Physics-informed ML for continuous S1 prediction in organic-rich shale

Integrating logs, mineralogy, and geochemical constraints

Roufeida Bennani[1], and Min Wang[1]

[1]School of Geosciences, China University of Petroleum (East China), Qingdao, China

EGU26-5306

Find me on LinkedIn

1. Introduction

❖ Motivation & Contribution

- S1 is critical for source rock evaluation and sweet spot identification.
- Direct measurement is expensive and spatially sparse.
- Existing models fail to capture nonlinear geological relationships.

→ **We propose a physics-informed ML framework integrating well logs, mineralogy, and geochemical constraints for continuous S1 prediction.**

2. Methodology

❖ Data

- **logs (n=10235), mineralogy (n=7606), S1** measurements **(n=357).**
- **Formation:** Upper Shahejie formation.

→ **Challenge:** Irregular and limited S1 sampling.

❖ Physics-Informed ML Workflow

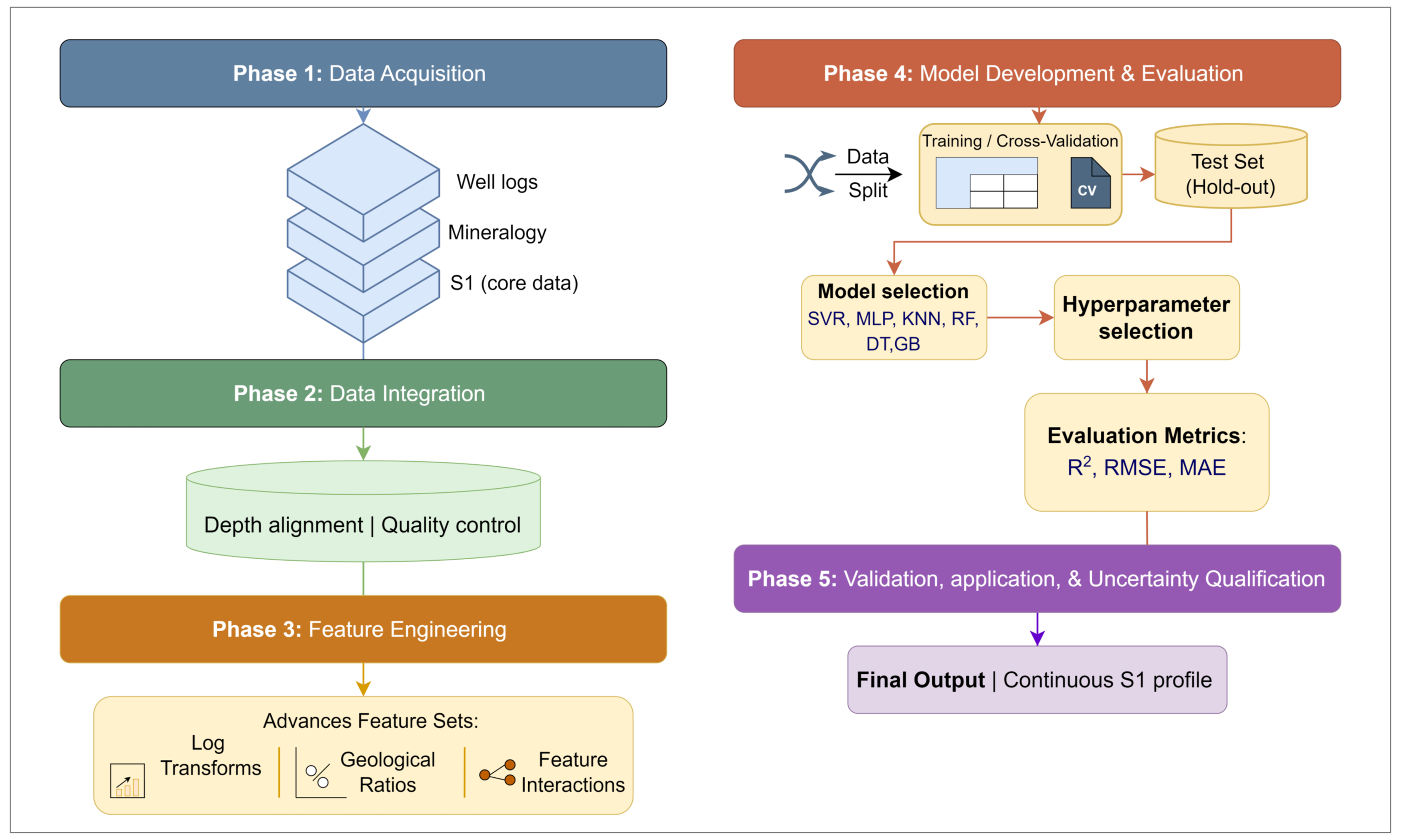

Fig.01. Physics-informed machine learning workflow for continuous S1 prediction, integrating multi-source data, depth alignment, feature engineering, and model validation.

3. Depth Alignment strategy

- Data aligned on a common grid (0.15 m).
- Interpolation (Logs/minerals).
- Nearest-neighbor (S1).

→ **Prevents data leakage.**

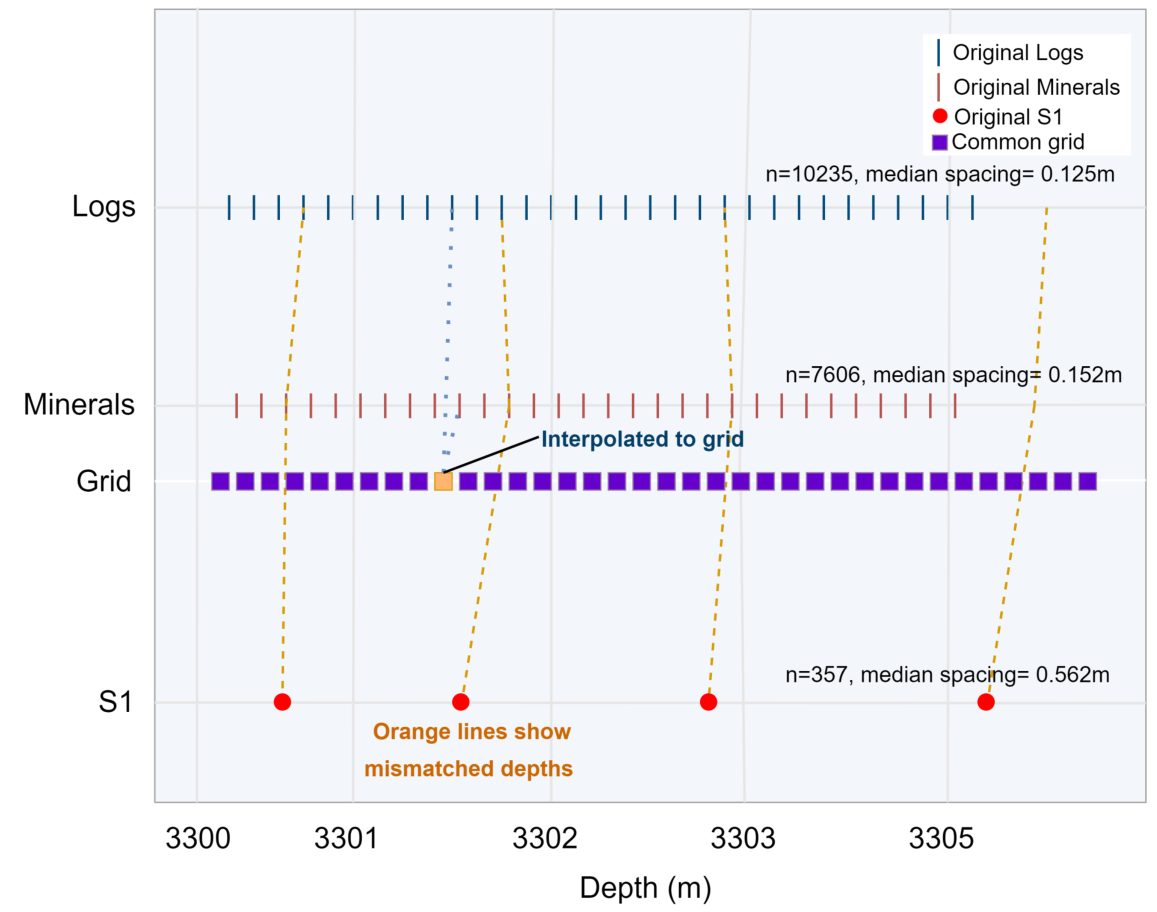

Fig. 02. Depth-alignment strategy integrating irregular core measurements with continuous log data on a common depth grid.

4. Physics-Informed Feature Engineering

- Features encode geological controls on S1, including organic richness, mineral composition, fluid-rock interactions, and depth-dependent maturity.

❖ Mineralogical Ratios

TOC/Clay, Quartz/Clay, Carbonates, Quartz fraction

➢ Capture organic enrichment and rock composition.

❖ Feature interactions

Porosity × Density, Rt × TOC, CNL × DEN

➢ Represent nonlinear fluid–rock relationships.

❖ Log Transforms

Rt log, Dolomite, DWTOC

➢ Enhances sensitivity to hydrocarbon related variations.

❖ Geological Feature

Normalized depth, Stratigraphic zones

➢ Captures maturity-driven trends.

5. Results

- The physics-informed ML model reliably reconstructs S1 with improved generalization and physically consistent feature sensitivity.

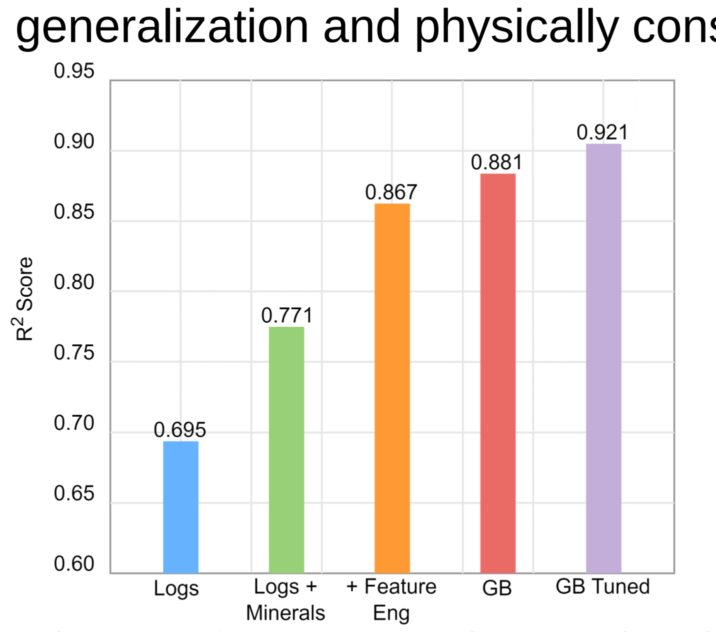

Fig.03. Performance evolution from baseline logs to physics-informed Gradient Boosting.

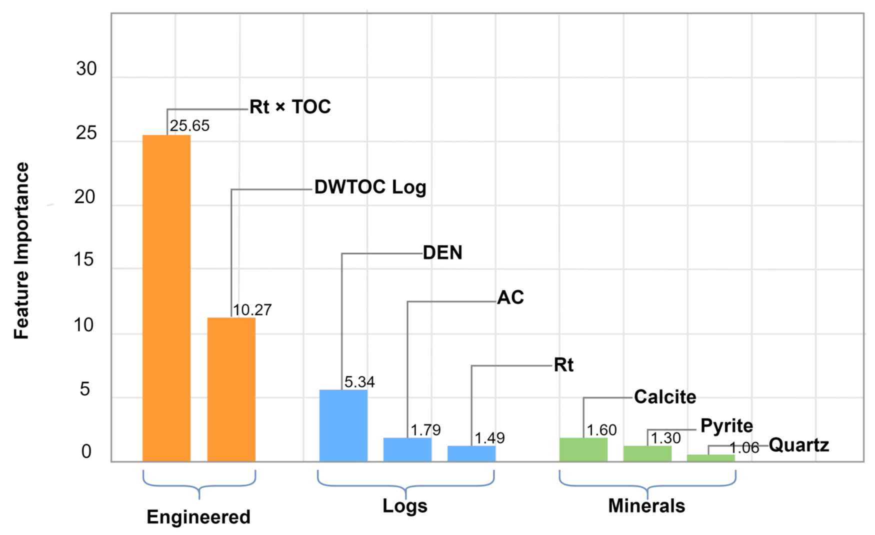

Fig.04. Feature importance analysis highlighting the dominant contribution of physics-informed features.

6. Interpretation

- Rt × TOC interaction reflects hydrocarbon saturation effects.
- Mineralogical features improve prediction beyond log-only inputs.

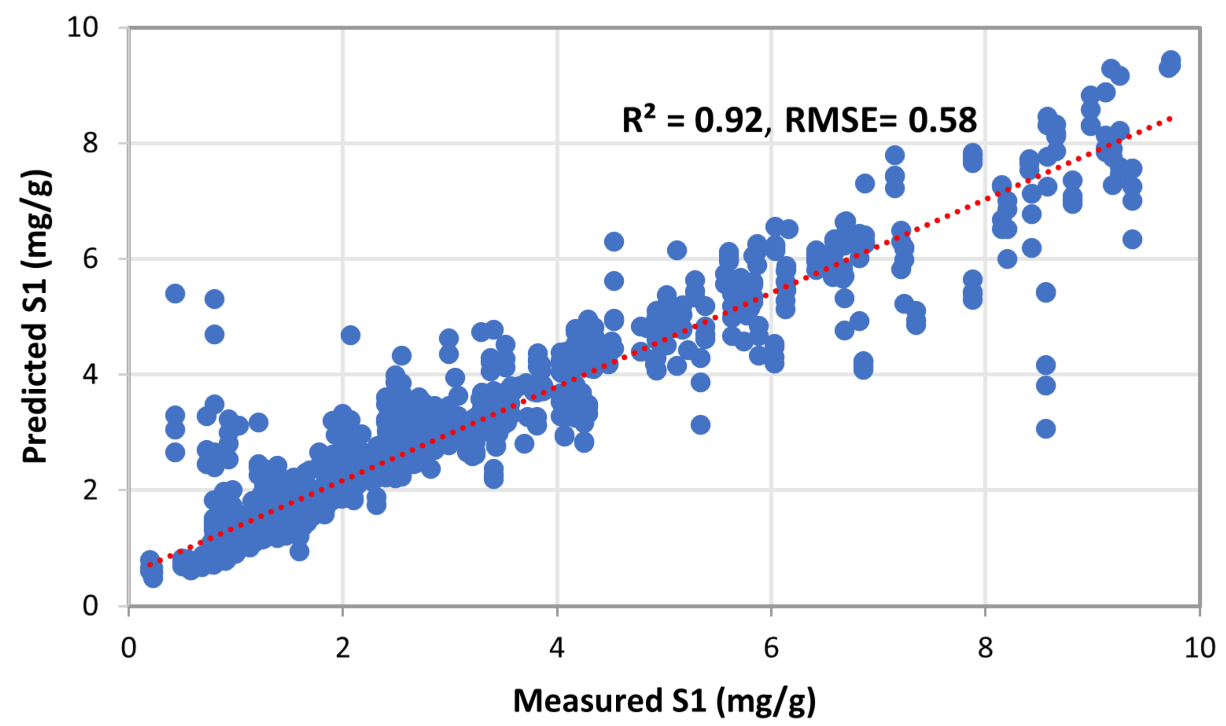

Fig.05. Predicted vs measured S1 values showing agreement and good accuracy of the GB model.

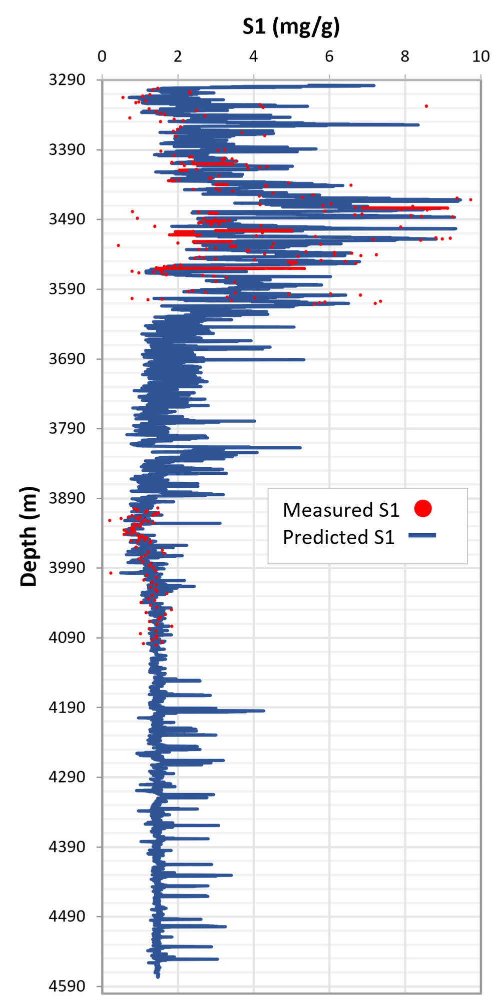

Fig. 06. Continuous S1 prediction along the depth.

7. Conclusion

- Gradient Boosting achieved high accuracy (R^2 = 0.92).
- Physics-informed features significantly improved performance.
- Interaction terms such as Rt × TOC are dominant predictors.
- Model shows stable generalization across depth.

→ **Reduces dependence on core data and supports shale reservoir evaluation.**

References: [1] Chen, F., Sun, L., Jiang, B., Huo, X., Pan, X., Feng, C., Zhang, Z., 2025. A Review of AI Applications in Unconventional Oil and Gas Exploration and Development. Energies 18, 391. https://doi.org/10.3390/en18020391

[2] Şen, Ş., 2022. Prediction of fluid oil and gas volumes of shales with a deep learning model and its application to the Bakken and Marcellus shales. Sci Rep 12, 20842. https://doi.org/10.1038/s41598-022-23406-3

[3] Sinha, U., Dindoruk, B., 2025. Review of physics-informed machine learning (PIML) methods applications in subsurface engineering. Geoenergy Science and Engineering 250, 213713. https://doi.org/10.1016/j.geoen.2025.213713