Qeios

Peer Review

Review of: "Uncertainty-Driven Modeling of Microporosity and Permeability in Clastic Reservoirs Using Random Forest"

Adedibu Sunny Akingboye[1]

1. Earth Sciences, Adekunle Ajasin University, Nigeria

The manuscript introduces a robust and innovative approach for predicting permeability and microporosity in clastic reservoirs using Random Forest, enhanced with uncertainty-driven resampling. The integration of accessible geological parameters and the aim to reduce reliance on costly laboratory techniques are commendable. However, several conceptual and methodological issues must be clarified to improve the manuscript's clarity, scientific rigor, and reproducibility. Below are major issues to address:

1. The manuscript includes the Th/K (thorium to potassium) ratio as one of the input variables but does not explain its relevance to microporosity or permeability prediction. Clarify the geochemical or petrophysical rationale for including this ratio. Is it used as a proxy for clay type or maturity? If so, support this with references and explain how it relates to fluid flow characteristics or diagenetic trends in your study area.

2. A significant methodological gap exists in the description of how the dataset expanded from 41 samples to 61,600 data points.

- Was this based on Monte Carlo simulation, bootstrapping, perturbation with Gaussian noise, or another resampling strategy?
- How was data correlation preserved during this augmentation?
- Were categorical variables (e.g., permeability class) held constant during sampling or also reclassified?

Include this information in the methodology section and consider illustrating the resampling pipeline with a schematic.

3 Despite the large synthetic dataset, the original sample size (41) is limited. Acknowledge this limitation more explicitly and discuss whether the model is potentially overfitting to noise within a small

 doi.org/10.32388/36HXBX

observational base. Was any form of outlier detection or data balance verification performed post-resampling?

4 The boundaries defining permeability categories (e.g., <15 mD = “Poor to Fair”) need clearer justification. Reference industry-standard classification schemes or regional analogues. Additionally, clarify whether the chosen thresholds align with the distribution of the original or augmented data.

5 Misclassifications were attributed to laminated/interbedded sandstone facies (Discussion section), yet these features were not incorporated numerically or modeled. Suggest encoding facies as categorical variables or using depositional texture indices to improve model accuracy. At a minimum, clarify why this information was excluded if it was available.

6 You used both 5-fold cross-validation and a holdout set, which is excellent. Please clarify whether stratification was applied during the splits, particularly given the categorical target variables.

7 Include precision, recall, and F1-score per class to complement accuracy and ROC-AUC metrics.

8 A feature importance ranking or SHAP plot from the Random Forest model is strongly encouraged to interpret the dominant predictors for each target class.

9 Improve consistency in terminology: Use “microporosity-dominated” and “macroporosity-dominated” throughout.

10 Consider streamlining repetitive content in the conclusion and discussion sections.

11 Improve the transition and integration of figure descriptions into the text body.

12 Uniform formatting of references and citation style is needed.

13 Final Recommendation: Major Revision — The manuscript presents a novel, practical, and scientifically sound approach to predicting key reservoir properties using RF models. However, clarity in methodology, explicit treatment of limitations, deeper statistical support, and refinement in language are essential before it can be considered for final publication.

Declarations

Potential competing interests: No potential competing interests to declare.