was modeled using Random Forest classification, while XGBoost was selected for sweet-spot prediction due to its ability to capture non-linear interactions and variable dependencies within the dataset.

The resulting Sweet Spot Prediction Map reveals spatial zones of elevated production potential and highlights pressure and rock quality as the dominant controls on productivity. Using peak-detection with spatial separation constraints, six high-confidence, non-overlapping drilling targets were identified. These results align with previous studies emphasizing the importance of integrated geological and data-driven workflows, while demonstrating the added value of uncertainty-aware spatial modeling.

Overall, this study demonstrates that combining spatial analytics, composite reservoir indices, and machine learning provides a robust, interpretable, and scalable framework for drilling optimization. The methodology offers practical implications for reducing exploration risk, improving capital allocation, and supporting data-informed decision-making in modern reservoir development.

# 2.0 EXECUTIVE SUMMARY

1. What is the problem?

Selecting optimal drilling locations in oil and gas fields is inherently uncertain and capital-intensive. Traditional decision-making relies heavily on geological interpretation and isolated production metrics, which can be subjective and difficult to scale spatially. This creates a risk of drilling in sub-optimal zones, increasing development costs and reducing recovery efficiency. ConocoPhillips sought a data-driven, spatially integrated workflow to improve confidence in sweet-spot identification using existing well data.

2. What was your approach?

- We developed an integrated geology–machine learning–spatial analytics workflow using well-level petrophysical, production, and spatial data. Our process included:
- Cleaning and exploring the raw well data to understand spatial and production behavior.
- Applying spatial kriging to visualize reservoir trends with uncertainty awareness.
- Constructing a composite Quality Index using the harmonic mean to create a balanced well-performance ranking.
- Modeling facies quality with Random Forest.
- Using XGBoost to capture non-linear relationships and generate a Sweet Spot Prediction surface.
- Applying peak-detection algorithms to identify high-confidence drilling targets while enforcing spatial separation.

3. What did you learn?

- Pressure and rock quality (porosity and permeability) are the strongest drivers of well productivity.

- Porosity and permeability are largely redundant with facies, confirming the importance of geological context.
- Machine learning improves consistency in ranking wells but must be constrained by physical geology.
- Spatial uncertainty increases near field boundaries, emphasizing the need for risk-aware interpretation.
- Sweet-spot patterns form consistent spatial clusters rather than isolated single-point highs.

4. What are your recommendations?

- Prioritize drilling in high-pressure, high-rock-quality zones identified by the Sweet Spot Map.
- Focus on the six highest-confidence drilling locations identified through peak-detection.
- Avoid over-reliance on kriging without raw well control—always interpret with observed data.
- Use the Quality Index for relative ranking, not absolute production forecasting.
- Expand well density in low-confidence areas to improve future spatial accuracy.
- Incorporate depletion-corrected pressure in future modeling to improve long-term reliability.

# 3.0 INTRODUCTION

## *3.1 Context*

Oil and gas exploration plays a critical role in meeting global energy demands, yet it remains one of the most uncertain and capital-intensive stages of field development. Drilling decisions are traditionally guided by geological interpretation, seismic data, and expert judgment. While these methods provide valuable subsurface insight, they rely on large volumes of complex, heterogeneous data and often require subjective integration across multiple disciplines. As reservoirs become more heterogeneous and development costs rise, the industry increasingly requires objective, scalable, and data-driven tools to support drilling decisions.

Recent advances in data science and machine learning have enabled more systematic analysis of geological, petrophysical, and production data. Prior studies show that reservoir properties such as porosity, permeability, depth, facies, and pressure are strong indicators of well productivity, and that machine learning models can outperform traditional statistical approaches in predicting production behavior (Ibrahim et al., 2022; Du et al., 2024). More recently, machine learning has also been applied directly to spatial sweet spot identification, demonstrating its potential to predict high-productivity zones using integrated reservoir and completion data (Machine Learning Sweet Spot Identification…, 2023). These advances highlight the growing role of analytics in modern reservoir evaluation and field development.

## *3.2 Problem*

Despite progress in predictive modeling, most existing studies focus primarily on well-level production prediction rather than direct spatial identification of optimal drilling locations. Traditional sweet spot identification often depends on isolated geological indicators or seismic attributes without fully integrating geological quality, production behavior, spatial variability, and model uncertainty into a unified framework. Additionally, subsurface data commonly exhibit strong variable correlation, spatial sparsity, pressure depletion effects, and extreme production outliers, all of which challenge reliable modeling and interpretation. As a result, there remains a critical need for a transparent, interpretable, and spatially explicit workflow that links machine learning predictions to physical reservoir controls in order to support high-confidence drilling decisions.

## *3.3 Response*

This study addresses this gap by developing an integrated, data-driven sweet spot identification workflow using a well-level dataset provided by ConocoPhillips, consisting of 55 wells and 14 petrophysical, production, and spatial variables. The workflow combines exploratory data analysis, feature engineering, machine learning, and spatial visualization to quantify reservoir quality and drilling potential. A composite Quality Index based on the harmonic mean was developed to provide a conservative, balanced representation of reservoir performance. Ensemble machine learning models were applied to capture non-linear geological–production

relationships and generate a Sweet Spot Prediction Score. Spatial kriging was then used as a visualization tool to interpolate model outputs across the field while preserving actual well control and uncertainty awareness.

The resulting framework directly translates raw well data into actionable drilling guidance while maintaining physical interpretability and risk awareness. This approach supports data-informed drilling decisions, reduces the reliance on subjective interpretation, and contributes to the growing application of machine learning in reservoir characterization and field development planning.

# 4.0 BACKGROUND

## *4.1 Prior Research in Sweet Spot Identification*

Previous research in petroleum engineering has extensively examined the geological and petrophysical factors that control hydrocarbon productivity. Rock properties such as porosity, permeability, facies, and total depth are widely recognized as key indicators of reservoir quality and fluid flow potential. Early efforts to identify productive zones relied primarily on statistical regression and petrophysical interpretation; however, these traditional approaches were often limited by subsurface heterogeneity and complex nonlinear relationships.

Ibrahim et al. (2022) demonstrated that predictive modeling can effectively capture production behavior across diverse formations using long-term regression analysis of oil, gas, and water outputs. Similarly, Du et al. (2024) applied deep learning techniques to forecast oil production in U.S. reservoirs, achieving improved predictive performance compared to conventional empirical models. While these studies provided valuable insight into production behavior at existing wells, they primarily focused on forecasting outputs rather than identifying new drilling locations.

More recent studies have begun shifting toward integrating petrophysical and spatial parameters to locate areas of enhanced reservoir potential, commonly referred to as “sweet spots.” However, many existing models still evaluate wells individually and do not fully incorporate coordinate-based spatial relationships that reveal geographic clusters of productivity. This limitation reduces

their usefulness for field-scale exploration planning. The present study builds upon this prior work by integrating geological, production, and spatial features into a unified framework for mapping zones of high production potential.

## *4.2 Machine Learning Applications in Reservoir Characterization*

Machine learning has become an increasingly important tool in reservoir characterization due to its ability to model nonlinear relationships and process large, multidimensional datasets. Techniques such as regression, random forests, and neural networks have been widely used to predict production trends, estimate reservoir properties, and automate geologic interpretation. These methods often outperform traditional statistical techniques, particularly when data are incomplete or noisy.

Lavi et al. (2024) introduced data-driven imputation strategies to address missing production data, demonstrating that machine learning can preserve data integrity while improving predictive accuracy. Li et al. (2024) developed a weighted productivity framework incorporating permeability, porosity, and depth, establishing permeability and porosity as dominant controls on reservoir performance. Their hierarchy of control factors provides a foundation for the multi-variable scoring approach adopted in this study.

Recent advances increasingly combine machine learning with spatial analysis, allowing models to move beyond isolated well prediction toward field-scale reservoir visualization. By integrating spatial coordinates with petrophysical and production variables, predictive models can identify geographic trends in productivity. This study advances this direction by applying machine learning to quantify relationships among key variables and visualizing these patterns through a three-dimensional Sweet Spot Map.

# 5.0 METHODS

This section describes the data sources, preprocessing steps, feature engineering procedures, analytical methods, machine learning algorithms, and spatial interpolation techniques used to identify high-potential drilling locations. All analyses are conducted in Python to ensure full reproducibility.

## *5.1 Data Collection*

The dataset is provided by ConocoPhillips and consists of measurements from 55 wells. The variables include petrophysical properties (porosity, permeability in x and y directions, facies, and total measured depth), production metrics (oil production, gas production, water production, and reservoir pressure), and spatial information (bottom-hole x and y coordinates). These data form the basis for all subsequent statistical, machine learning, and geospatial analyses.

## *5.2 Data Cleaning and Munging*

### 5.2.1 Missing and Zero Values

The dataset is first inspected to identify missing and zero values. Five wells (Wells 8, 9, 27, 28, and 47) contain zero entries in oil and gas production columns. Prior to imputation, summary statistics and distribution plots (including boxplots) are generated for all production variables to verify that zero values reflect valid operational conditions rather than data-entry errors. These visual diagnostics reveal strong skewness and outlier sensitivity, motivating the use of median-based imputation.

Following Lavi et al. (2024), missing or zero entries are treated as operational artifacts such as temporary well shutdowns or sensor errors rather than random omissions. Median imputation is applied because it is robust to extreme values and preserves the empirical distribution of the data while maintaining the small sample size.

### 5.2.2 Outlier Detection and Transformation

Outliers are detected using the interquartile range (IQR) method. Water production displays strong positive skewness with several extreme values. To stabilize variance while preserving geological meaning, a MinMax scaling was applied to all numerical variables.

Outlier behavior is also assessed separately within each facies group to identify facies-specific anomalies in both production and petrophysical variables. Facies-stratified boxplots are

generated to determine whether transformations should be applied uniformly across the dataset or selectively by geological classification, ensuring consistent scaling across all rock types.

### 5.2.3 Feature Selection and Redundancy Control

Non-predictive identifiers such as well name and well number are removed. The future pressure prediction variable is excluded to prevent data leakage. Statistical hypothesis testing is used to evaluate feature significance; variables with limited predictive power, such as future_pressure, were evaluated for removal when appropriate.

To capture total energy output rather than treating oil and gas separately, a “Predicted Production” metric was created by combining normalized oil and gas production. This composite variable represents unified energy production and accounts for wells that are predominantly oil- or gas-driven. The metric is used as an alternative exploratory target variable to evaluate unified productivity behavior during preliminary modeling.

## *5.3 Feature Engineering*

### 5.3.1 Variable Renaming and Encoding

All variables are standardized using snake-case naming conventions to ensure computational consistency. The categorical facies variable is one-hot encoded into binary features for integration into regression and machine learning models.

### 5.3.2 Correlation and Multicollinearity Handling

Pearson correlation analysis identifies strong dependencies between permeability and porosity, and between depth and pressure, indicating potential multicollinearity. To mitigate instability caused by correlated predictors, quality scores were created as a way of testing data with more than one variable in a single metric.

### 5.3.3 Normalization

Continuous variables are normalized using min–max scaling to the [0,1] interval:

$$X_{norm} = \frac{X - Xmin}{Xmax - Xmin}$$

This ensures all parameters contribute proportionally regardless of native measurement units. Categorical and dimensionless parameters remain in encoded format for interpretability.

### 5.3.4 Exploratory Statistical Analysis

As part of the exploratory workflow, descriptive statistics and boxplots are generated for all key variables including porosity, permeability, oil, gas, and water production. Wells are grouped by facies and facies-stratified production plots are constructed for oil, gas, water, and derived ratios. These procedures document geological variability patterns prior to modeling and support later physical interpretation of machine learning outputs.

## *5.4 Analysis Methods*

### 5.4.1 Regression Modeling and Validation

Rather than focusing on linear regression as our initial thought process entailed, we modeled production using an XGBoost regressor trained on spatial coordinates, depth, porosity, harmonic-mean permeability, past pressure, and an engineered facies-quality index. This allows the model to capture threshold effects and interactions which linear regressions can't explain or show.

We also used Kriging to interpolate depth and other continuous properties onto a regular grid. The XGBoost-based Sweet Spot Score was then evaluated on this grid and combined with kriged fields to generate continuous maps, on which spatial clustering, anisotropy, and connectivity

were interpreted. Minimum-distance filters relative to existing wells and peak-finding algorithms were also applied to identify candidate infill locations.

### 5.4.2 Sweet Spot and Quality Index Formulation

An initial composite Sweetspot Score is explored using XGBoost feature importance weightings. It ends up prioritizing pressure and depth as dominant controls. The final Sweetspot Score is a combination of pressure, depth, permeability, and porosity to create a variable for the synthetic wells created called “Predicted Production.”To further isolate geological quality from production noise, a Quality Index (QI) is constructed using harmonic aggregation of porosity, permeability, pressure, and GOR. A Facies Quality Score (FQS) is also computed to isolate lithologic influence independent of depth and production. These indices serve as inputs to subsequent spatial prediction.

## *5.5 Algorithms*

- Random Forest Regressor for feature importance and exploratory imputation
- Ordinary Kriging for spatial interpolation
- XGBoost (final model) for Sweet Spot prediction

XGBoost is selected due to its ability to capture nonlinear feature interactions, robustness to correlated predictors, and superior performance under small-sample conditions.

## *5.6 Spatial Modeling and Visualization*

The Sweetspot Scores are visualized using bottom-hole coordinates (bh_x, bh_y) and total depth. To transform discrete well measurements into continuous fields, Ordinary Kriging is applied using the PyKrige library.

Kriging estimates values at unsampled locations using weighted averages of known points while modeling spatial autocorrelation through a variogram. A spherical variogram model is automatically fit using least-squares optimization. A 150 × 150 spatial grid is defined across the

field bounds, and kriging is applied independently to permeability, porosity, depth, pressure, and Sweetspot Score.

KNN interpolation was evaluated as a computational benchmark but was rejected due to its lack of uncertainty quantification and poor spatial continuity. Kriging remains the primary interpolation method due to its optimal statistical properties and uncertainty estimation.

Sweet Spot surfaces were visualized using the Kriging mapping package with depth contours overlaid. These maps reveal spatial clustering of high-scoring wells and allow identification of prospective drilling targets between existing wells based on the warmness of the color shown.

Prediction uncertainty is largest near field boundaries outside the convex hull of wells and lowest near dense well clusters.

# 6.0 RESULTS

## *6.1 Sweet Spot Score Distribution*

An interpolation of the Sweetspot Score was created to visualize the spatial variation of reservoir productivity across the study area (Figure 13). The Sweetspot Score, calculated as a weighted linear combination of permeability, porosity, production, and depth parameters, provides a unified measure of reservoir quality and flow potential. The imputed weighting scheme was calculated from the XGBoost model, which identifies pressure and depth as dominant controls on well oil and gas output.

The resulting Sweetspot Score distribution differentiates high- and low-productivity wells across the dataset. Higher scores generally occur within intermediate depth intervals and regions exhibiting stronger permeability and porosity values. In contrast, lower scores are concentrated in deeper or low-permeability zones where compaction and diagenesis likely reduce pore connectivity.

Overall, the Sweetspot Score successfully integrates petrophysical and production variables into a single interpretable metric that reflects the underlying physical properties governing hydrocarbon flow.

## *6.2 Interpolated Map of Reservoir Productivity*

The Kriging Sweetspot Map (Figure 13) combines borehole spatial coordinates (`bh_x`, `bh_y`) with total measured depth to visualize the Sweetspot Score as a color mapped surface. Linear interpolation converts discrete well data into a continuous surface, revealing both lateral and vertical continuity in reservoir productivity potential.
Warmer colors (orange to red) represent higher Sweetspot Scores and indicate zones where permeability and porosity combine with favorable production metrics, suggesting enhanced hydrocarbon flow potential. Cooler colors (yellow) correspond to lower-quality zones with reduced pressure, depth, permeability, or porosity.

Several high-score clusters are visible, laterally connected at intermediate depths. These areas likely correspond to depositional or structural features such as channel sands or high permeability stratigraphic units that promote greater fluid connectivity. Shallower and mid-depth regions consistently exhibit stronger scores than deeper sections, where compactional effects and diagenetic cementation may diminish permeability.

The resulting visualization highlights spatially coherent regions of high reservoir potential and provides a foundation for interpreting geological trends that control productivity.

# 7.0 DISCUSSION/INTERPOLATION

## *7.1 Geological Implication of High-Score Zones*

The Sweet Spot Score, built from the XGBoost production model and the facies-probability–based facies quality index, highlights coherent high-score clusters that coincide with wells in higher-porosity, higher-permeability facies. These zones are concentrated in structurally

lower areas where kriged depth, facies quality, and historical production all agree, suggesting thicker, less compacted reservoir intervals with better preserved pore networks. Laterally continuous high-score trends follow the mapped facies architecture rather than appearing as isolated artifacts, which is consistent with channel-like or sheet-sand bodies that maintain reservoir quality across multiple locations. This behavior indicates that the Sweet Spot Score is capturing integrated rock and fluid-flow behavior rather than simply mirroring a single input variable.

## *7.2 Reservoir Connectivity and Anisotropy*

Since the score uses harmonic-mean permeability ($perm_{hm}$) and a facies classifier trained on $perm_x$, $perm_y$, porosity, depth, and pressure, the sweet-spot map more clearly expresses directional permeability trends. Elongated high-score corridors align with the dominant permeability direction inferred from $perm_x/perm_y$ and from the facies map, implying preferential flow pathways along depositional fabrics or paleoflow directions. These anisotropic corridors link multiple high-quality wells and coincide with smoother pressure gradients in the XGBoost predictions, supporting an interpretation of laterally connected flow units rather than isolated "good" wells.

## *7.3 Compartmentalization and Flow Barriers*

Where high Sweet Spot Scores terminate abruptly against low-score regions, the updated maps often coincide with facies transitions, rapid depth changes, or low-quality facies identified by the facies-quality model. These boundaries are interpreted as potential intra-reservoir baffles or compartments (shaly intervals, tight facies, or stratigraphic pinch-out) that limit lateral pressure communication. In contrast, areas where high-score clusters remain contiguous across several grid cells and wells are likely to represent more continuous reservoir bodies with better connectivity. The combination of raw-data facies maps, kriged depth contours, and the sweet-spot surface makes it easier to distinguish structural versus stratigraphic controls on connectivity and to identify where additional wells would most efficiently connect to existing high-quality flow units.

## *7.4 Model Limitations and Sources of Uncertainty*

Despite the methodological improvements such as explicit facies-quality modeling, separation of training and validation data, and clearer feature engineering, the workflow remains constrained by the limited dataset (55 wells) and the strong correlations among key predictors (facies, porosity, permeability, and pressure). The XGBoost feature-importance and facies-model analyses show that rock-quality and pressure effects are intertwined; as a result, statements about "dominant" drivers must be interpreted as predictive, not strictly causal. Spatial interpolation via kriging still introduces uncertainty away from well control, especially near the model edges where the sweet-spot surface is extrapolated. In addition, water production and some pressure behaviors remain noisy and occasionally counterintuitive, indicating that more targeted EDA and perhaps separate water-handling models would be beneficial. Despite these constraints, the overall workflow demonstrates that even small datasets can yield physically interpretable insights when combined with transparent modeling and sound geological reasoning.

## *7.5 Support for the Study Hypothesis*

The central hypothesis of this study is that integrating geological, petrophysical, production, and spatial features within a unified machine-learning and geostatistical framework can improve the identification of high-potential drilling zones. The results support this hypothesis. High Sweet Spot Scores consistently align with wells located in high-porosity, high-permeability facies and with zones of stronger historical production. The spatial clustering of predicted sweet spots further confirms that the model is capturing coherent geological patterns rather than random production variability. These outcomes demonstrate that the integrated workflow successfully translates multi-source data into a physically interpretable prediction of reservoir potential.

## *7.6 Unexpected Results and Their Interpretation*

Several unexpected behaviors were observed during analysis. In some areas, elevated Sweet Spot Scores occurred where historical oil production was moderate rather than extreme. This behavior

is attributed to the model's emphasis on balanced reservoir quality through the harmonic mean–based Quality Index, which penalizes single-parameter dominance. Additionally, localized high water production occasionally overlapped high permeability zones, reducing Sweet Spot Scores in regions that might otherwise appear attractive based solely on rock quality. These outcomes highlight the importance of incorporating multiple production indicators and demonstrate that high permeability alone does not guarantee favorable oil productivity.

## *7.7 Unexpected Results and Their Interpretation*

Consistent with prior studies by Ibrahim et al. (2022), Du et al. (2024), and Li et al. (2024), this project confirms that permeability and porosity remain the dominant controls on reservoir productivity. However, unlike earlier studies that primarily focused on well-level prediction, this work advances the literature by extending machine-learning predictions into a spatially continuous sweet spot map using ordinary kriging. While previous studies emphasized production forecasting, the present framework explicitly targets **geographic drilling optimization**, bridging a key gap between predictive modeling and practical field development planning.

## *7.8 Implications for Field Development and Decision-Making*

The spatial Sweet Spot framework developed in this study provides a practical tool for guiding drilling decisions in data-limited reservoirs. By ranking potential drilling zones based on integrated rock quality, pressure behavior, and production indicators, the approach supports more efficient well placement and reduces the risk associated with exploratory drilling. The methodology is directly transferable to other fields with limited well control and can be extended to include seismic attributes, time-dependent production, and geomechanical constraints. Ultimately, this workflow demonstrates how data-driven modeling can complement traditional geological interpretation to improve reservoir management and capital allocation.

# 8.0 CONCLUSION

This study develops an integrated geologic–machine learning workflow to identify production-driven sweet spots by combining spatial interpolation, facies analysis, and gradient-boosted prediction. The central hypothesis of this work is that sweet spots can be reliably identified using a unified framework that integrates rock quality, pressure, and spatial context rather than relying on production alone. The results support this hypothesis.

Kriging maps of oil, gas, water, porosity, permeability, pressure, gas–oil ratio, facies, facies quality, and the final Sweet Spot Score reveal a structurally coherent reservoir in which production highs systematically track structural position and associated rock-quality trends rather than appearing as isolated anomalies. Facies-conditioned boxplots demonstrate that Facies 1 and 2 consistently exhibit the highest porosity and permeability, confirming their role as the primary reservoir facies and establishing a physical foundation for the engineered machine-learning features.

By explicitly addressing the strong correlation between facies, porosity, and permeability, the workflow uses a facies quality score and a harmonic-mean-based quality index to compress redundant petrophysical information into stable, outlier-robust predictors. After this correction, XGBoost feature importance and kriged production surfaces show that pressure and depth, modulated by rock quality, emerge as the dominant predictive controls on the Sweet Spot surface, while extreme water producers are deliberately excluded to prevent optimization bias. Facies-specific uncertainty analysis further reveals elevated classification risk within Facies 3, identifying zones where additional geological control is needed before development decisions.

The final Sweet Spot Map, combined with peak local maxima detection and spatial spacing constraints, yields six spatially independent drilling candidates located within the structurally aligned “sweet-spot ring.” These candidates represent relative, not absolute, production forecasts, but they are supported by mutually consistent trends across pressure, facies, petrophysical properties, and historical production, providing a defensible basis for ranked drilling decisions within the convex hull of existing wells.

Overall, this work demonstrates that when kriging-based spatial context, facies-aware feature engineering, and nonlinear machine-learning models such as XGBoost are carefully constrained by geology, artificial intelligence can move beyond black-box prediction to deliver interpretable, uncertainty-aware sweet-spot recommendations for field development planning.

Future work should include:

1. Expanding the dataset with additional wells or time-lapse production to improve generalization and uncertainty control.
2. Integrating seismic or log-derived attributes to strengthen spatial and stratigraphic resolution.

# 9.0 REFERENCED


**1. Du, J., Zheng, J., Liang, Y., & Shahzad, K. 2024.**
A deep learning-based approach for predicting oil production: A case study in the United States. ***Energy*** 288: 129688.
https://doi.org/10.1016/j.energy.2023.129688

**2. Ibrahim, N. M., Alharbi, A. A., Alzahrani, T. A., Abdulkarim, A. M., Alessa, I. A., Hameed, A. M., Albabtain, A. S., Alqahtani, D. A., Alsawwaf, M. K., & Almuqhim, A. A. 2022.**
Well-performance classification and prediction: Deep learning and machine-learning long-term regression experiments on oil, gas, and water production. ***Sensors*** 22(14): 5326.
https://doi.org/10.3390/s22145326

**3. Li, J., Chen, Z., & Zhang, M., 2024.** "Quantitative Evaluation of Main Control Factors Affecting Well Productivity in Tight Sandstone Reservoirs." **Processes**, 12(9): 1890. https://doi.org/10.3390/pr12091890

**4. Lavi, B., Bertini Jr., J.R., Pires, L.C.O., and Schiozer, D.J. 2024.**
Handling Missing Data on Short-term Oil Rate Forecasting in a Pre-salt Offshore Field Benchmark. ***ROG.e 2024: Technical Paper***, 1–11.
https://doi.org/10.48072/2525-7579.roge.2024.3726

**5. Leem, J., & Chen, H. 2023.**
Machine Learning Sweet Spot Identification and Performance Validation Utilising Reservoir and Completion Data from Unconventional Reservoir in British Columbia, Canada. ***SPE-215220-MS***.
https://www.researchgate.net/publication/374503924

**6. Meyer, R., and Krause, F. F. 2006.**
Permeability anisotropy and heterogeneity of a sandstone reservoir analogue: An

estuarine to shoreface depositional system in the Virgelle Member, Milk River Formation, Writing-on-Stone Provincial Park, Southern Alberta. ***Bulletin of Canadian Petroleum Geology*** 54(4): 301–318. https://doi.org/10.2113/gscpgbull.54.4.301

**7. ScienceDirect Topics. 2025.** Reservoir Compartmentalization—An Overview. ***Earth and Planetary Sciences***. https://www.sciencedirect.com/topics/earth-and-planetary-sciences/reservoir-compartmentalization

# 10.0 FIGURE CAPTIONS

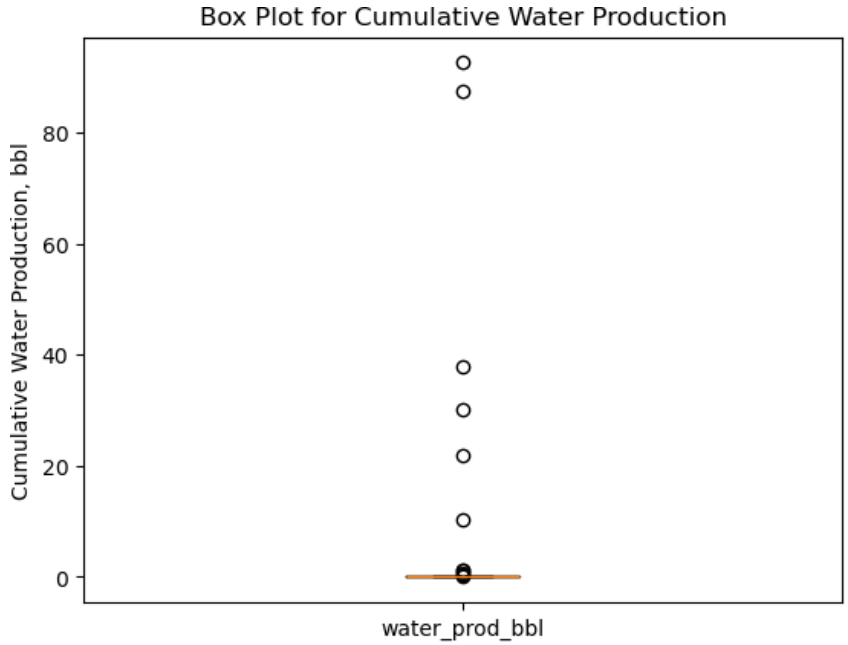



Figure 1: Box Plot for Water Production

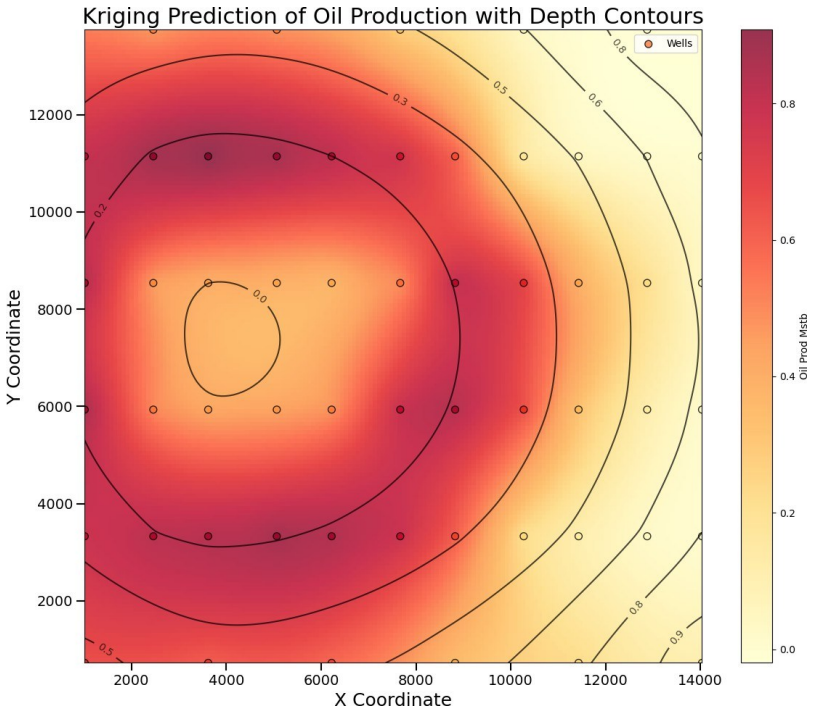



Figure 2: Kriging Map of Oil Production

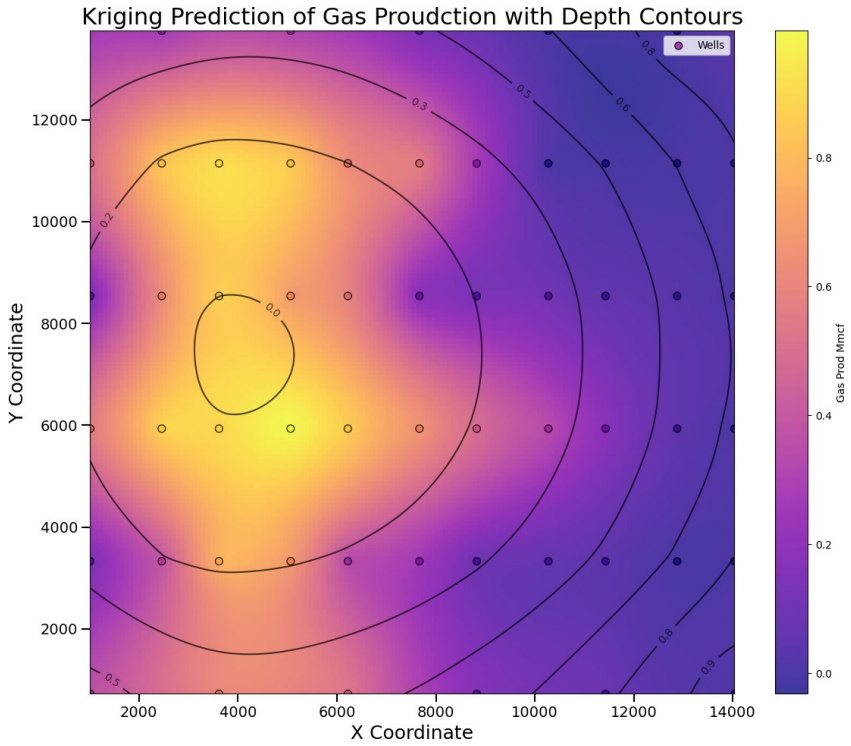



Figure 3:Kriging Map of Gas Production

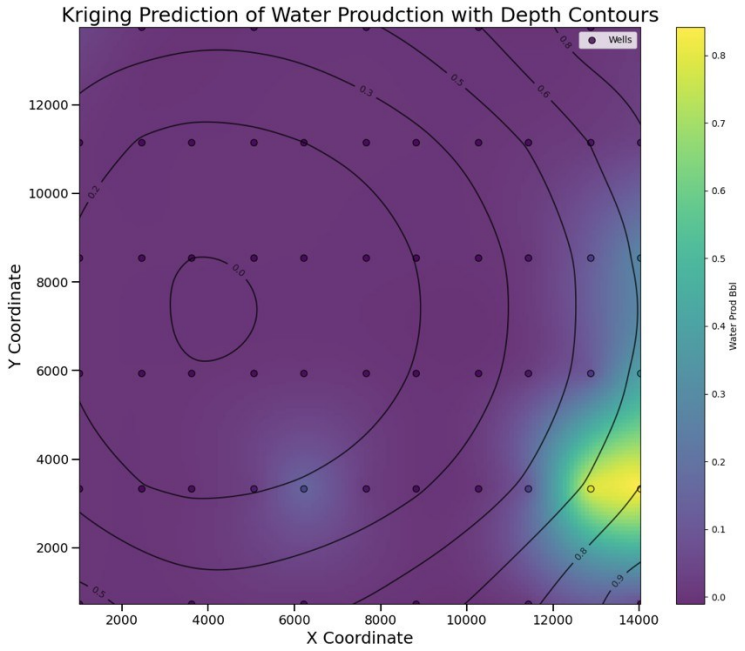
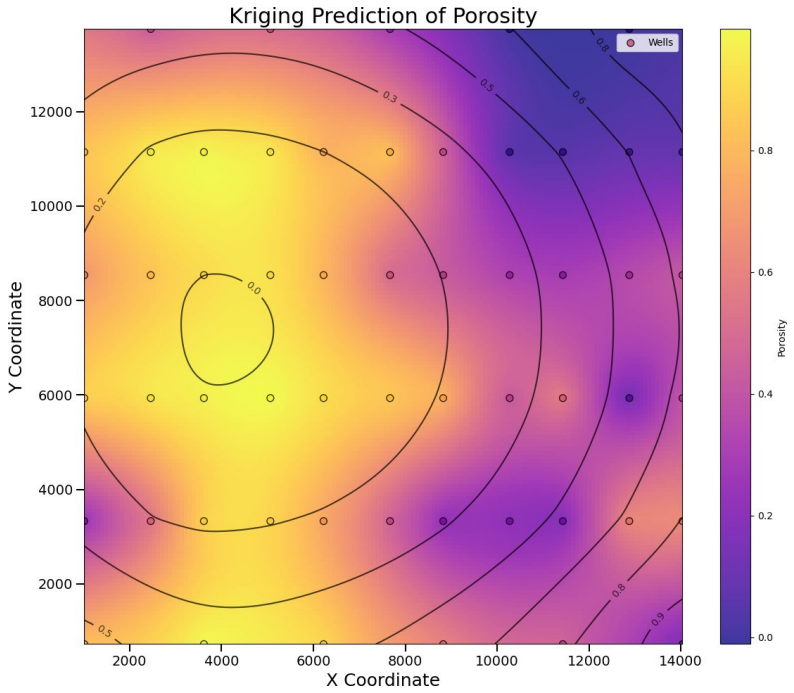


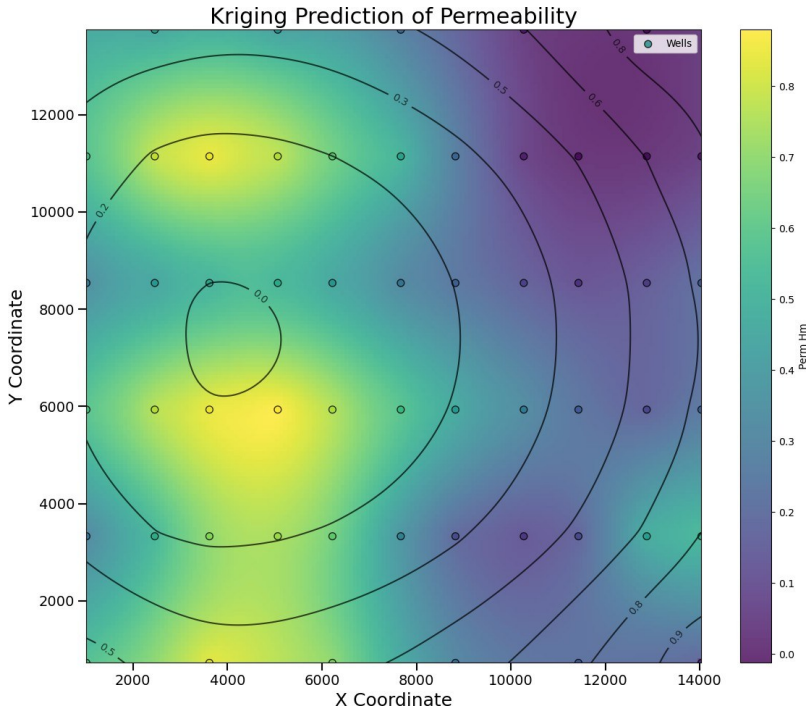

Figure 4: Kriging Map of Water Production




462

463 Figure 5: Kriging Map of Porosity

464




465

466 Figure 6: Kriging Map of Permeability

467

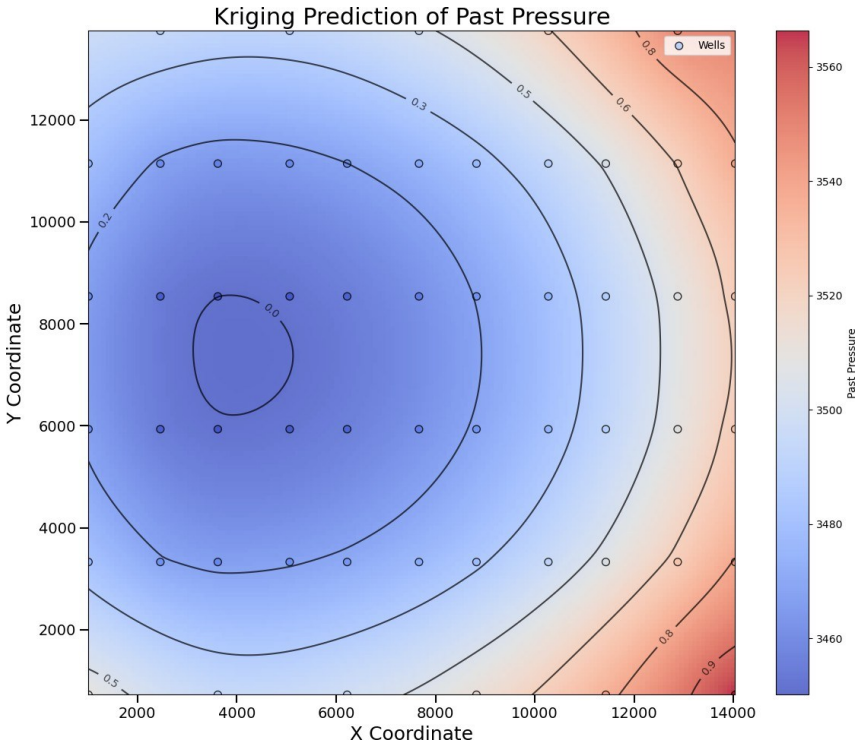



Figure 7: Kriging Map of Pressure

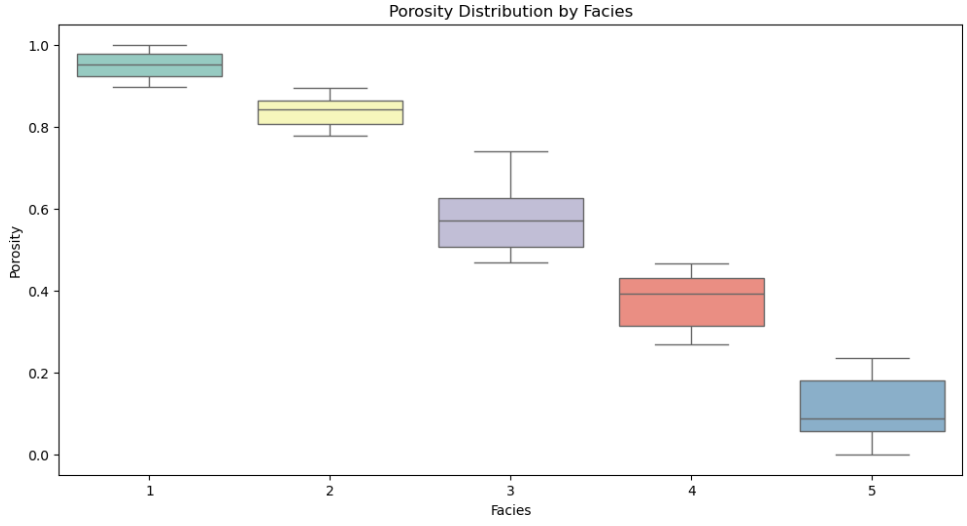



Figure 8: Porosity Distribution by Facies

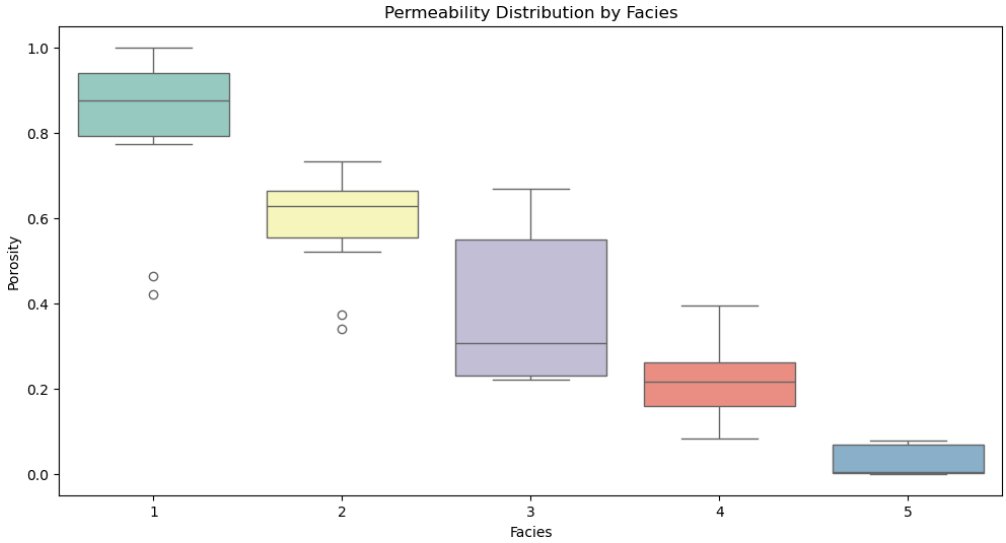



Figure 9: Permeability Distribution by Facies

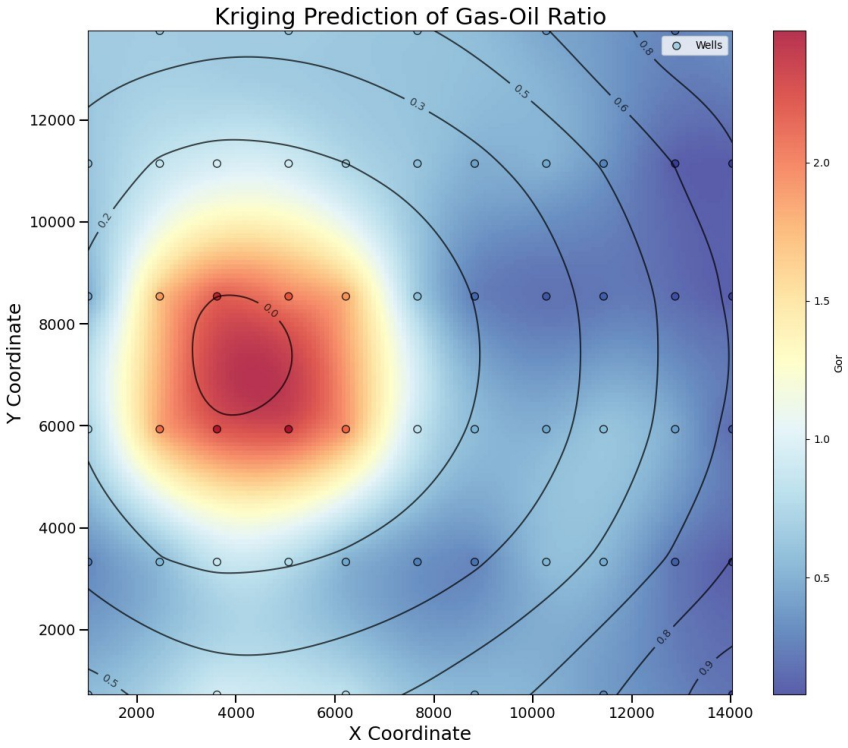



Figure 10: Kriging Map of Gas-Oil Ratio

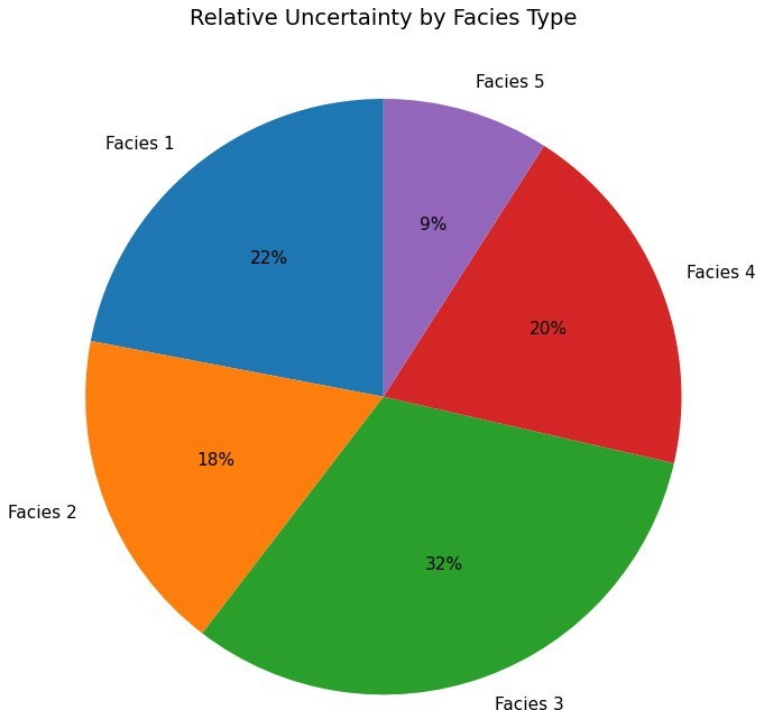



Figure 11: Relative Uncertainty by Facies (Random Forest)

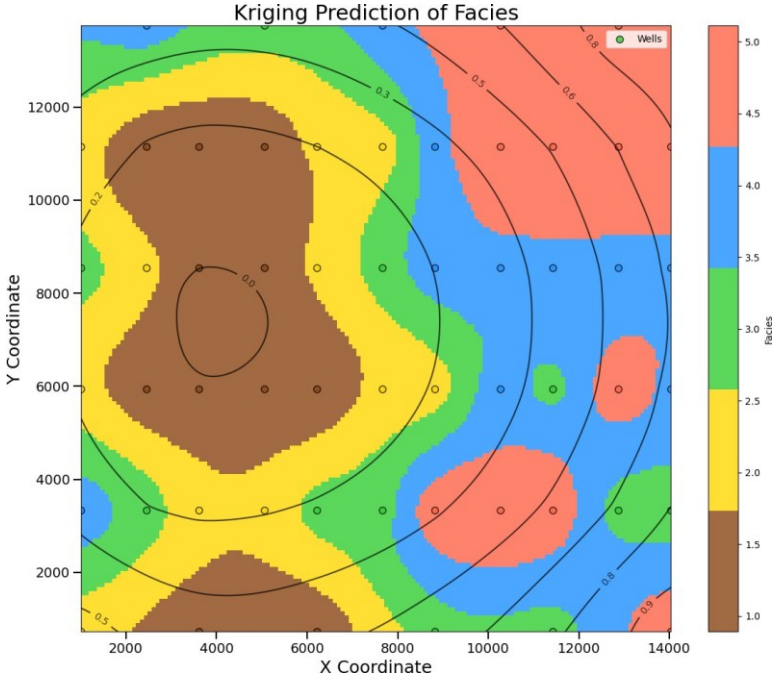



Figure 12: Kriging Map of Facies Distribution

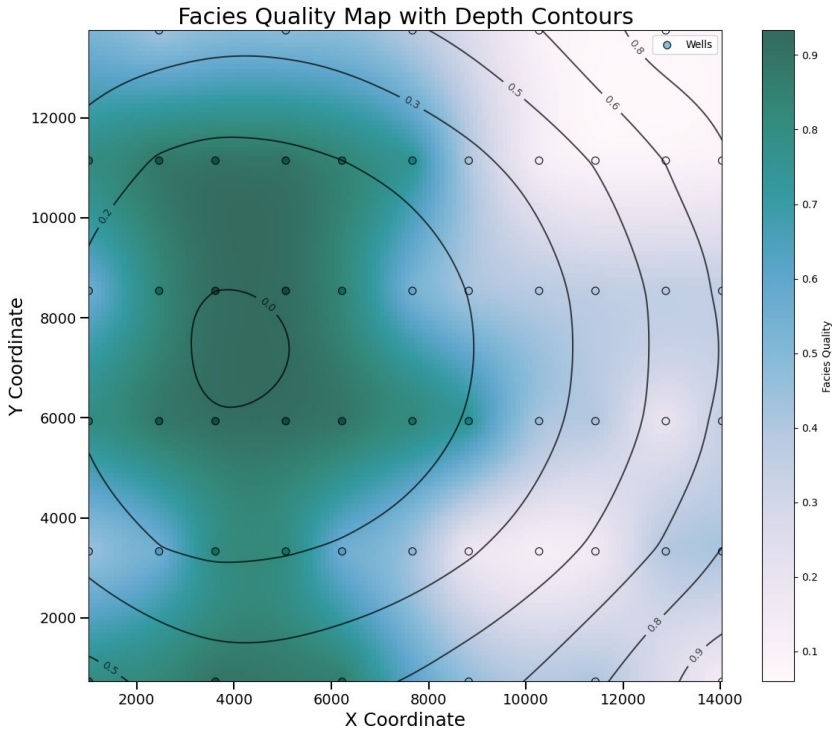



Figure 13: Kriging Map of Facies Quality Score

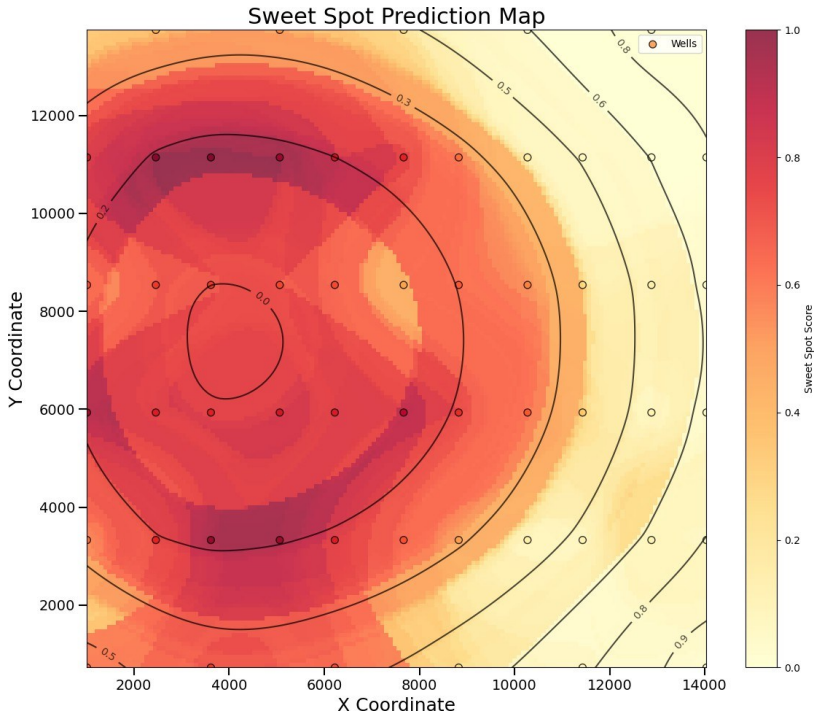
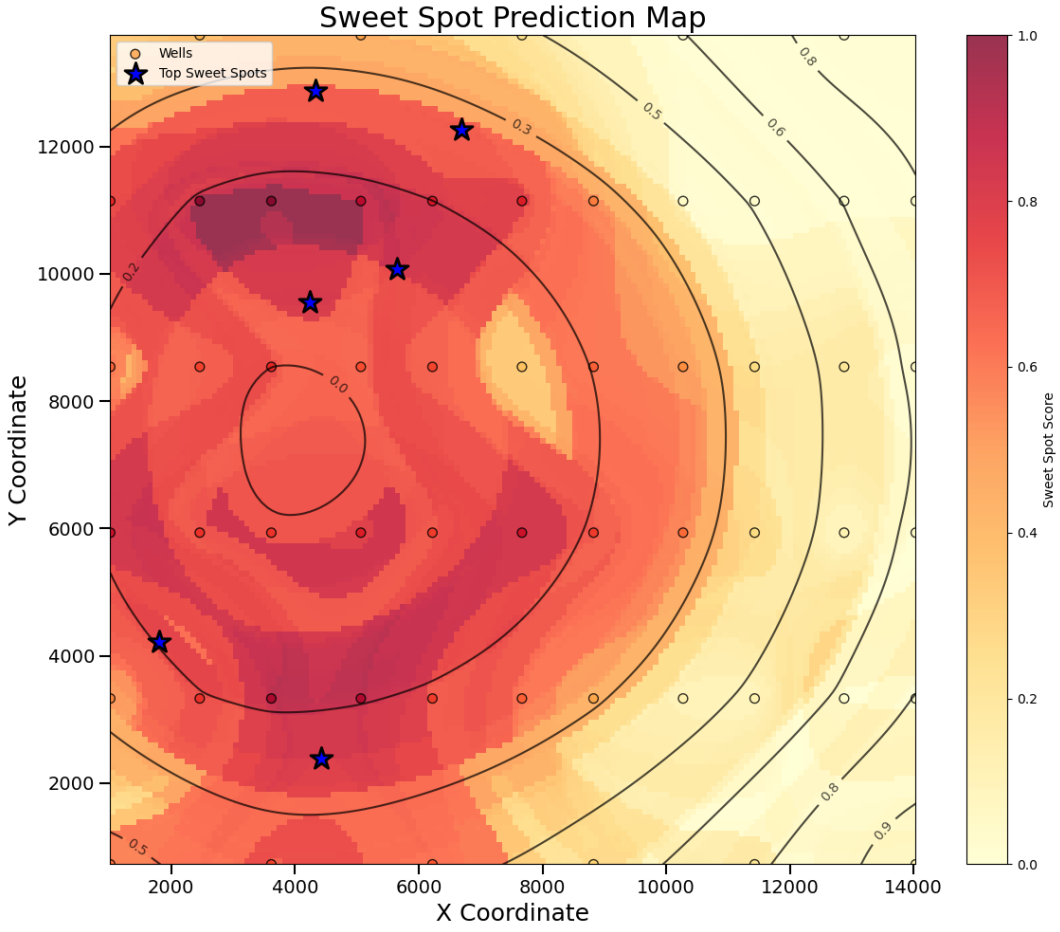



Figure 14: Kriging Map of Sweetspot Scores

493




494

495 Figure 15: Kriging Map of Sweetspot Scores overlaid with the six highest scores