SWEET SPOT IDENTIFICATION

Team Sweet Spot - Problem 2

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

ConocoPhillips

PROBLEM STATEMENT

Goal: Identify the most promising drilling locations (“sweet spots”) that maximize oil & gas production.

Core Questions:

- Which geological and reservoir properties drive productivity?
- Where on the field map do these optimal conditions overlap?
- How can we quantify confidence and risk in our recommendations?

PROJECT ROADMAP

Data Exploration

Understand reservoir & production behavior

Spatial & Geological Analysis

Identify structural patterns

Modeling

Quantify productivity drivers

Final Recommendation

Select optimal drilling locations

DATA OVERVIEW

Sourced from ConocoPhillips			Team Produced
Petrophysical	**Production**	**Spatial**	**Derived**
• Depth • Porosity • Permeability • Facies	• Oil Output • Pressure • Locations	• X/Y Well Locations • Bottom-hole Depth • 3D Coordinates	• Harmonic Averages • Quality Index • Facies Quality Score

METHODOLOGY OVERVIEW

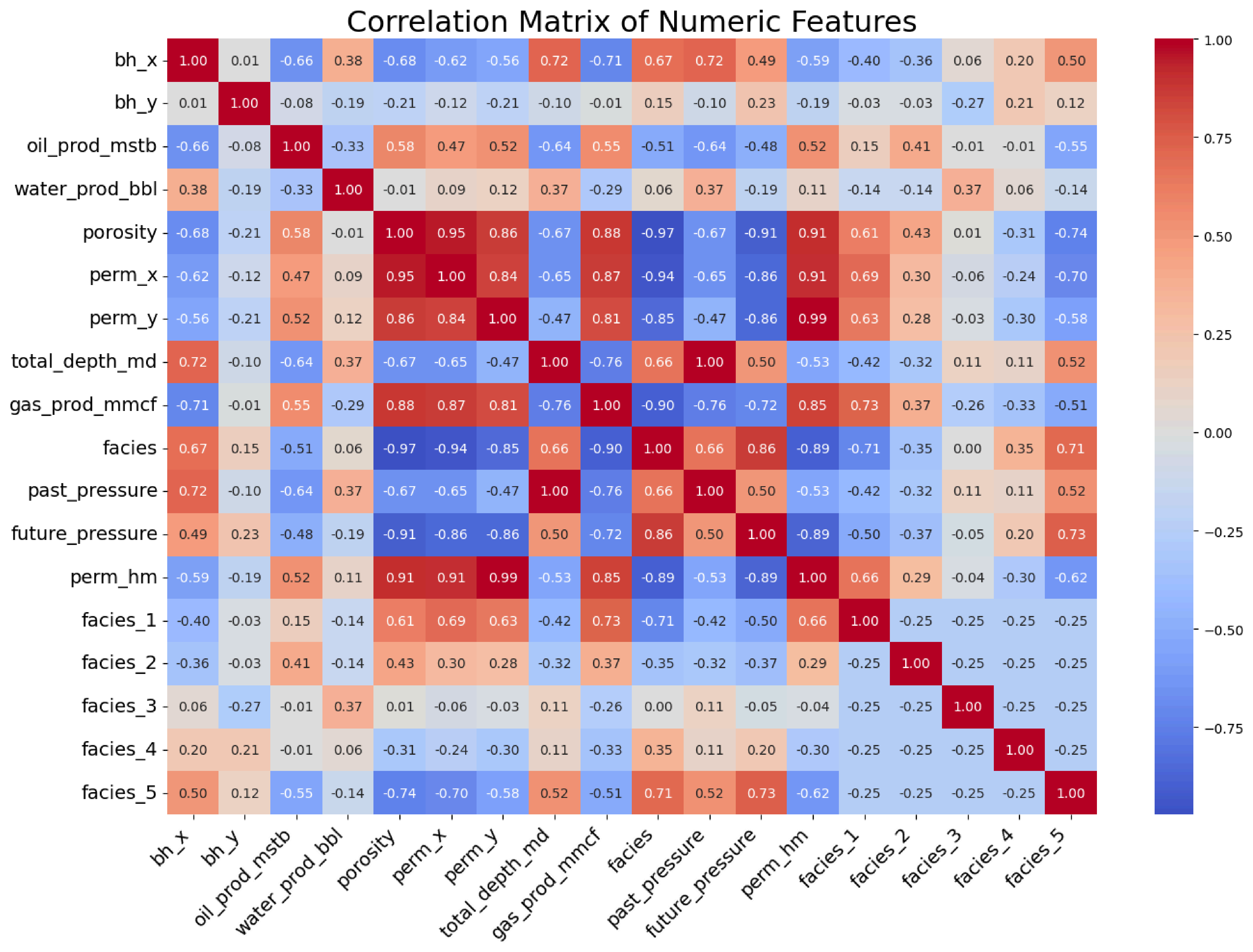
Our pipeline integrates geology, production, and spatial modeling:

1. Data Cleaning & Outlier Handling
2. Spatial Modeling (Kriging & Facies Interpolation)
3. Facies Modeling
   - Spatial classification of dominant rock types
4. Feature Engineering
   - Harmonic Averages
   - Composite Quality Index
   - Facies Quality Score
5. Machine Learning Models
   - XGBoost for Sweet Spot Prediction
   - Random Forest for Facies Quality Scoring
6. Uncertainty Quantification

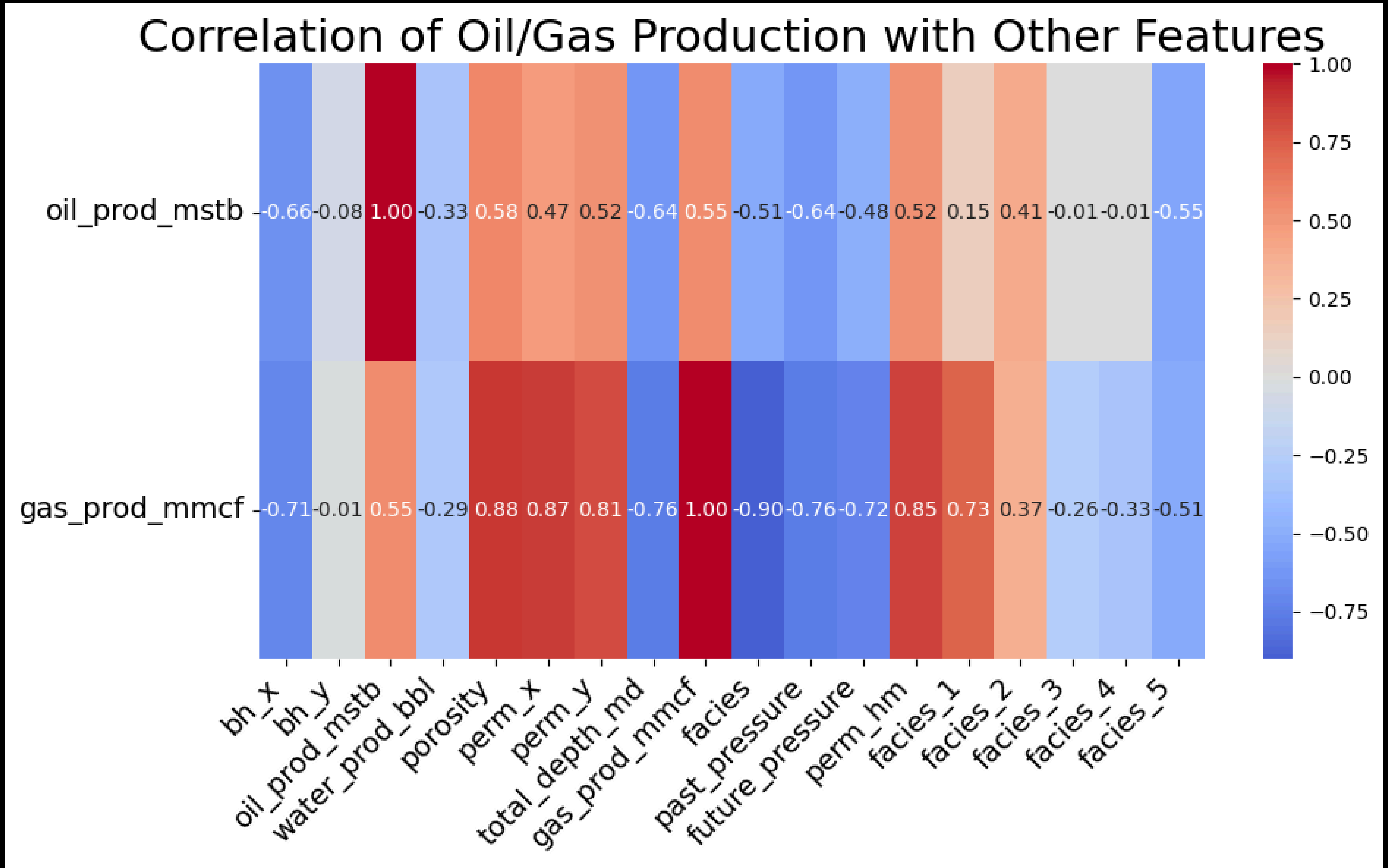
DATA EXPLORATION

Understanding the raw behavior of the field before modeling

HEATMAPS


Correlation Matrix of Numeric Features
bh_x 1.00 0.01 -0.66 0.38 -0.68 -0.62 -0.56 0.72 -0.71 0.67 0.72 0.49 -0.59 -0.40 -0.36 0.06 0.20 0.50
bh_y 0.01 1.00 -0.08 -0.19 -0.21 -0.12 -0.21 -0.10 -0.01 0.15 -0.10 0.23 -0.19 -0.03 -0.03 -0.27 0.21 0.12
oil_prod_mstb -0.66 -0.08 1.00 -0.33 0.58 0.47 0.52 -0.64 0.55 -0.51 -0.64 -0.48 0.52 0.15 0.41 -0.01 -0.01 -0.55
water_prod_bbl 0.38 -0.19 -0.33 1.00 -0.01 0.09 0.12 0.37 -0.29 0.06 0.37 -0.19 0.11 -0.14 -0.14 0.37 0.06 -0.14
porosity -0.68 -0.21 0.58 -0.01 1.00 0.95 0.86 -0.67 0.88 -0.97 -0.67 -0.91 0.91 0.61 0.43 0.01 -0.31 -0.74
perm_x -0.62 -0.12 0.47 0.09 0.95 1.00 0.84 -0.65 0.87 -0.94 -0.65 -0.86 0.91 0.69 0.30 -0.06 -0.24 -0.70
perm_y -0.56 -0.21 0.52 0.12 0.86 0.84 1.00 -0.47 0.81 -0.85 -0.47 -0.86 0.99 0.63 0.28 -0.03 -0.30 -0.58
total_depth_md 0.72 -0.10 -0.64 0.37 -0.67 -0.65 -0.47 1.00 -0.76 0.66 1.00 0.50 -0.53 -0.42 -0.32 0.11 0.11 0.52
gas_prod_mmcf -0.71 -0.01 0.55 -0.29 0.88 0.87 0.81 -0.76 1.00 -0.90 -0.76 -0.72 0.85 0.73 0.37 -0.26 -0.33 -0.51
facies 0.67 0.15 -0.51 0.06 -0.97 -0.94 -0.85 0.66 -0.90 1.00 0.66 0.86 -0.89 -0.71 -0.35 0.00 0.35 0.71
past_pressure 0.72 -0.10 -0.64 0.37 -0.67 -0.65 -0.47 1.00 -0.76 0.66 1.00 0.50 -0.53 -0.42 -0.32 0.11 0.11 0.52
future_pressure 0.49 0.23 -0.48 -0.19 -0.91 -0.86 -0.86 0.50 -0.72 0.86 0.50 1.00 -0.89 -0.50 -0.37 -0.05 0.20 0.73
perm_hm -0.59 -0.19 0.52 0.11 0.91 0.91 0.99 -0.53 0.85 -0.89 -0.53 -0.89 1.00 0.66 0.29 -0.04 -0.30 -0.62
facies_1 -0.40 -0.03 0.15 -0.14 0.61 0.69 0.63 -0.42 0.73 -0.71 -0.42 -0.50 0.66 1.00 -0.25 -0.25 -0.25 -0.25
facies_2 -0.36 -0.03 0.41 -0.14 0.43 0.30 0.28 -0.32 0.37 -0.35 -0.32 -0.37 0.29 -0.25 1.00 -0.25 -0.25 -0.25
facies_3 0.06 -0.27 -0.01 0.37 0.01 -0.06 -0.03 0.11 -0.26 0.00 0.11 -0.05 -0.04 -0.25 -0.25 1.00 -0.25 -0.25
facies_4 0.20 0.21 -0.01 0.06 -0.31 -0.24 -0.30 0.11 -0.33 0.35 0.11 0.20 -0.30 -0.25 -0.25 -0.25 1.00 -0.25
facies_5 0.50 0.12 -0.55 -0.14 -0.74 -0.70 -0.58 0.52 -0.51 0.71 0.52 0.73 -0.62 -0.25 -0.25 -0.25 -0.25 1.00
bh_x bh_y oil_prod_mstb water_prod_bbl porosity perm_x perm_y total_depth_md gas_prod_mmcf facies past_pressure future_pressure perm_hm facies_1 facies_2 facies_3 facies_4 facies_5
1.00 0.75 0.50 0.25 0.00 −0.25 −0.50 −0.75

HEATMAPS



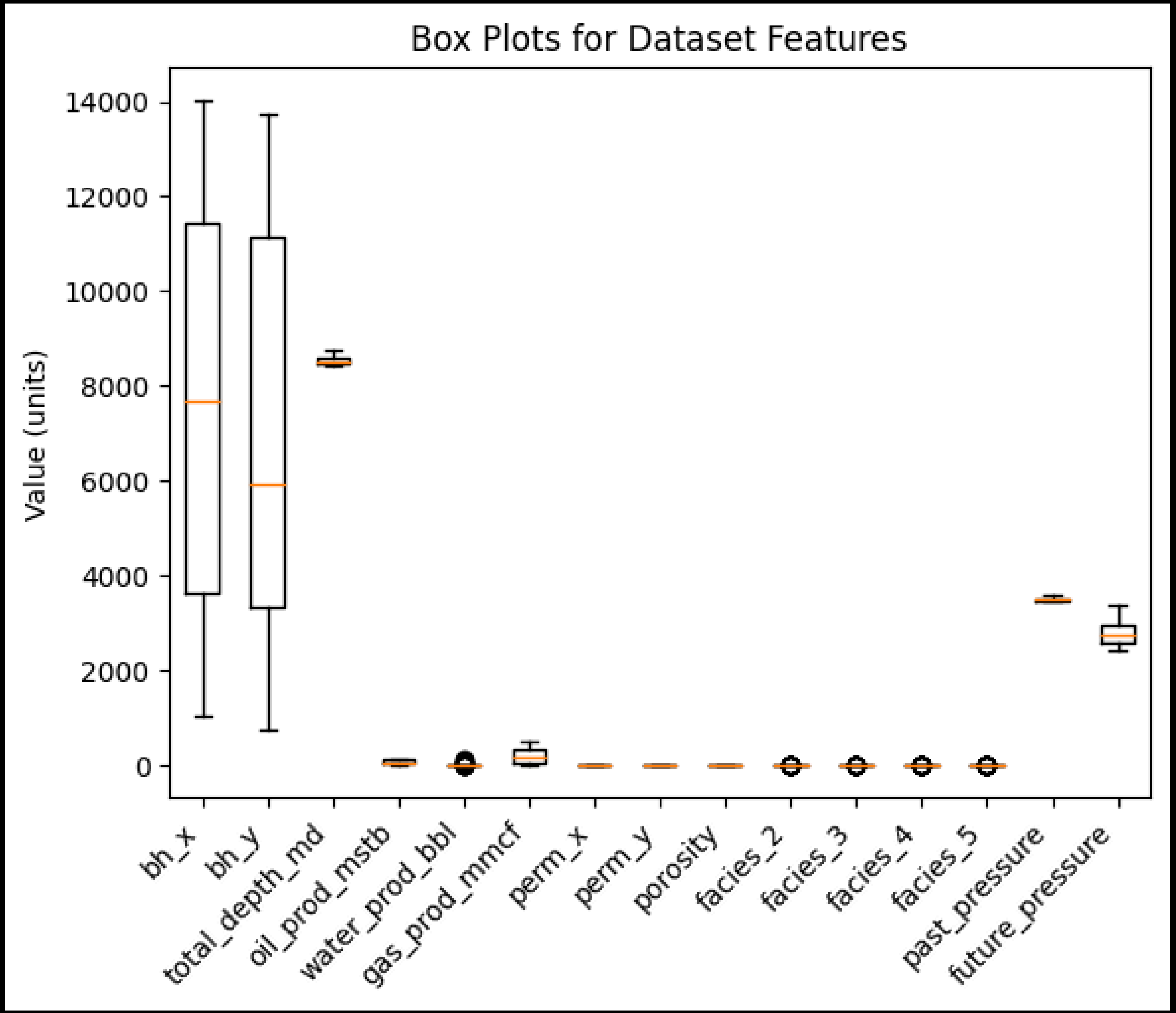
Key Takeaway: Production is most strongly controlled by pressure and rock quality (porosity and permeability), justifying their use as primary drivers in spatial and sweet-spot modeling

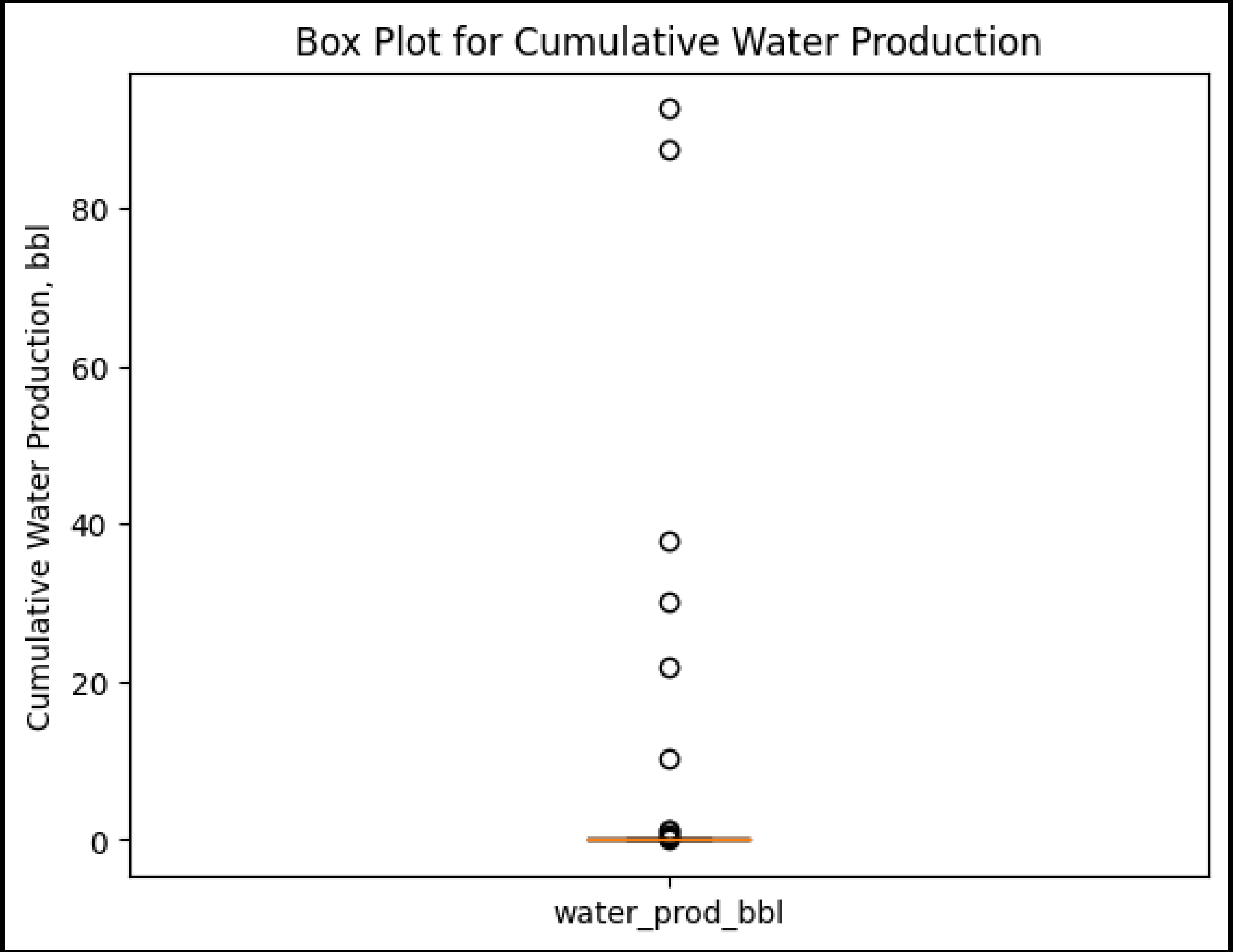
MISSING VALUES

- Printing the amount of 0s in the dataset revealed 5 data points with missing data, which means that 9% of wells have missing data.
  - Oil and gas production (columns Co [MSTB] and Cw (bbl))
  - Likely reflect measurement, reporting, or operational gaps rather than missing geological properties
- Implemented median imputation on oil and gas production due to outlier robustness.

	Well Number	Well Name	Bottomhole X	Bottomhole Y	Co [MSTB]	Cw (bbl)	POROS	KX	KY	TD(MD)	Cg (mmcf)	FACIES	P_2020-1-6	P_2029-1-1
7	8	PO1_8	12877.11	11141.88	0.0	1.24460	0.074247	0.000270	0.000270	8638.6	0.0	5	3520.804323	2949.434450
8	9	PO1_9	14034.61	11141.88	0.0	1.30679	0.073380	0.000284	0.000284	8684.3	0.0	5	3535.967335	2954.485548
26	27	PO1_27	14034.61	8537.52	0.0	30.15660	0.116111	0.033130	0.010039	8645.4	0.0	4	3523.060526	2676.168143
27	28	PO1_28	14034.61	5933.16	0.0	37.77570	0.112872	0.012573	0.012573	8647.7	0.0	4	3523.823653	2694.598799
46	47	PO1_47	14034.61	3328.80	0.0	87.57740	0.143755	0.038000	0.038000	8691.2	0.0	3	3538.256717	2515.127738

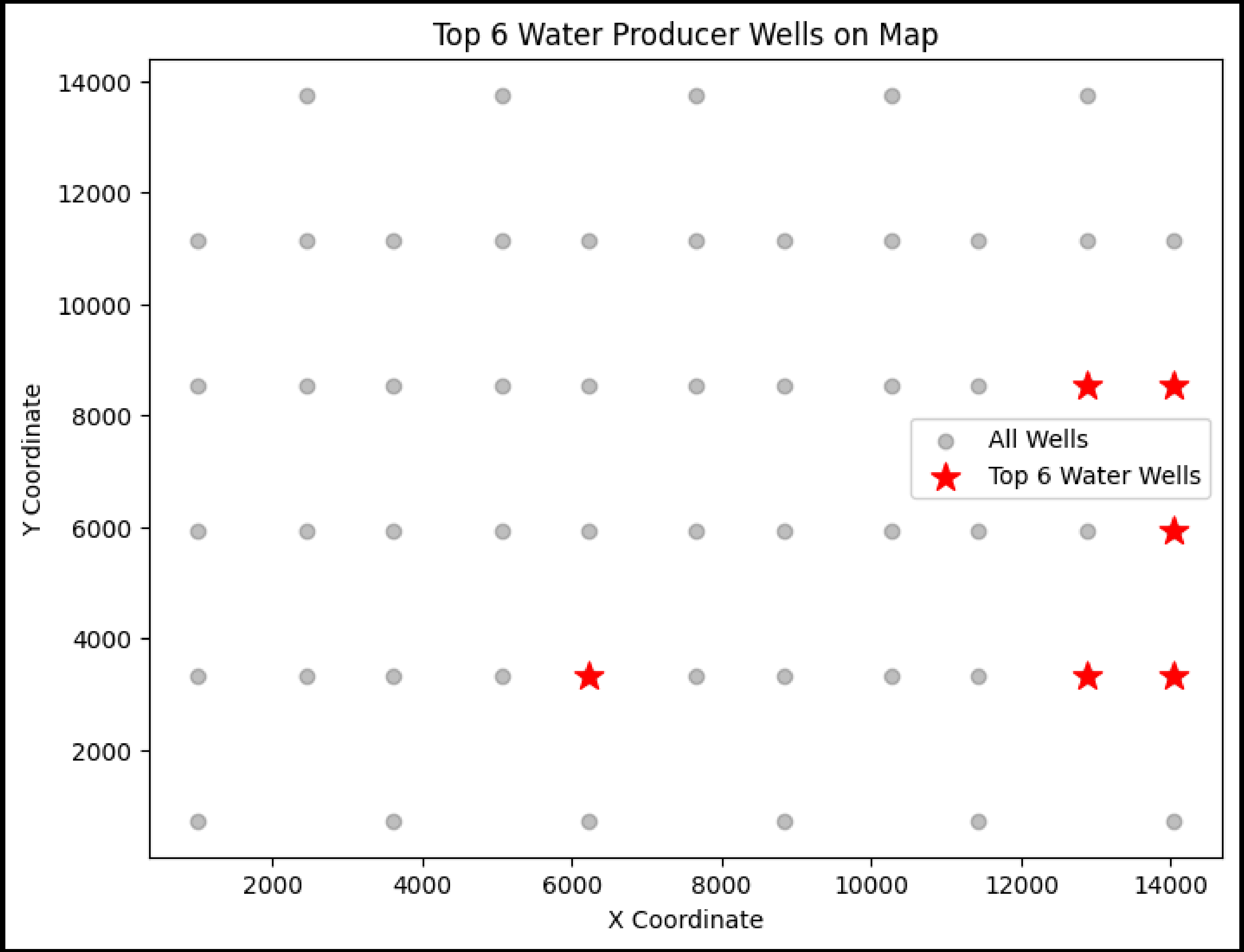
OUTLIERS





Key Takeaway: Extreme outliers occur only in water production, not in geological properties, and were isolated for spatial analysis to avoid biasing oil-productivity results

OUTLIERS

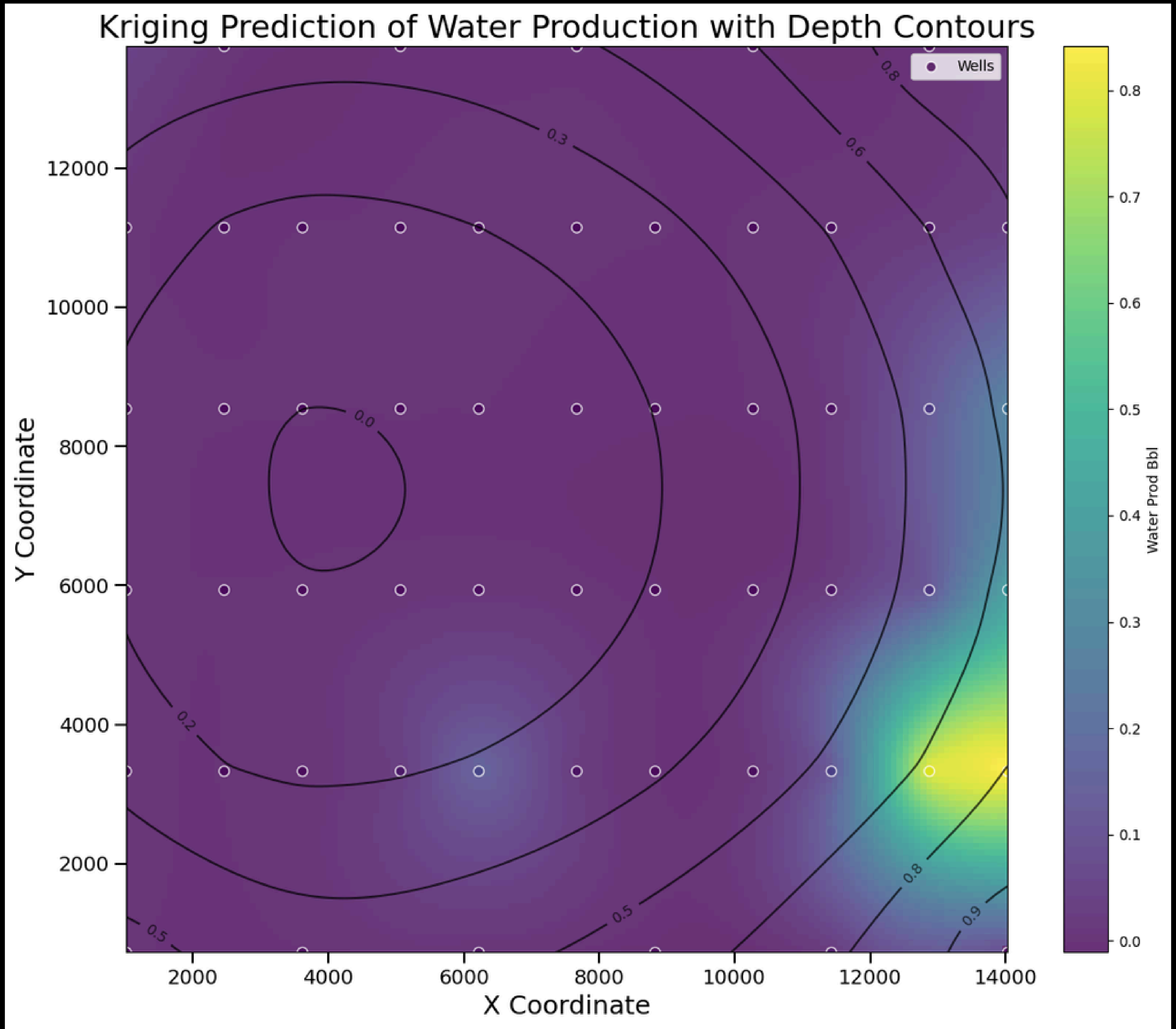


Key Takeaway: Extreme water-production outliers are spatially clustered, motivating the need for spatial kriging to understand their broader field-wide impact

SPATIAL & GEOLOGICAL ANALYSIS

How reservoir characteristics vary across the field

WATER PRODUCTION - KRIGING MAP

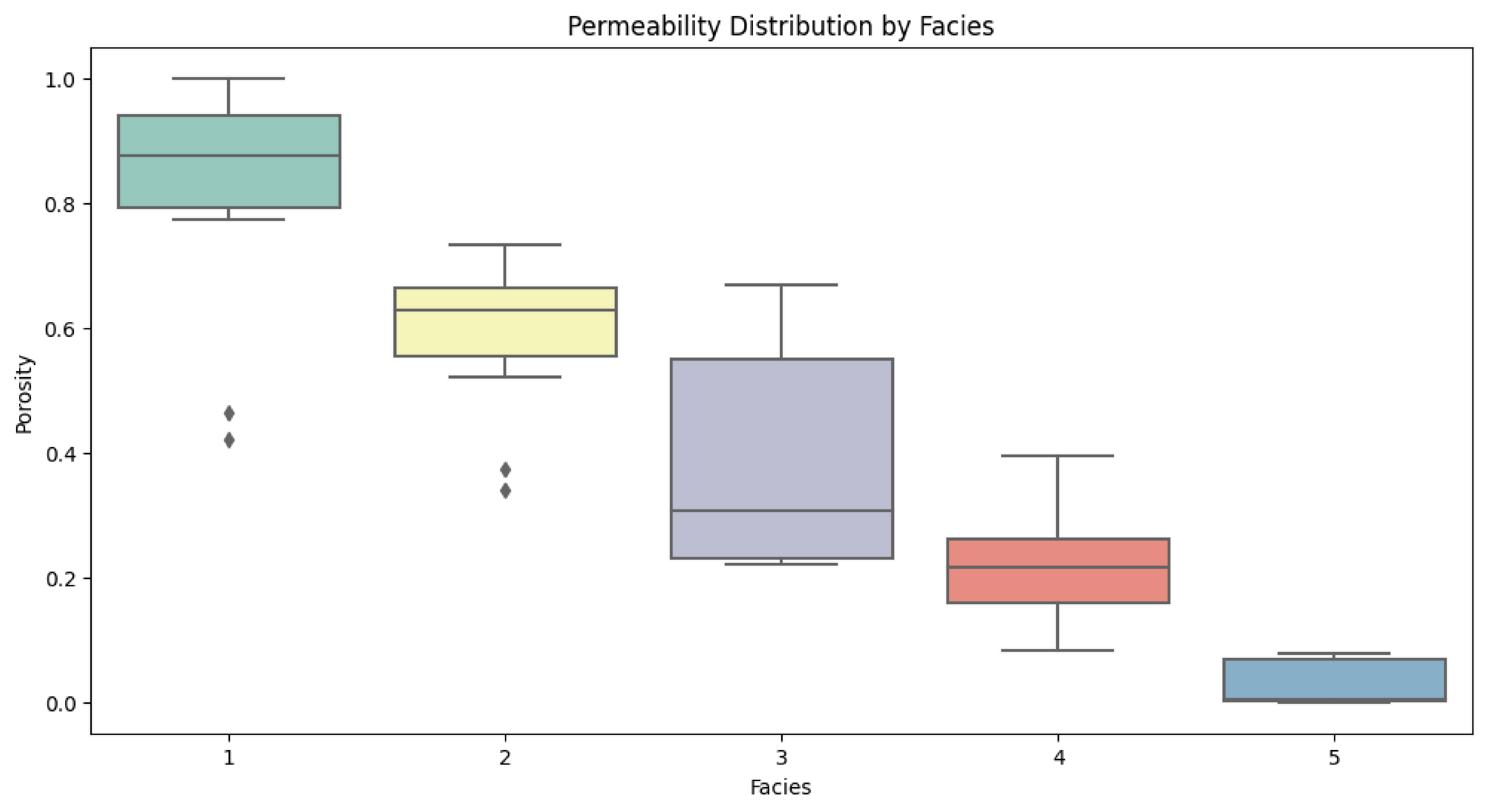


Key Takeaway

Spatial interpolation reveals localized high-water zones that can distort oil-productivity interpretation if not explicitly accounted for

***Kriging Maps are Interpolated Visualizations (Not Direct Measurements)*

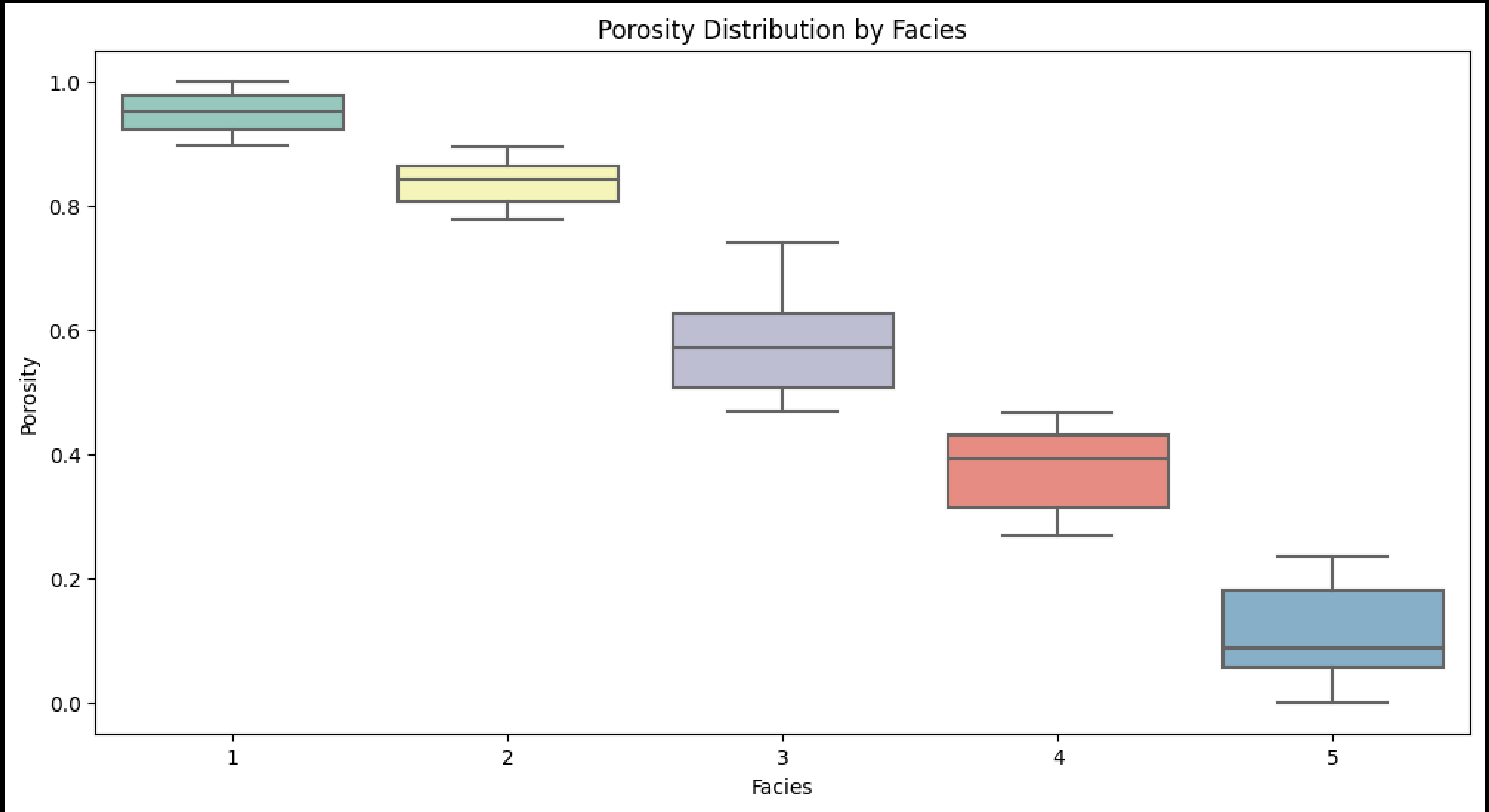
PERMEABILITY BY FACIES



- Facies 1 and 2 exhibit the highest median permeability
- Facies 3 shows moderate permeability with high variability
- Facies 4 and 5 consistently display low permeability

Key Takeaway: Facies 1 exhibits the strongest and most consistent permeability, making it the most favorable facies for sustained fluid flow

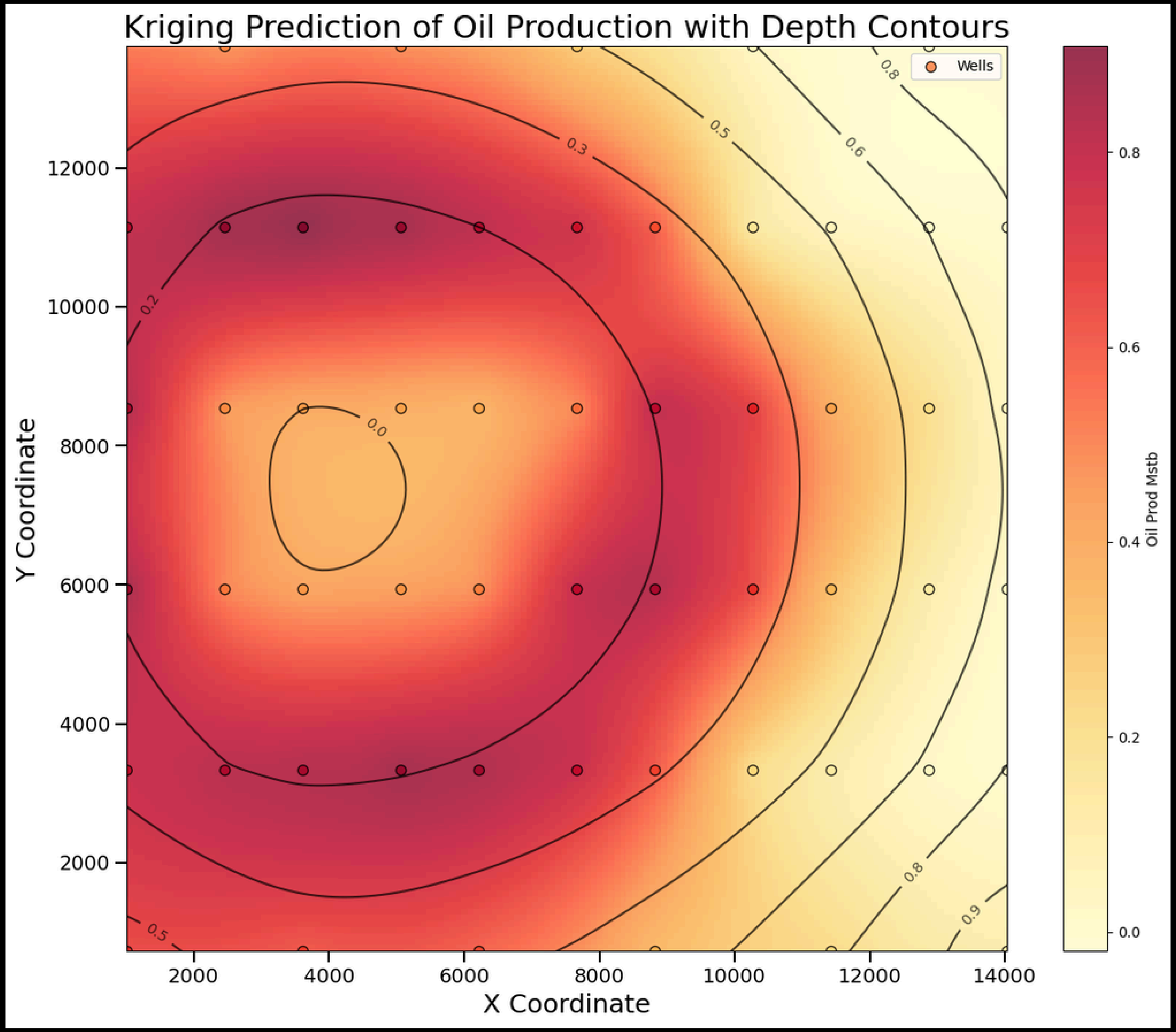
POROSITY BY FACIES



- Facies 1 and 2 exhibit the highest median porosity
- Facies 3 shows moderate porosity with noticeable variability
- Facies 4 and 5 consistently display low porosity

Key Takeaway: Facies 1 provides the greatest storage capacity based on its consistently high porosity, supporting its role as a prime drilling target

OIL PRODUCTION - KRIGING MAP

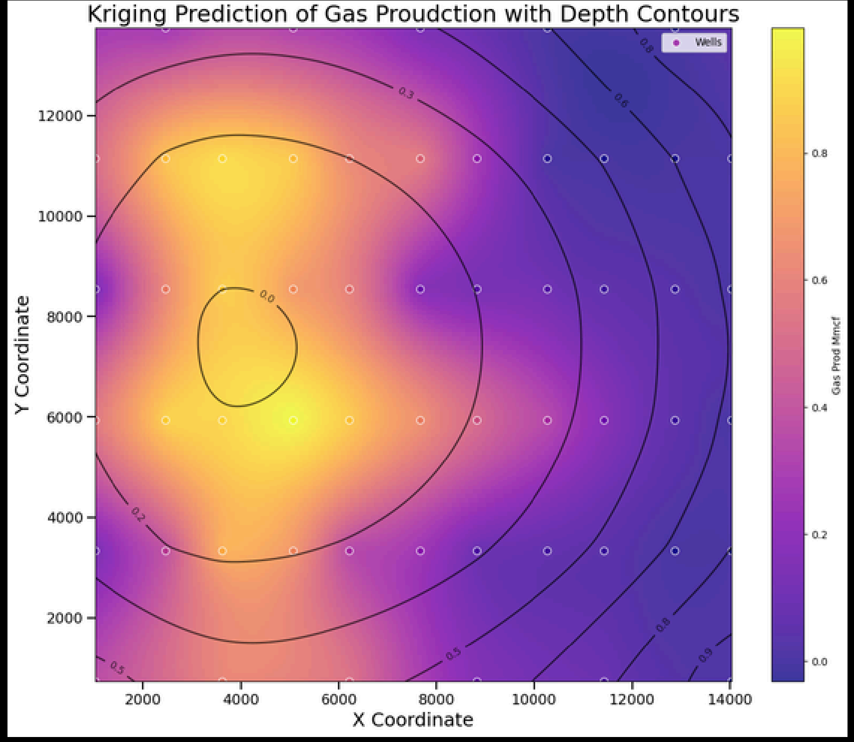


Key Takeaway

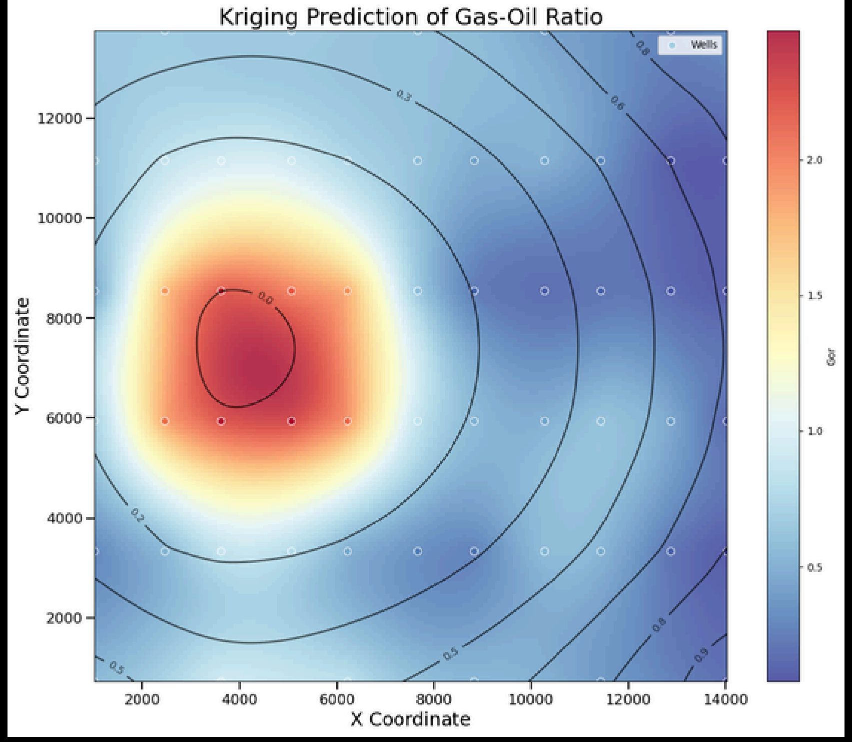
The kriging surface reveals elevated oil production surrounding the structural high, suggesting a strong link between structure and productivity

MORE KRIGING MAPS

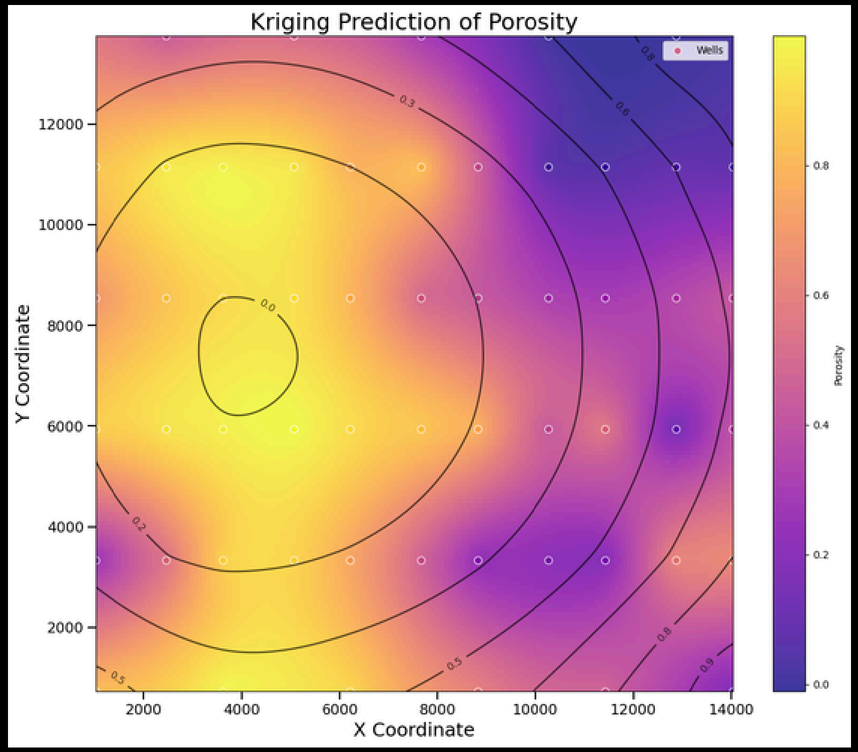
GAS PRODUCTION

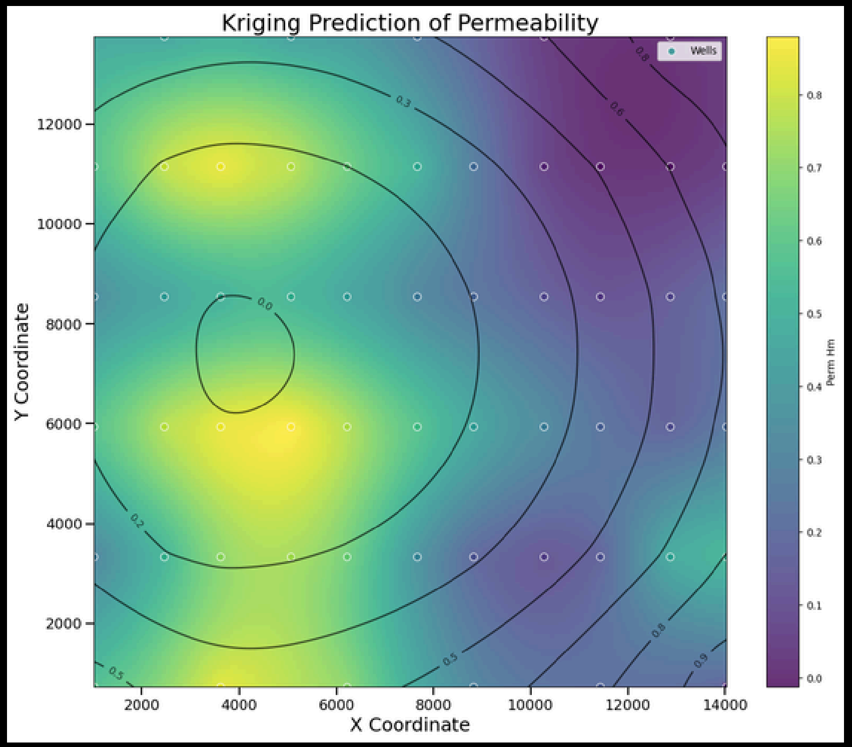


GAS/OIL RATIO PRODUCTION

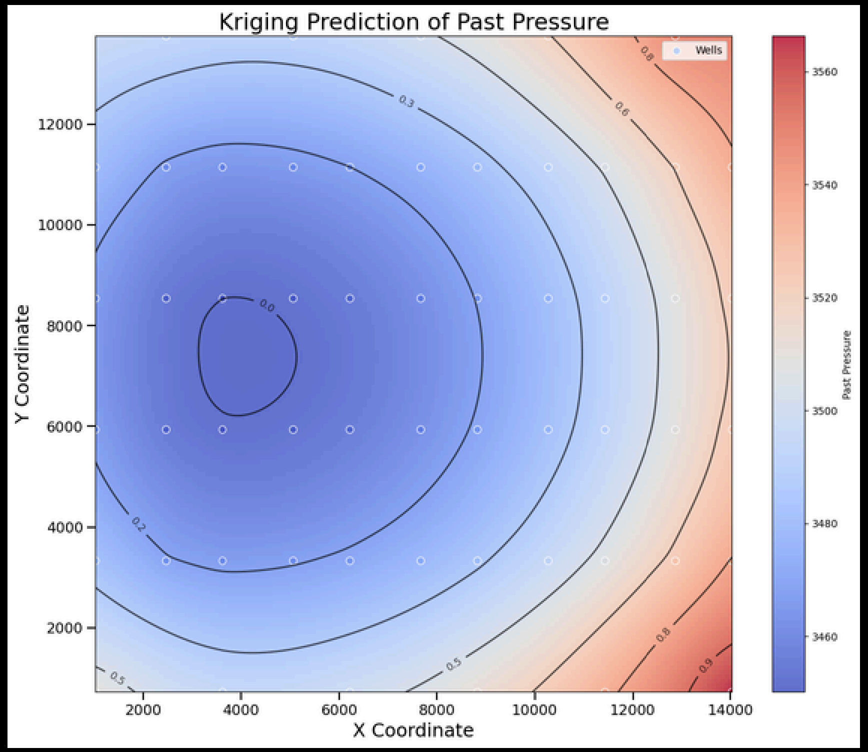


POROSITY





PERMEABILITY



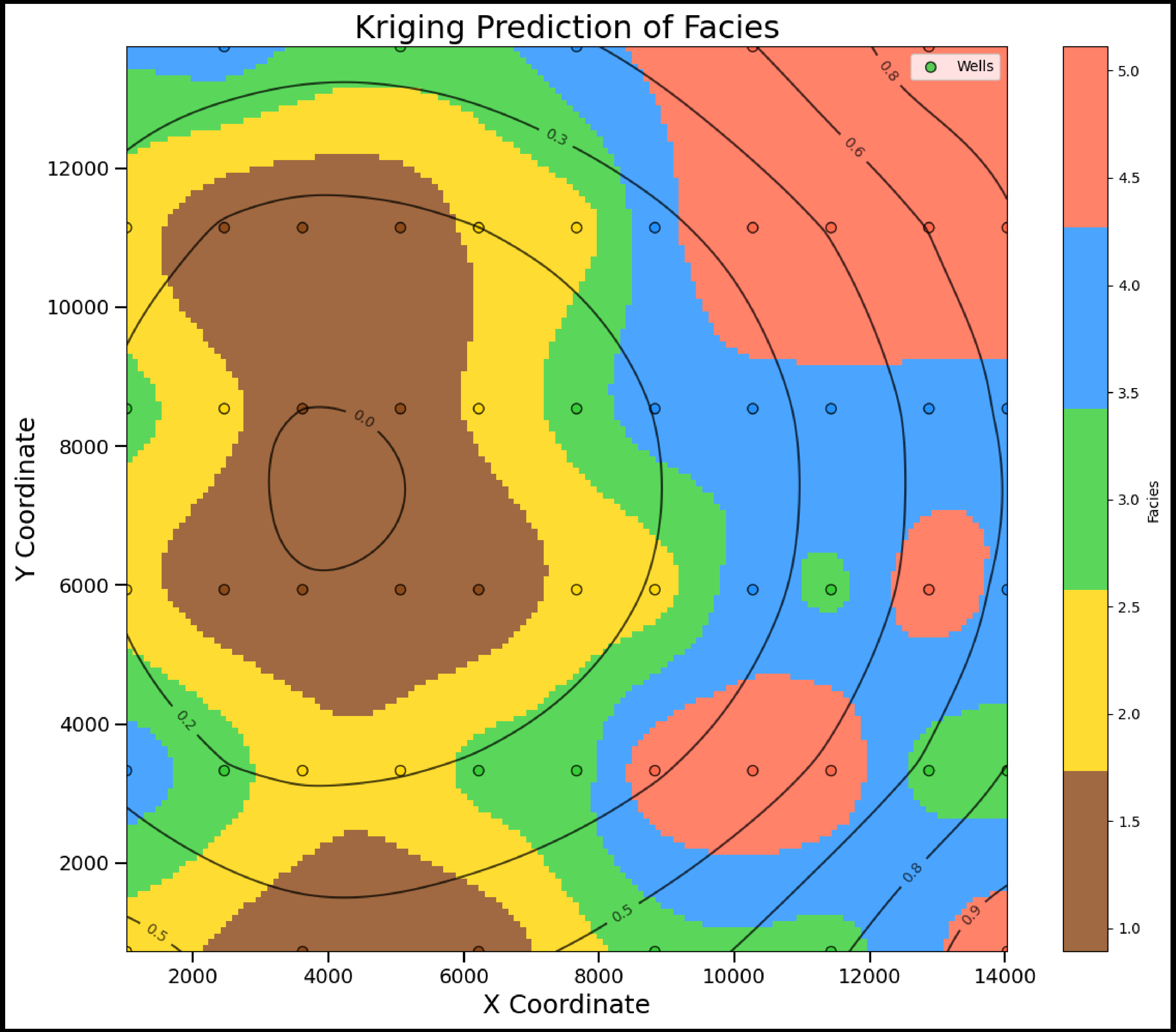
PRESSURE

Key Takeaway

Consistent structural alignment across rock and fluid properties confirms a coherent reservoir system that supports the model-driven sweet-spot targeting

All kriging surfaces shown for spatial visualization only.

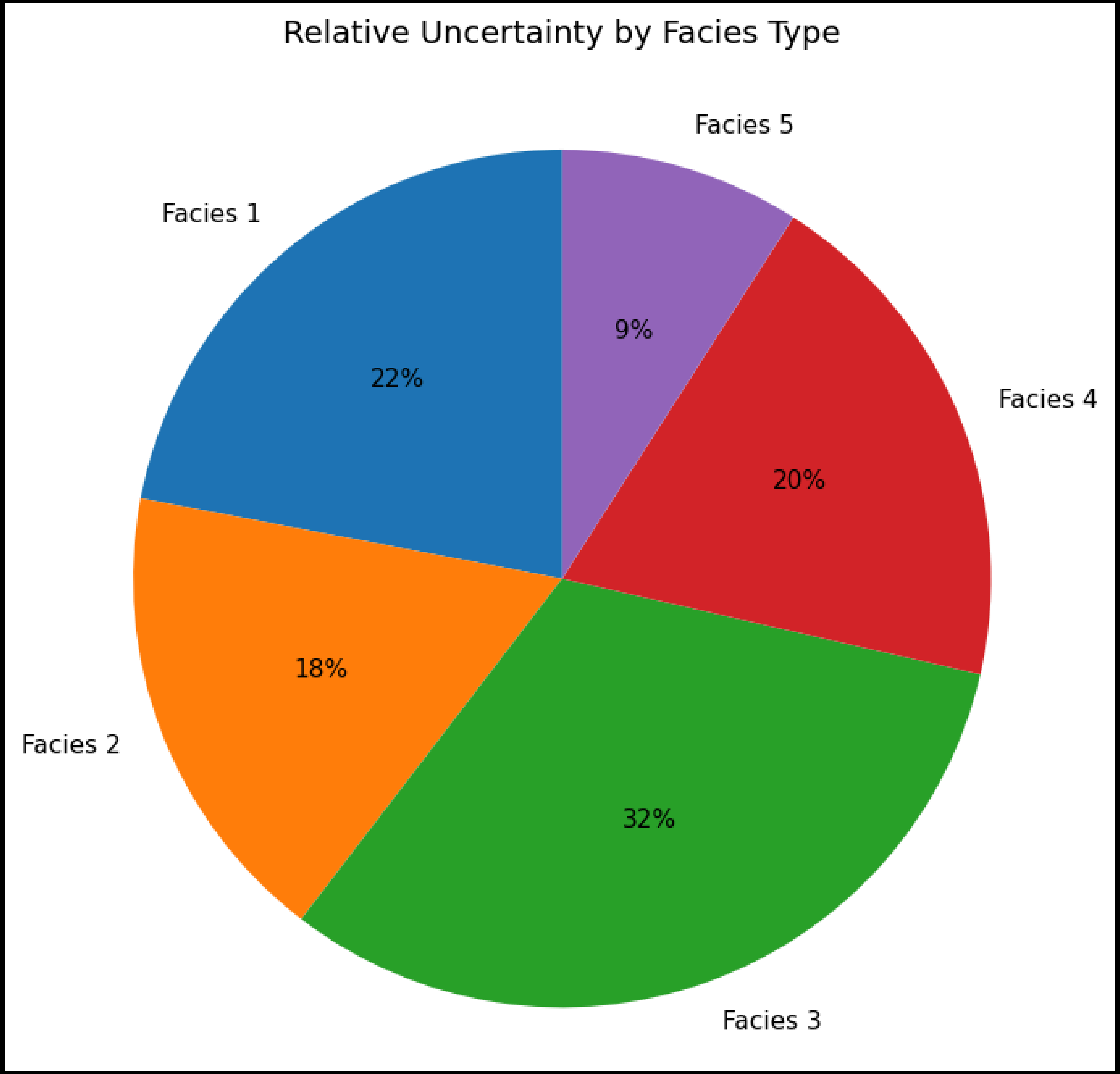
FACIES DISTRIBUTION - KRIGING MAP



Key Takeaway

Facies exhibit coherent spatial structure across the field, providing geological context for production patterns but serving as an input to the model rather than a direct drilling decision tool

FACIES UNCERTAINTY DISTRIBUTION



Key Takeaway

Facies 3 carries the highest classification uncertainty, indicating greater geological risk in those zones relative to the more confidently resolved facies

OVERALL OBSERVATIONS

- Oil and gas production exhibit strong spatial organization, consistent with underlying structural and geological controls
- Production patterns align with systematic variations in depth and facies, indicating that both structure and rock type influence productivity
- Facies 1 consistently emerges as the highest-quality reservoir unit, supported by:
    - High porosity
    - High permeability
- Pressure and fluid behavior vary coherently across the field, further reinforcing the structural control on reservoir performance

HARMONIC AVERAGE

- Best suited for rates and ratio-based variables (e.g., permeability, production efficiency)
- Computed using the reciprocal of values, then averaged
- Down-weights extreme high outliers and emphasizes consistently low values
- Used in our feature engineering to create stable, outlier-resistant reservoir metrics

$$\text{Harmonic Mean} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

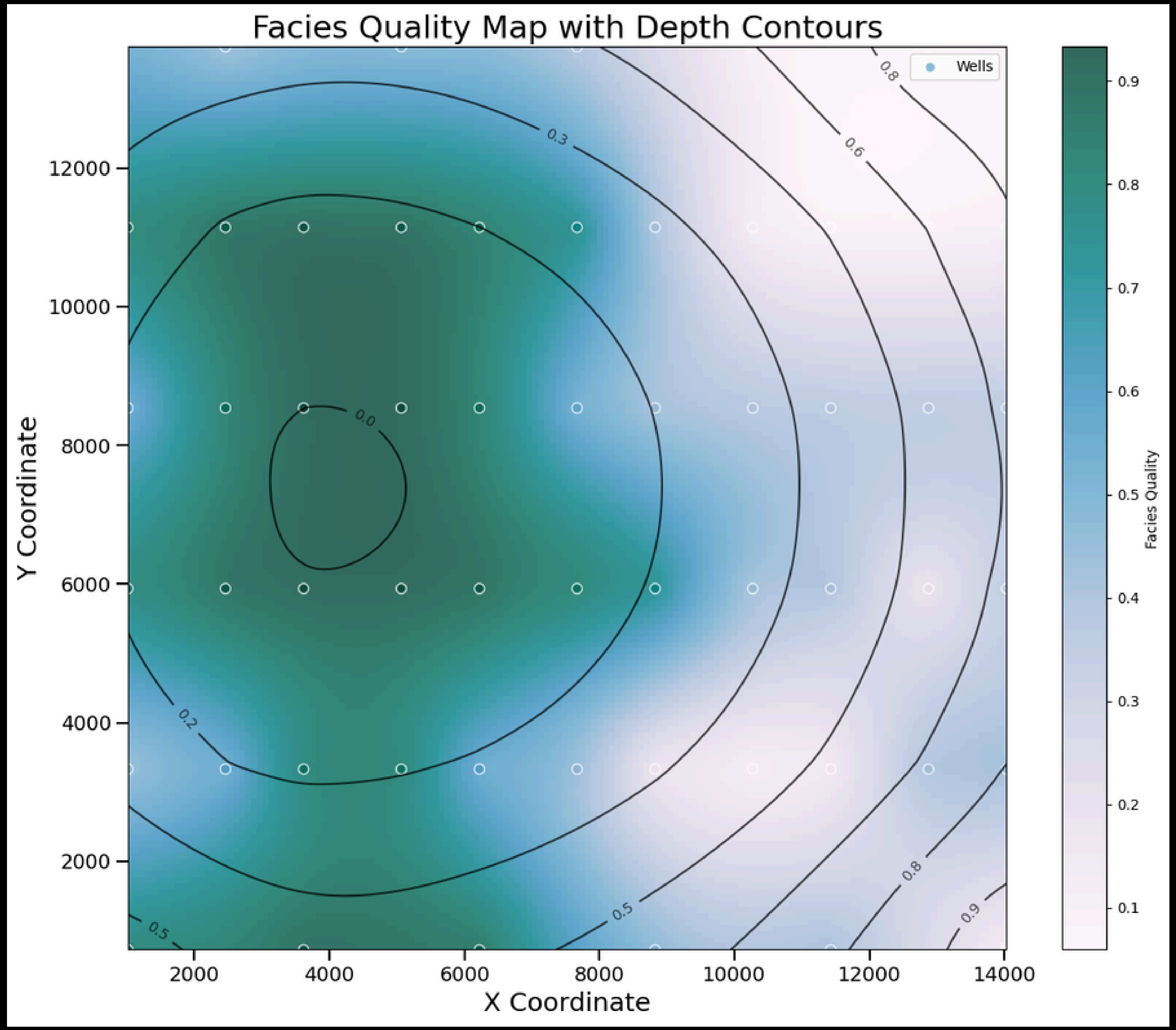
Key Takeaway: The harmonic average provides a conservative, outlier-robust measure of reservoir performance used directly in our feature engineering pipeline

QUALITY INDEX

- A single composite score that integrates multiple reservoir performance variables into one comparable metric
- Constructed using the harmonic average to ensure robustness to extreme values
- Prevents any single strong metric from dominating the overall ranking
- Ensures wells are ranked based on balanced reservoir quality, not isolated strengths

Key Takeaway: The Quality Index produces a conservative, balanced ranking of wells that reflects overall reservoir performance rather than isolated high values

FACIES QUALITY - KRIGING MAP



Why the Facies Quality score was created:

- Facies were found to control productivity more strongly than depth
- Compresses rock-type flow capacity into one spatial variable for sweet spot prediction

Why Random Forest was selected:

- Captures non-linear geology–productivity relationships
- Robust to class imbalance
- Provides interpretable feature importance

Key Takeaway: Facies quality reveals productivity-driven geological patterns that cannot be explained by depth alone and serves as a critical spatial input to the Sweet Spot model

MODEL & FEATURE ENGINEERING

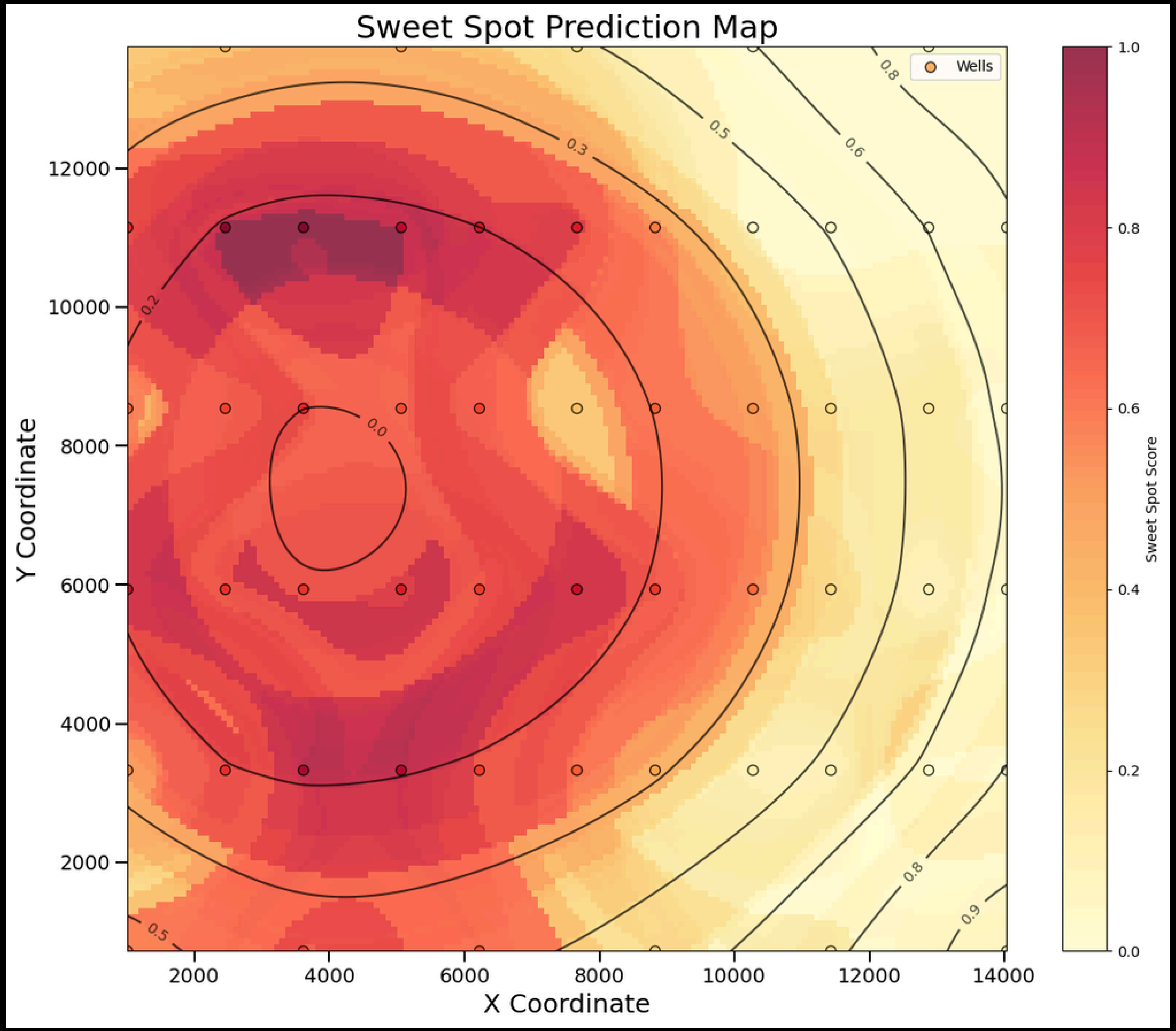
Translating geology into predictive insight

MODEL IMPLEMENTATION

- XGBoost model used for feature importance and production prediction
- Well-suited for medium-sized tabular datasets
- Captures non-linear interactions between geospatial and petrophysical variables
- Linear regression was tested but underperformed, failing to capture strong non-linear spatial effects
- Model output used to generate a "Predicted Production" variable
- Production Model Cross Validated R^2: 0.785 (+/- 0.134)

Key Takeaway: XGBoost effectively captures complex non-linear geological and spatial relationships that linear models cannot, enabling accurate production-driven sweet spot prediction

SWEET SPOT - KRIGING MAP



- **Model Approach**
  - Gradient Boosting model used to generate the Sweet Spot Score
  - Feature importance extracted after correcting for geological redundancy
  - High-confidence zones are defined using both the Sweet Spot score and spatial uncertainty thresholds
- **Feature Importance**
  - Depth → Porosity → Permeability → Pressure

Key Takeaway: After correcting for facies-driven redundancy, production and depth emerge as the dominant controls on sweet spot, with rock quality being a factor of all aforementioned variables

TOP DRILLING LOCATIONS

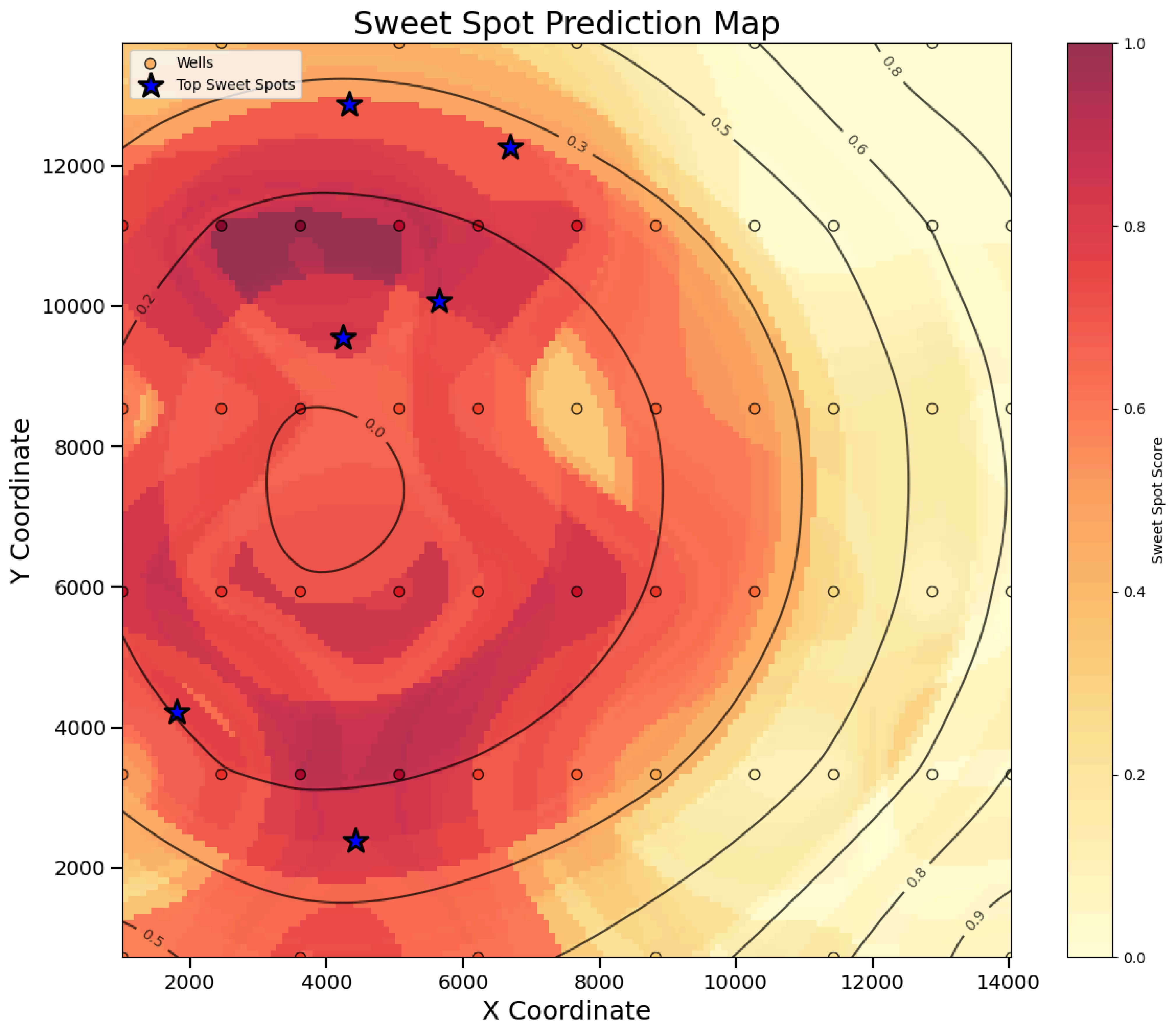
- Applied scikit-image's peak_local_max algorithm to the Sweet Spot Score surface to identify statistically significant local maxima
- Enforced a minimum spacing constraint between existing and synthetic wells to prevent duplicate targeting of the same reservoir feature
- Resulting coordinates represent spatially independent, production-optimized drilling candidates

All proposed locations satisfy both geological favorability and operational spacing constraints

FINAL RECOMMENDATION

Translating analysis into drilling decisions

TOP DRILLING LOCATIONS



- Top drilling candidates were identified as the six highest local maxima on the Sweet Spot Prediction surface

X-coordinate	Y-coordinate	Sweet Spot Score
5645	10076	0.882
4421	2385	0.835
4246	9551	0.824
1799	4220	0.763
6693	12261	0.690
4334	12872	0.686

CHALLENGES & SOLUTIONS

Challenges	Solutions
• Facies, porosity, and permeability were highly correlated, obscuring true rock quality • Sparse well spacing increased spatial interpolation uncertainty • Extreme water-production outliers distorted productivity interpretation • Pressure depletion vs. original reservoir pressure was difficult to separate • Oil vs. gas dominance complicated production-based targeting	• Developed a Facies Quality Score to isolate geology from depth and petrophysics • Applied spatial kriging with uncertainty awareness and overlaid actual well data • Removed water production from the Sweet Spot model to prevent bias • Used pressure normalization and drawdown-aware interpretation • Integrated GOR and production filtering to distinguish oil- vs. gas-driven zones

CRITICAL DISCUSSION

What we Achieved

- Built an integrated geology–machine learning workflow
- Identified a consistent sweet-spot ring pattern
- Created interpretable scores aligned with physical geology

Done Differently

- Increase well density to improve kriging accuracy
- Test alternatives to kriging (e.g., RBF interpolation)
- Explore time-dependent production trends

Lessons Learned

- Rock quality is a stronger driver of productivity than initially expected
- Pressure must be corrected for depletion to avoid bias
- Machine learning improves consistency but must remain constrained by geology
- Model outputs are best used for relative ranking, not absolute production prediction

Predictions are most reliable within the convex hull of observed wells and become less certain near field boundaries.

THANK YOU

ANY QUESTIONS?