Petroleum Science 22 (2025) 2307–2311

Contents lists available at ScienceDirect

Petroleum Science

journal homepage: www.keaipublishing.com/en/journals/petroleum-science

Short Communication

Extracting useful information from sparsely logged wellbores for improved rock typing of heterogeneous reservoir characterization using well-log attributes, feature influence and optimization

David A. Wood

DWA Energy Limited, Lincoln, United Kingdom

A R T I C L E I N F O

Article history:
Received 26 January 2025
Received in revised form
27 February 2025
Accepted 18 March 2025
Available online 20 March 2025

Edited by Jie Hao and Min Li

Keywords:
Petrophysical/geomechanical rock typing
Log attribute calculations
Heterogeneous reservoir characterization
Core-well-log-seismic integration
Feature selection influences

A B S T R A C T

The information from sparsely logged wellbores is currently under-utilized in reservoir simulation models and their proxies using deep and machine learning (DL/ML). This is particularly problematic for large heterogeneous gas/oil reservoirs being considered for repurposing as gas storage reservoirs for CH_4, CO_2 or H_2 and/or enhanced oil recovery technologies. Lack of well-log data leads to inadequate spatial definition of complex models due to the large uncertainties associated with the extrapolation of petrophysical rock types (PRT) calibrated with limited core data across heterogeneous and/or anisotropic reservoirs. Extracting well-log attributes from the few well logs available in many wells and tying PRT predictions based on them to seismic data has the potential to substantially improve the confidence in PRT 3D-mapping across such reservoirs. That process becomes more efficient when coupled with DL/ML models incorporating feature importance and optimized, dual-objective feature selection techniques.

1. Introduction

The gas and oil industry typically drills exploration and discovery appraisal wells with comprehensive well-logging with whole-rocks cores of the prospective reservoir sections often recovered. However, for cost and time reasons, it does not do this for most of the field development wells that are subsequently drilled to develop its discovered resources. Most field development wells are characterized by not being cored and recording only a basic suite of petrophysical well logs. The sparse petrophysical logs recorded in such wells are typically caliper (CAL), gamma ray (GR), compressional acoustic travel time (DT), and sometimes bulk density (PB). CAL is required to quantify borehole width and conditions for casing cement programs; GR is used for determining the main lithological formation boundaries and depth calibrations for the placement of perforations; and DT and PB logs are used to calibrate velocities and acoustic impedance with the available seismic data. Around the world there are hundreds of thousands of suspended and abandoned wells, drilled during the past seventy years or more, for which such sparse logs, together with the mudlogging and drilling records, form the only available subsurface information. This huge data resource is currently underutilized at a time when there is a greater need than ever to improve the detailed characterization of large previously exploited reservoirs for repurposed applications such as enhanced oil recovery (EOR), long-term, carbon dioxide (CO_2) storage and short-term hydrogen (H_2) storage.

2. Requirements of heterogeneous reservoir characterization

Unfortunately, the sparse petrophysical well-log suites of development wells, and many unsuccessful exploration wells, are inadequate to meet the input requirements of most modern deep and machine learning (DL/ML) and reservoir simulation models applied to heterogeneous reservoirs. To adequately characterize rock formations information is required regarding lithology, geomechanical properties, pressures and sub-surface stress distributions. While laboratory analysis of whole-rock cores can provide such information with limited uncertainty for a few wellbores spread across a formation of interest, there is never enough core recovered to adequately determine these properties across the entire areas of commercial-scale gas/oil-bearing reservoirs. Petrophysical well-logs, calibrated with the available but limited core data, are now routinely used to predict those properties in non-

E-mail address: dw@dwasolutions.com.

https://doi.org/10.1016/j.petsci.2025.03.023
1995-8226/

cored wellbores. However, this is difficult in wellbores where only two or three petrophysical logs have been recorded in addition to drilling data.

Many large subsurface reservoirs, particularly those associated with carbonate and coal formations, are composed of heterogeneous and anisotropic lithologies. To capture the complexity of such reservoirs in 3-D models, reservoir-property data must be incorporated from as many spatial locations as possible. A common problem is that inadequate well-log data exists from most of the wellbores drilled to provide detailed lateral and vertical distribution maps of the heterogeneities present. The focus of industry and academia has shifted in recent years more towards the evaluation of enhanced gas and oil production developments, and the redeployment of depleted, and partially depleted, gas/oil reservoirs for CO_2 or H_2 storage. The storage capacities and efficiency of CO_2 storage is substantially influenced by reservoir heterogeneity (Sundal et al., 2014), and the same is also the case for H_2 storage (Davoodi et al., 2025). Capillary trapping of gases (CH_4, CO_2, or H_2) injected into reservoirs, which depends to an extent on the buoyancy of the gas involved relative to the prevailing reservoir fluids and can be substantially influenced by small-scale reservoir heterogeneities (Krishnamurthy et al., 2017). To understand and optimize reservoirs for such processes and usages, it is imperative to understand in detail how reservoir porosity, permeability, capillary pressure, fluid saturations and movements are impacted by heterogeneities. Hence, methods that make it possible to extract more information from sparsely logged historical wells are highly desirable.

Petrophysical rock typing (PRT) has become a popular technique for characterizing the heterogeneity of reservoirs in terms of their core-derived rock, geomechanical and formation fluid properties (Mohammadian et al., 2022). PRT, once defined, can be used to map the efficiency of fluid movements in different reservoir zones, as defined by hydraulic flow units (HFU) based on flow zone indicators (FZI) (Kadkhodaie and Kadkhodaie, 2018). Key variables used to distinguish PRT/HFU are porosity, permeability, capillary pressure, fluid saturations and irreducible fluid saturations. Extrapolating PRT/HFU across an entire reservoir using DL/ML techniques offers an effective way of improving the performance of reservoir simulation models (Djebbas et al., 2023).

High-pressure mercury intrusion capillary pressure (MICP) laboratory analysis of available cores, coupled with unsupervised cluster analysis, is now widely used to initially define PRT zones that can be correlated from well to well (Jiao et al., 2020; Saki et al., 2020). MICP data not only provides accurate determinations of capillary pressure, but it also enables fractal dimensions of the pore/fracture space to be calculated (Wang et al., 2024), which is particularly useful for certain EOR and determining the potential of a reservoir for repurposing as a CO_2/H_2 storage facility. As well as defining PRT and HFU, MICP data can be used to predict reservoir permeability, water saturation and specific reservoir geomechanical properties separately, and define PRT or HFU simply in terms of porosity versus permeability relationships (e.g., Lucia plots; Jennings and Lucia, 2003). However, due to the limited availability of cores, those core-defined PRT zones and their associated geomechanical properties require a suite of well-log data to predict them in non-cored wells. In most sparsely logged development wells, the available well-logs are insufficient to do this, making it difficult to map reservoir heterogeneities in detail throughout a reservoir.

The definitive determination of most reservoir geomechanical properties requires core and/or specialist well-logs data (e.g. shear-wave acoustic logs). Specialist logs are expensive to run and, consequently, are only recorded in relatively few wells. Nevertheless, shear-wave acoustic logs can be reconstructed from a suite of other well-logs by DL/ML methods (Wood, 2020; Rajabi et al., 2023). A number of geomechanical variables (e.g., Young's modulus, Poison's ratio) and rock mechanics stress and strength variables, obtained from uniaxial and triaxial laboratory core stress test (e.g., uniaxial compressive stress (UCS), confined compressive strength(CCS)) can also be predicted from suites of well logs (petrophysical and drilling data) calibrated with core data (Davoodi et al., 2023; Talkhouncheh et al., 2024). UCS and CCs are important reservoir properties of interest in heterogeneous reservoirs as their values vary non-linearly as porosity increase and fluid saturation change (Shi et al., 2015). These variables are also influenced by density and pore fluid saturation (Hassanvand et al., 2018). Hence, the mentioned geomechanical variables are potentially useful variables for characterizing reservoir PRT/HFU. Several geomechanical and rock mechanics variables are also useful for tying core-derived reservoir characteristics, predicted by well-logs, to seismic data and attributes used for seismic amplitude-versus-offset (AVO) analysis extracted from seismic data. Variables such as compressional-wave and shear wave acoustic impedances, compressional-wave/shear-wave velocity ratio, and Lamé-parameter relationships between those variables and rigidity (μ) and incompressibility (λ) (e.g., Lambda-mu-rho, LMR variables; Goodway, 2001) form useful links between core-derived PRT/HFU and seismic attributes. Again, wells with only sparse suites of recorded well logs are not generally involved in reservoir-wide distribution mapping of the mentioned geomechanical/AVO properties.

3. Extracting more useful data from sparsely logged wellbores

An effective way to extract more information from the few petrophysical logs recorded in sparsely logged wells is to calculate a range of mathematical attributes for those logs available. This approach was introduced to improve lithofacies predictions in clastic reservoirs using only a single recorded gamma-ray (GR) log and a selection of six well log attributes (Wood, 2022a). The technique was later expanded to calculate attributes for wells that had recorded three well logs (GR, compressional acoustic log (DTc), and bulk density (PB) and using the recorded logs plus the six attributes for each of them (21 features in total) to improve lithofacies in a heterogeneous carbonate reservoir (Wood, 2022b). The technique has subsequently been successfully applied to predict other rock characteristics, e.g., brittleness (Wood, 2023) and total organic carbon (Wood, 2025). The success to date with the well-log-attribute technique suggests that it is suitable for use to predict core-measured and clustered PRT/HFU and a range of geomechanical rock properties from a sparse suite of well logs. Such predictions could then be used to establish PRT/HFU mapping across heterogeneous reservoirs incorporating data from a larger number of wellbores and correlating PRT/HFU clusters to seismic attributes. This offers the potential to substantially reduce the uncertainty that is currently associated with reservoir simulation models of large, heterogeneous/anisotropic reservoirs.

Fig. 1 describes a workflow for calculating the attributes of recorded well-logs from sparsely logged wells and using them as input variables for DL/ML models to reduce the uncertainty in extrapolating PRT/HFU spatially across extensive, sparsely logged, heterogeneous reservoirs. The purpose and value of calculating well-log attributes to compliment the data available from just a few recorded well logs is particularly relevant to heterogeneous reservoir formations.

The six calculated well-log attributes extracted in applications of the method to date, involve three related to the recorded well-log's derivative and simple moving average of the first derivative, and

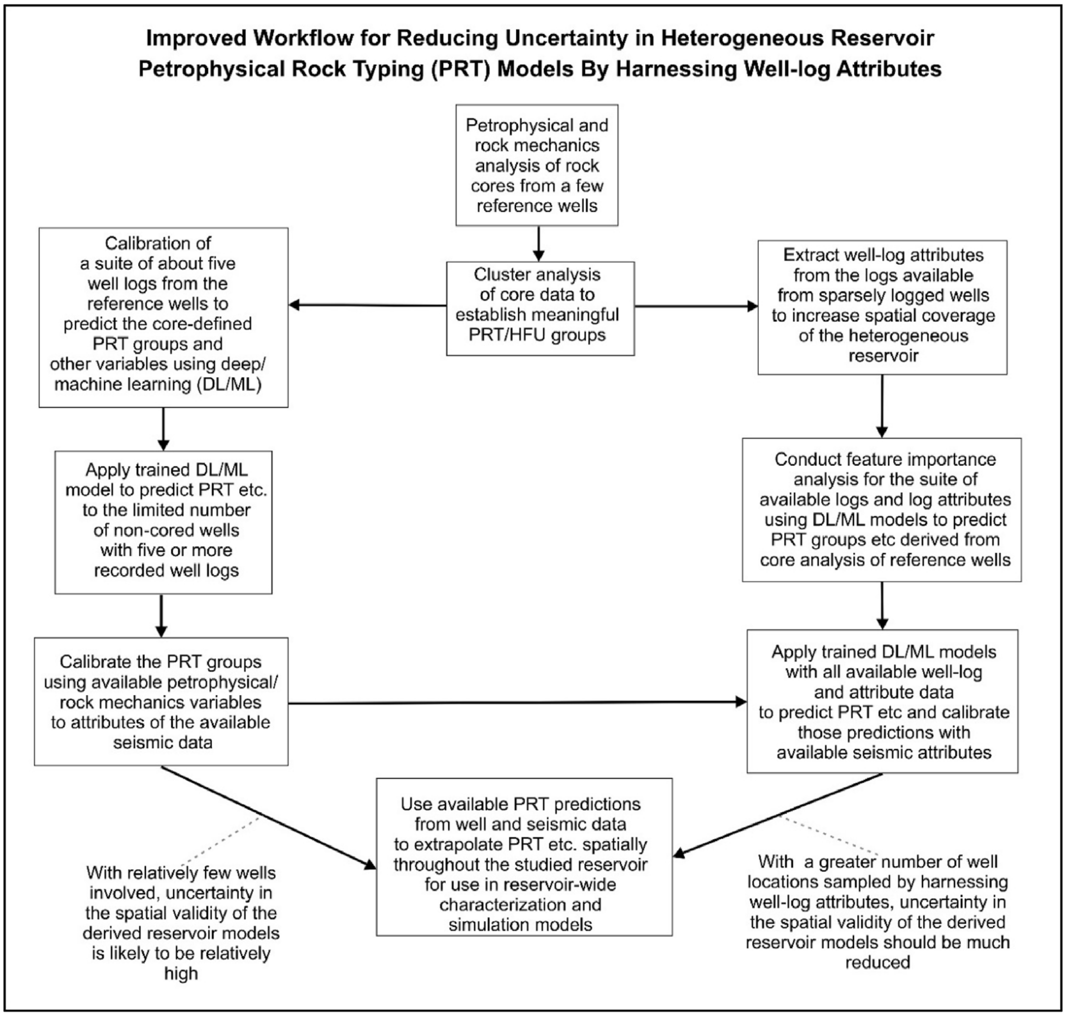

Fig. 1. Workflow diagram for incorporating well-log attributes into PRT characterization of heterogeneous reservoirs. PRT refers to petrophysical rock type; HFU refers to hydraulic flow unit.

three related to the recorded well-log's volatility. These attributes can be calculated rapidly using simple mathematical formula (Wood, 2022a), and can be applied to any well-log recorded over a continuous depth interval. The formulas include some flexibility that enable them to be adjusted to suit the degree of variability observed in a recorded well log. There is also potential to calculate additional attributes defined by simple mathematical relationships. For instance, weighted moving average provides a flexible way of assigning more, or less, weight to certain values in a specified weighted-average interval. Alternatively, exponentially weighted moving averages or centralized moving averages could also be exploited as attributes. Each of these attributes captures slightly different periodic characteristics from the recorded well logs. DL/ML models can exploit these details when predicting, lithofacies, geomechanical variables, PRT/HFU and other required well logs, such as shear-wave acoustic travel time (DTs), when calibrated with data from cores and/or specialist well-logs from a reference well from a studied reservoir,

If a sparsely logged well has only recorded logs for DTc, GR and PB available then calculating the six attributes for each of recorded well logs plus the recorded log can generate twenty-one features for evaluation as input variables to DL/ML models. It is likely, depending on the target variable to be predicted and the range reservoir characteristics contributing to the PRT/HFU distinguished, that only some of those features will substantially influence the prediction models. Hence, it is worthwhile calculating multiple well-log attributes and then conducting feature-influence and selection analysis with respect to specific target variable predictions. Such analysis establishes which combination of the well-log attributes calculated and the recorded logs should be used to predict each target variable of interest.

4. Feature influence and selection of well-log attributes

There are several methods available to conduct feature importance analysis on DL/ML datasets with a view to reducing their dimensionality. Five models that generate different perspectives on feature importance are:

- Least absolute shrinkage and selection operator(LASSO) linear regression
- Mean decline in impurity (MDI) from tree-ensemble ML models
- Permutation feature importance (PFI)
- Local Interpretable model-agnostic explanations (LIME)
- Shapley additive explanations (SHAP)

LASSO provides insight into feature selection because its L1 regularization acts assign zero coefficients to the least influential features (MuthukrishnanRohini and Rohini, 2016). However, it linear/parametric assumptions are not able to capture the non-linear relationships typically associated with heterogeneous reservoir dataset.

MDI is extracted from tree-ensemble ML algorithms by assessing the frequency of involvement of each feature in the splitting criteria applied at each tree branch (Scornet, 2021). However, MDI feature values can be skewed by model scaling biases and over-fitting of model training datasets.

PFI randomly shuffles the data value sequence of a dataset, one feature at a time and assesses its impacts on prediction performance metrics considering all data records. Developed originally for the random forest algorithm (Breiman, 2001) it has subsequently been generalized (Fisher et al., 2018) and can be executed rapidly for most datasets.

LIME involves introducing disturbances at certain points in the feature space and monitoring their impacts on prediction performance, by fitting linear models locally within the feature space and weighting the impacts of each feature (Ribeiro et al., 2016). This approach has some limitations with non-linear datasets, and unlike SHAP does not identify the impacts of each data point, but it can be executed more rapidly than SHAP (Mane et al., 2024).

SHAP quantifies the degree to which each input feature influences a model's predictions for every data point in a dataset (Lundberg and Lee, 2017). It provides comprehensive model-specific feature importance rankings (Lundberg et al., 2019). However, it is complex and time-consuming to run on large datasets. This downside has led recently to the development of the simplified SHAPG version (Zhao et al., 2024) which executes more rapidly.

It is prudent to consider the feature importance results from at least two of the mentioned methods and consider that LASSO and LIME make parametric assumptions which may not work well with heterogeneous reservoir datasets. Although these feature importance techniques provide useful guidance regarding feature selection for ML models, they do not establish exactly which combination of features specific ML models would use to make their optimum predictions. Hence, the is value in combining optimizers with ML model to make such selections. Using multi-objective optimizers to optimize two targets, prediction performance of the dependent variable and the number of features (Wood, 2022c). The output from such optimizers displayed as a Pareto frontier comparing the performance of different feature combinations.

The incorporation of such information derived from well-log attributes, feature-influence analysis, and optimized feature selection has the potential to substantially improve reservoir characterization of producing reservoirs for potential repurposing as EOR and gas storage facilities. The proposed methodology is particularly appropriate for large, extensively drilled, heterogeneous reservoirs in which only a few wells have been cored, and the log suites recorded in most wellbore are limited to less than about five recorded well logs.

Declaration of competing interest

The author declares that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32. https://doi.org/10.1023/A:1010933404324.

Davoodi, S., Mehrad, M., Wood, D.A., Rukavishnikov, V.S., Bajolvand, M., 2023. Predicting uniaxial compressive strength from drilling variables aided by hybrid machine learning. Int. J. Rock Mech. Min. Sci. 170, 105546. https://doi.org/10.1016/j.ijrmms.2023.105546.

Davoodi, S., Al-Shargabi, M.A., Wood, D.A., Longe, P.O., Mehrad, M., Rukavishnikov, V.S., 2025. Underground hydrogen storage: a review of technological developments, challenges, and opportunities. Appl. Energy 381, 125172. https://doi.org/10.1016/j.apenergy.2024.125172.

Djebbas, F., Ameur-Zaimeche, O., Kechiched, R., Heddam, S., Wood, D.A., Zohreh, M., 2023. Integrating hydraulic flow unit concept and adaptive neuro-fuzzy inference system to accurately estimate permeability in heterogeneous reservoirs: case study Sif Fatima Oilfield, Southern Algeria. J. Afr. Earth Sci. 206, 105027. https://doi.org/10.1016/j.jafrearsci.2023.105027.

Fisher, A., Rudin, C., Dominici, F., 2018. All models are wrong, but many are useful : learning a variable's importance by studying an entire class of prediction models Simultaneously (arXiv:1801.01489). arXiv. https://doi.org/10.48550/arXiv.1801.01489.

Goodway, W, 2001. AVO and Lamé constants for rock parameterization and fluid detection. Recorder (Canadian Society of Exploration Geophysicists) 26 (6) Online publication. https://csegrecorder.com/articles/view/avo-and-lame-constants-for-rock-parameterization-and-fluid-detection (Accessed 18 January 2025).

Hassanvand, M., Moradi, S., Fattahi, M., Zargar, G., Kamari, M., 2018. Estimation of rock uniaxial compressive strength for an Iranian carbonate oil reservoir: modeling vs. artificial neural network application. Petrol. Res. 3 (4), 336–345. https://doi.org/10.1016/j.ptlrs.2018.08.004.

Jennings, J.W., Lucia, F.J., 2003. Predicting permeability from well logs in carbonate with a link to geology for interwell permeability mapping. SPE Reservoir Eval. Eng. 6, 215–225. https://doi.org/10.2118/84942-PA.

Jiao, L., Andersen, O., Zhou, J., Cai, J., 2020. Applications of mercury intrusion capillary pressure for pore structures: a review. Capilliarity 3 (4), 62–74. https://doi.org/10.46690/capi.2020.04.02.

Kadkhodaie, A., Kadkhodaie, R., 2018. A review of reservoir rock typing methods in carbonate reservoirs: relation between geological, seismic, and reservoir rock types. Iranian J. Oil Gas Sci. Technol. 7 (4), 13–35.

Krishnamurthy, P.G., DiCarlo, D.A., Bryant, S.L., 2017. Understanding the effects of small-scale heterogeneity on buoyancy-driven CO_2 migration for capillary-trapped storage capacity estimation. Energy Proc. 114, 2037–2044. https://doi.org/10.1016/j.egypro.2017.03.1637.

Lundberg, S., Lee, S.-I., 2017. A Unified approach to interpreting model predictions. arXiv 1705, 07874. https://doi.org/10.48550/arXiv.1705.07874.

Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.-I., 2019. Explainable AI for trees: from local explanations to global understanding. arXiv:1905.04610. https://doi.org/10.48550/arXiv.1905.04610.

Mane, D., Magar, A., Khode, O., Koli, S., Bhat, K., Korade, P., 2024. Unlocking machine learning model decisions : a comparative analysis of LIME and SHAP for enhanced interpretability. J. Electrical Syst. 20 (2s), 598–613. https://doi.org/10.52783/jes.1480.

Mohammadian, E., Kheirollahi, M., Ostadhassan, M., Sabet, M., 2022. A case study of petrophysical rock typing and permeability prediction using machine learning in a heterogenous carbonate reservoir in Iran. Sci. Rep. 12, 4505. https://doi.org/10.1038/s41598-022-08575-5.

Muthukrishnan, M., Rohini, R., 2016. LASSO: a feature selection technique in predictive modeling for machine learning. IEEE International Conference on Advances in Computer Applications (ICACA). Coimbatore, India, pp. 18–20. https://doi.org/10.1109/ICACA.2016.7887916.

Rajabi, M., Hazbeh, O., Davoodi, S., Wood, D.A., Tehrani, P.S., Ghorbani, H., Mehrad, M., Mohamadian, N., Rukavishnikov, V.S., Radwan, A.E., 2023. Predicting shear wave velocity from conventional well logs with deep and hybrid machine learning algorithms. J. Pet. Explor. Prod. Technol. 13, 19–42. https://doi.org/10.1007/s13202-022-01531-z.

Ribeiro, M.T., Singh, S., Guestrin, C., 2016. Why should I trust you? explaining the predictions of any classifier. arXiv 1602, 04938. https://doi.org/10.48550/arXiv.1602.04938.

Saki, M., Siahpoush, S., Khaz'ali, A.R., 2020. A new generalized equation for estimation of sandstone and carbonate permeability from mercury intrusion porosimetry data. J. Pet. Explor. Prod. Technol. 10, 2637–2644. https://doi.org/10.1007/s13202-020-00900-w.

Scornet, E., 2021. Trees, forests, and impurity-based variable importance. arXiv.2001.04295. https://doi.org/10.48550/arXiv.2001.04295.

Shi, X., Meng, Y., Li, G., Li, J., Tao, Z., Wei, S., 2015. Confined compressive strength model of rock for drilling optimization. Petroleum 1 (1), 40–45. https://doi.org/10.1016/j.petlm.2015.03.002.

Sundal, A.V., Jahren, J.S., Aagaard, P., Oelkers, E.H., 2014. Mineral trapping of CO_2 in a natural CO_2 field: implications for CO_2 storage. Energy Proc. 63, 5362–5369. https://doi.org/10.1016/j.egypro.2014.11.536.

Talkhouncheh, M.Z., Davoodi, S., Wood, D.A., Mehrad, M., Rukavishnikov, V.S., Bakhshi, R., 2024. Robust machine-learning predictive models for real-time determination of confined compressive strength of rock using mudlogging data. Rock Mech. Rock Eng. 57, 6881–6907. https://doi.org/10.1007/s00603-024-03859-w.

Wang, F., Hue, H., Wang, L., Zhu, W., 2024. Permeability prediction and rock typing for unconventional reservoirs using high-pressure mercury intrusion and fractal analysis. Energy & Fuels 38 (22), 1–11. https://doi.org/10.1021/acs.energyfuels.4c03123.

Wood, D.A., 2020. Bakken stratigraphic and type well log learning network exploited to predict and data mine shear wave acoustic velocity. J. Appl. Geophys. 173, 103396. https://doi.org/10.1016/j.jappgeo.2019.103936.
Wood, D.A., 2022a. Gamma-ray log derivative and volatility attributes assist facies characterization in clastic sedimentary sequences for formulaic and machine learning analysis. Adv. Geo-Energy Res. 6 (1), 69–85. https://doi.org/10.46690/ager.2022.01.06.
Wood, D.A., 2022b. Carbonate/siliciclastic lithofacies classification aided by well-log derivative, volatility and sequence boundary attributes combined with machine learning. Earth Sci. Inform. 15 (3), 1699–1721. https://doi.org/10.1007/s12145-022-00829-0.
Wood, D.A., 2022c. Optimized feature selection assists lithofacies machine learning with sparse well log data combined with calculated attributes in a gradational fluvial sequence. Artificial Intell. Geosci. 3, 132–147. https://doi.org/10.1016/j.aiig.2022.11.003.
Wood, D.A., 2023. Geomechanical brittleness index prediction for the Marcellus Shale exploiting well-log attributes. Results Eng. 17, 100846. https://doi.org/10.1016/j.rineng.2022.100846.
Wood, D.A., 2025. Regression models to estimate total organic carbon (TOC) from well-log data. Chapter 2 in, Implementation and Interpretation of Machine and Deep Learning to Applied Subsurface: Problems Prediction Models Exploiting Well-Log Information. Elsevier, pp. 43–77. https://doi.org/10.1016/B978-0-443-26510-5.00002-5.
Zhao, C., Liu, J., Parilina, E., 2024. ShapG : new feature importance method based on the Shapley value. arXiv 2407, 00506. https://doi.org/10.48550/arXiv.2407.00506.