

RESEARCH

Open Access



Geologically informed machine learning for neutron porosity reconstruction from legacy well logs: implications for high-temperature reservoir thermal energy storage

Amr Talaat Tolba^{1*} , Xheni Garipi^{1,2}, Florian Bauer¹, Thomas Kohl¹ and Eva Schill³

*Correspondence:

Amr Talaat Tolba
amr.talaat@kit.edu

¹Institute of Nuclear Waste Disposal,
Karlsruhe Institute of Technology
Geoenergy,
Eggenstein-Leopoldshafen
D-76344, Germany

²Institute of Applied Geosciences,
Technical University of Darmstadt,
Schnittspahnstrasse 9,
Darmstadt 64287, Germany

³Lawrence Berkeley National
Laboratory, 1 Cyclotron Road,
Berkeley 94720, CA, USA

Abstract

High-temperature thermal energy storage in subsurface reservoirs (HT-RTES) can help balance seasonal heat demand and supply by storing excess thermal energy in subsurface reservoirs. Mature hydrocarbon fields are promising storage targets because their reservoir properties are relatively well constrained; however, many legacy fields lack modern petrophysical data. In the Leopoldshafen oil field (northern Upper Rhine Graben), complete log coverage is limited to legacy self-potential (SP) and short-normal (SN) resistivity logs, whereas neutron porosity (NPHI) logs are available only in two wells (LH-5 A and Stutensee-1). Here, we present an integrated workflow to reconstruct NPHI in legacy wells using machine-learning methods. SP and SN logs from 24 boreholes were digitized, standardized, and cleaned to remove acquisition artefacts. Lithological classes were derived from SP–SN cross-plot clustering and used as a rock-physics-guided lithology flag. Five supervised algorithms were tested using two input configurations: Model 1, based on scaled SP and SN logs, and Model 2, which additionally included the lithology flag. Random Forest with Model 2 yielded the best performance in the training well Stutensee-1 and generalized well to the independent validation well LH-5 A. The main prediction errors occurred in gas-bearing intervals, but a lithology-specific post-processing correction markedly improved agreement with measured NPHI. The results show that lithological information substantially improves porosity prediction from sparse legacy logs and enables geologically consistent NPHI reconstruction for reservoir characterization at Leopoldshafen.

Keywords Machine learning, Neutron porosity, High-temperature reservoir thermal energy storage, Leopoldshafen oil field

Introduction

Reservoir Thermal Energy Storage (RTES) represents a subsurface thermal storage concept in which heat is injected into and subsequently recovered from permeable geological reservoirs at intermediate to deep depths (Lee 2010; Witter et al. 2026). In contrast to

conventional shallow Underground Thermal Energy Storage (UTES), including Aquifer Thermal Energy Storage (ATES), RTES target deeper geological formations where higher storage temperatures and larger storage volumes may be achievable (Fleuchaus et al. 2020; Zhang et al. 2026). High-Temperature RTES (HT-RTES) refers here to reservoir storage systems designed to operate at temperatures exceeding those typically considered in conventional shallow ATES applications (Fleuchaus et al. 2020). Such systems may utilize depleted hydrocarbon reservoirs or deep saline formations, where pre-existing geological and petrophysical information can reduce exploration uncertainty and support reservoir-scale assessment (Stricker et al. 2020; Zhang et al. 2026).

In this study, High-Temperature Reservoir Thermal Energy Storage (HT-RTES) refers to reservoir thermal energy storage systems designed for injection temperatures exceeding approximately 100 °C, consistent with the temperature range considered for high-temperature subsurface heat storage applications (Banks et al. 2021; Stricker et al. 2024). Injection temperatures of up to 250 °C are targeted near Bakersfield (California) where excess solar thermal energy will be stored for electricity generation (Akindipe et al. 2024). The strategic repurposing of depleted hydrocarbon reservoirs (Stricker et al. 2020) for HT-RTES presents a “brownfield” advantage: it leverages existing infrastructure and capitalizes on historical geological data, thereby mitigating the exploration risks associated with greenfield geothermal projects (Schill et al. 2024). However, repurposing as an HT-RTES site bears challenges such as porosity changes over time during operation (Rosenbrand et al. 2014). The DeepStor research infrastructure at the Karlsruhe Institute of Technology (KIT) is designed to investigate the scientific and technological feasibility of utilizing a depleted hydrocarbon reservoir as a high-temperature seasonal subsurface heat storage system, with planned injection temperatures of up to 140 °C. The research infrastructure specifically addresses the coupled thermo-hydro-chemical processes associated with high-temperature aquifer thermal energy storage, including critical reactive transport phenomena that may influence reservoir performance, operational sustainability, and long-term storage efficiency (Banks et al. 2021). More broadly, DeepStor seeks to advance the fundamental understanding and practical implementation of converting depleted hydrocarbon fields into large-scale seasonal thermal energy storage facilities. The reliable characterization of porosity, permeability, and geomechanical properties can substantially reduce uncertainty in the assessment and design of HT-RTES systems, particularly for simulating thermal plume migration, fluid flow, or caprock integrity under cyclic thermal loading (Stricker et al. 2020, 2024).

The depleted Leopoldshafen oil field in the Upper Rhine Graben (Fig. 1) retains a dense legacy dataset from boreholes and a 3D seismic survey. Geomechanical measurements on analogue samples add to this dataset (David et al. 2026). The reservoir rock is a calcareous sandstone interbedded in the Oligocene Froidefontaine to Niederrödern Formations. The proven productive porosity and permeability combined with borehole temperatures in the order of about 80 °C (Banks et al. 2021) makes these sandstone layers good candidates for seasonal heat injection and extraction.

The repurposing of the Leopoldshafen field for HT-RTES typifies the modern data-scarcity challenge inherent in mature or depleted assets: the majority of wells here were drilled between 1950 and 1963, a period predating modern NPHI logging suites. The dataset that is available in all wells of the field includes electric self-potential (SP) and

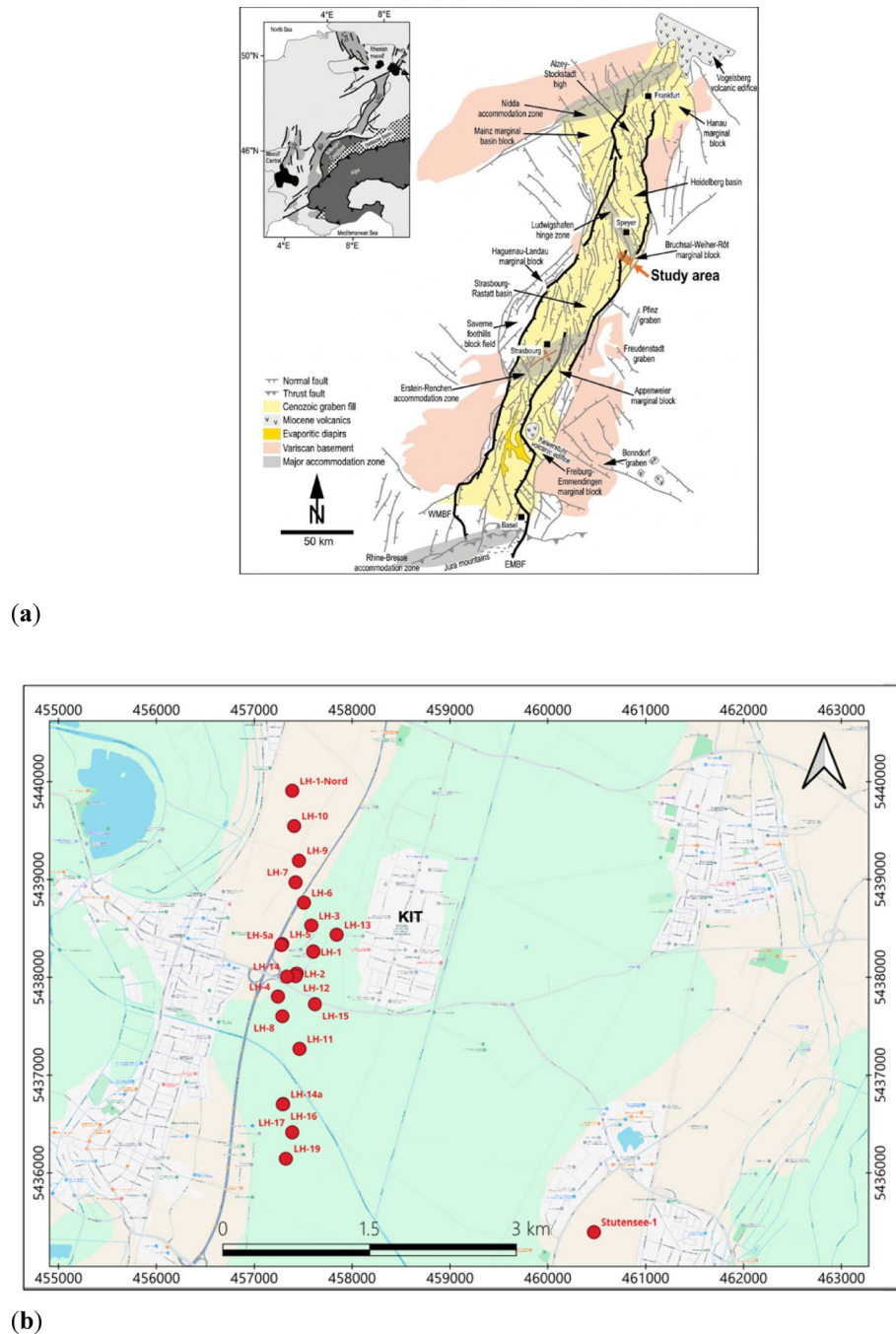


Fig. 1 (a) Location and geological context of the depleted Leopoldshafen oil field in the central Upper Rhine Graben (URG), southwest Germany (modified after Bauer et al. 2025). (b) Location of the Leopoldshafen oil field wells LH-X and the adjacent Stutensee-1 well. The campus north of Karlsruhe Institute of Technology (KIT), where the research infrastructure DeepStor will be installed, is located to the east of the Leopoldshafen oil field

short normal (SN) resistivity logs, while NPHI data is restricted to only two wells drilled in the early 1980s, LH-5 A and the adjacent Stutensee-1 wells.

In permeable sandstone units, electric potential differences between the borehole mud filtrate and formation water typically generate a measurable deflection relative to the shale baseline. Where fresher mud filtrate commonly invades more saline formation water, this response is often manifested as a pronounced negative deflection, which is

diagnostic of sand-prone intervals (Asquith and Krygowski 2004; Rider and Kennedy 2011). Conversely, shale units generally exhibit subdued SP responses due to their low permeability and high clay content, which suppress electrochemical filtration effects. Complementarily, the SN log captures variations in electrical resistivity in the near-borehole region. Sandstones typically display higher resistivity than shales as a consequence of reduced clay-bound water and, in some cases, the presence of hydrocarbons or moderately saline formation fluids. In the Leopoldshafen wells, the identified clay minerals are predominantly non-swelling, although their cation exchange capacity was not specifically investigated. Although the SN measurement is influenced by invasion effects, it remains a robust discriminator between clean sands and conductive shale intervals when interpreted in conjunction with SP data (Rider 2006).

Empirical relationships between the electric properties and porosity are described by (Archie 1942). Linear empirical relationships, often fail to extract reliable porosity estimates from such limited and heterogeneous legacy datasets (Anifowose et al. 2017). Furthermore, in multi-well petrophysical studies involving multi-decadal drilling campaigns, data standardization is a pre-requisite since raw log responses are often incomparable due to variations in tool calibration, mud filtrate properties, and logging vendors over time (Rider and Kennedy 2011).

The lack of standardized tool calibration and the non-linear response of resistivity to complex lithologies necessitate a more advanced approach. Neural network-based approaches were successfully applied to estimate subsurface temperature from geophysical measurements in the Soultz-sous-Forêts geothermal field in France (Spichak et al. 2015). Machine learning (ML) in general has emerged as a transformative tool in formation evaluation, capable of reconstructing missing log curves by learning high-dimensional non-linear mappings between electrical properties and reservoir parameters (Anifowose et al. 2017). Yet purely data-driven models are prone to overfitting and physically inconsistent predictions when trained on limited or unevenly distributed data — a condition that applies directly to the present dataset, where training is restricted to a single well and sandstone facies are strongly under-represented (Karniadakis et al. 2021; Willard et al. 2023). To address this, recent scholarship advocates for physics-guided ML workflows, where geological constraints—such as lithological classification derived from rock physics—are integrated into the learning architecture to ensure interpretability and reduce overfitting (Chakraborty et al. 2024; Suleymanov et al. 2023; Zhang et al. 2025).

With the objective to utilize legacy electrical measurements to reconstruct the NPHI logs using ML techniques, this study developed and applied an integrated workflow specifically designed to unlock the value of legacy analog well data in the Leopoldshafen field. By bridging the gap between mid-20th-century analog logs and modern digital reservoir modeling needs, we aim to reduce uncertainty in HT-RTES development where the principal value of reconstructing porosity across the legacy well network is not merely the generation of a missing log curve but the reduction of uncertainty in the spatial distribution of reservoir properties. Porosity controls pore volume and therefore influences the volume available for fluid circulation and heat storage. When combined with reservoir thickness, areal extent, fluid properties, and thermal parameters, the reconstructed porosity distribution can be propagated into storage-capacity and thermal-flow models. (Heldt et al. 2024, 2026; Stricker et al. 2020).

Geological setting

The study area is located in the Upper Rhine Graben within Southwest Germany, the Upper Rhine Graben is an approximately north–south-oriented Cenozoic rift basin in southwestern Germany that forms part of the European Cenozoic Rift System (URG; Fig. 1a), a north–south trending basin extending for approximately 300 km in length and up to 40 km in width and forming part of the European Cenozoic Rift System (Schumacher 2002). Graben development began in the late Eocene under a regional extensional stress regime (Hinsken et al. 2007, 2011), resulting in a complex structural architecture characterized by rotated fault blocks and tilted sedimentary successions. The graben is bounded by major trans-tensional faults and filled with up to about 4,000 m of Cenozoic sediments (Schumacher 2002). Within this tectonic framework, the Leopoldshafen reservoir forms an eastward-dipping, fault-bounded structural trap, with hydrocarbon production occurring in here from 1957 to 1986.

The reservoir comprises a sandstone succession within the Froidefontaine and Niederrödern Formations, both of which belong to the Oligocene sedimentary sequence (Fig. 2). In the central URG, the Rupelian Froidefontaine Formation is subdivided into the Karlsruhe Subformation (Cyrena Marls & Meletta beds) and the Rupel Clay, which overlies the upper, mid, and low “beds” of Pechelbronn Formation, a unit that contains a basin-wide seismic marker horizon. The Froidefontaine Formation is dominated by grey to dark grey, locally bituminous marls and clayey marlstones. Marine to freshwater fossil assemblages occur throughout the formation, and calcareous sandstone beds are present as intercalations. The Karlsruhe Subformation is dominated by grey marlstones with brackish to freshwater fossil assemblages and local sandstone interbeds. From base to top, three units are recognized within the Karlsruhe Subformation (LGRB, 2016). The Lower Karlsruhe Marls form the basal unit and are slightly sandy. Above this, the Karlsruhe Calcareous Sandstones (Meletta sandstones). consist of two calcareous sandstone beds that extend across the entire Leopoldshafen field These were deposited in a shallow-marine to deltaic environment under syn-rift tectonic influence and are interbedded with marl and shale layers. The Upper Karlsruhe Marls (Cyrena Marls) complete the succession and are characterized by fine-grained marlstones with only minor sandstone interbeds (Geyer et al. 2023).

The overlying Niederrödern Formation was deposited beginning at approximately 28 Ma in a meandering fluvial system (Tolba et al. 2026) and comprises several distinct facies associations. The sandstones are fine-grained, micaceous, and calcareous, and locally show distinct oil impregnation; they occur together with marlstone interbeds. Cumulative sandstone thickness ranges from 18 to 65 m, at depths of 700–1,040 m within the study area. Directly measured porosity and permeability values obtained from plugs of Leopoldshafen and Stutensee 1 well, which are considered particularly valuable for reservoir characterization, as they provide direct petrophysical data on the reservoir interval of the Leopoldshafen oil field, making the Niederrödern Formation a suitable target for high-temperature reservoir thermal energy storage (HT-RTES).

Data and methods

To obtain optimal results in this study we apply a workflow that compares different supervised machine learning (ML) algorithms. It includes the following steps:

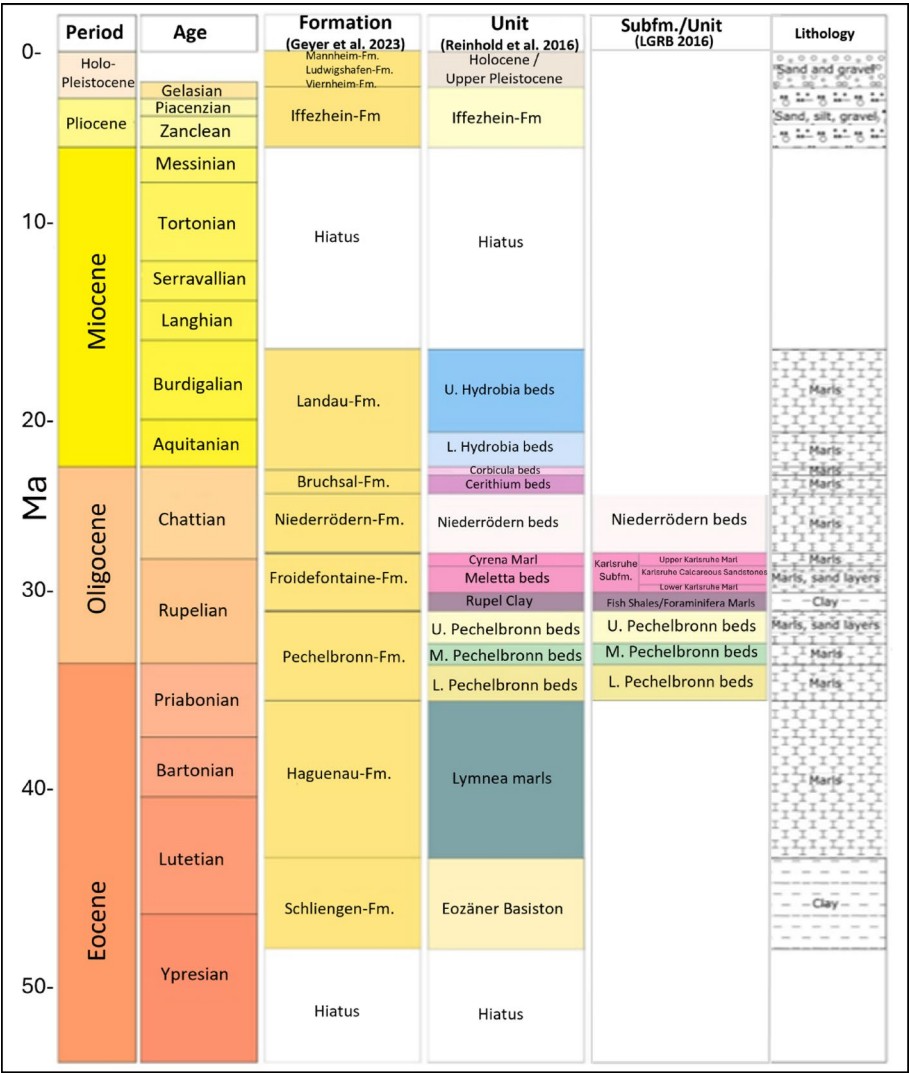


Fig. 2 Stratigraphic units and lithological representation of the Neogene and Paleogene sedimentary sequences in the study area (modified after Bauer et al. 2025)

1. The implementation of a rigorous digitalization and standardization protocol for legacy SP and SN logs to harmonize data across decades of drilling.
2. Flagging clusters of scaled SP and SN responses on cross-plots that characterize the lithology model (sand vs. shale).
3. The comparison of supervised ML algorithms (including Random Forest, Gradient Boosting, and XGBoost) that integrate the lithological flags to predict Neutron Porosity.
4. The validation of the model using the adjacent Stutensee-1 well as a training vehicle and a blind-test on the well LH-5 A.

Logging data and pre-processing

The available data inventory is predominantly composed of analogue legacy well logs and comprises records from 24 boreholes. Of these, 23 wells are located within the depleted Leopoldshafen oil field, whereas the Stutensee-1 well is situated in the adjacent tectonic block to the east (Fig. 1b). Most of the wells were drilled and logged during the principal

exploration and development phase of the Leopoldshafen oil field between 1956 and 1961. In contrast, the two wells selected as control wells, LH-5 A and Stutensee-1, were drilled and logged during the 1980s and therefore provide the most suitable datasets for the present study.

The selection of Stutensee-1 is motivated by its geological and structural setting rather than solely by its geographic proximity to the Leopoldshafen field. The well is located within the same structural domain for the Leopoldshafen oil field between the major Leopoldshafen and Stutensee fault systems and penetrates the same target reservoir interval as the Leopoldshafen wells (Fig. 1b). Consequently, Stutensee-1 provides an independent but geologically relevant dataset for evaluating the transferability of the developed petrophysical relationships beyond the immediate well control within the oil field. LH-5 A provides an independent validation of model performance directly within the Leopoldshafen field. This distinction is particularly important for the subsequent application of the predicted petrophysical properties to the DeepStor project, where the objective is to develop a reservoir-scale characterization framework for the assessment and potential deployment of High-Temperature Reservoir Thermal Energy Storage (HT-RTES).

The paucity of modern logging suites represents a critical challenge in the evaluation of the storage capacity in this brownfield environment. The pervasive data availability is limited to SP and SN (Fig. 3). Gamma ray (GR) logs are available only in nine boreholes. Advanced porosity logs, specifically NPHI and density (RHOB), are restricted to the two wells LH-5 A and Stutensee-1. Ensuring the applicability of the predictive model across the entire field, the input feature space was restricted to the ubiquitous SP and SN logs.

During the pre-processing stage, extreme values were objectively identified using statistical outlier detection criteria and subsequently excluded to reduce noise and prevent bias in downstream analyses. This step was essential to mitigate the influence of non-representative measurements arising from acquisition artefacts, tool malfunctions, or digitization errors, while preserving the underlying geological signal (Hastie et al. 2009). The analogue legacy data required an analog-to-digital conversion. The following quality control workflow was established for pre-processing to detect and eliminate inherently

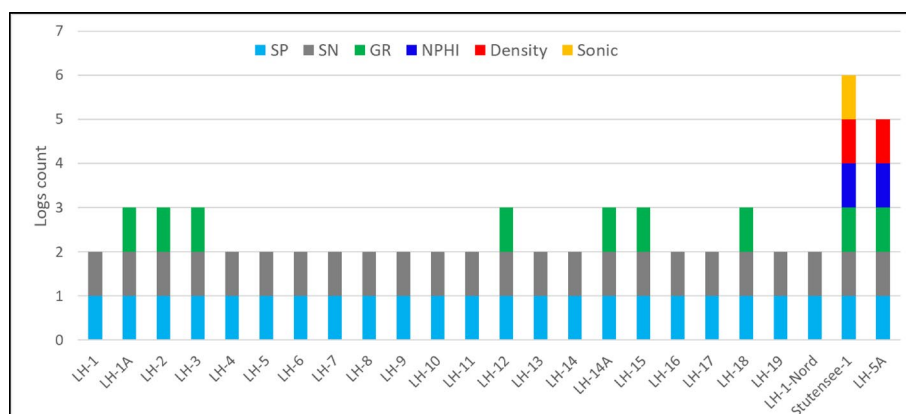


Fig. 3 Overview of well-by-well wireline logs in the Leopoldshafen oil field and the adjacent Stutensee-1 well. The distribution highlights the dominance of legacy measurements (SP and SN) and the limited availability of modern petrophysical logs such as NPHI. (SP: Self-Potential, SN: Short-Normal, GR: Gamma-Ray, NPHI: Neutron Porosity Index)

introduced artifacts including depth shifts, rasterization noise, and baseline drift, e.g., due to paper degradation.

- **Depth Alignment:** Curves were depth-matched to ensure stratigraphic correlation across the field.
- **Outlier Removal:** Extreme outliers were mitigated using an Interquartile Range (IQR)-based statistical filter, flagging any measurements exceeding the 1.5-times IQR threshold. This step was essential to eliminate anomalies such as cycle skips or paper degradation artifacts while preserving the underlying geological signal necessary for accurate reservoir characterization. The interquartile range (IQR) criterion was not applied as a standalone measure of geological invalidity. Instead, observations identified as statistical outliers were subjected to subsequent quality-control assessment based on log continuity, local geological trends, and evidence of acquisition or digitization artifacts. Only data points associated with isolated spikes, abrupt discontinuities, or demonstrable digitization artifacts were removed. This approach was adopted to distinguish non-geological measurement artifacts from genuine geological variability and, therefore, to minimize the risk of inadvertently removing physically meaningful reservoir heterogeneity (Honório et al. 2012). Outlier removal using the IQR-based filter ($1.5 \times$ IQR threshold) resulted in the exclusion of approximately 1.8% of the total data points across all wells. This proportion is consistent with typical quality control protocols for legacy log data and was deemed sufficient to remove obvious acquisition artifacts while preserving representative geological signal (Hastie et al. 2009).

Subsequently, a systematic quality-controlled analysis of the spontaneous potential (SP) and short normal resistivity (SN) well-log datasets was performed for all available wells. The dataset was filtered to include only log measurements acquired at depths greater than 650 m, thereby narrowing the analysis to the stratigraphic interval of primary geological interest where this interval contains the principal reservoir and caprock succession relevant to the planned DeepStor HT-RTES system. Shallower sections do not contain the primary target formations and therefore were excluded to focus the ML analysis on the stratigraphic interval for which reconstructed porosity is required. Subsequently, inconsistencies in the scaling of certain wells were identified and corrected to align their measurements with the appropriate physical units and expected log behavior. After rectifying these well-specific scale errors, normalization and scaling procedures were applied to the dataset. The objective of preprocessing was not to impose an absolute SP or SN magnitude across wells, but to harmonize relative log responses while preserving the within-well stratigraphic patterns and contrasts between lithological units.

To mitigate these systematic variances and harmonize the dataset for a ML application, a data-driven normalization strategy was employed (Eq. 1). This standardization ensures that the input features represent relative geological changes rather than acquisition artifacts. The methodology aligns with established petrophysical data handling protocols described by (Asquith and Krygowski 2004), and promotes the statistical robustness required for the subsequent ML workflow.

A feature transformation that rescales numerical data into a bounded interval while preserving the shape of the original distribution was applied using a min-max scaling technique. This step is critical for algorithms sensitive to the magnitude of input vectors,

such as Neural Networks and Gradient Boosting, as it ensures equal contribution from all features during the training process (Anifowose et al. 2017; Han et al. 2012; Rider 2006).

The general formulation of the min-max scaling is defined by

$$X' = X_{min_target} + \frac{(X - X_{min}) \cdot (X_{max_target} - X_{min_target})}{X_{max} - X_{min}} \quad (1)$$

where X is the original value of SP or SN, X_{min} and X_{max} are the statistical limits of the raw curve, X_{min_target} and X_{max_target} represent the desired range and X' is the scaled value.

To standardize the deflection baselines across the field and ensure consistent lithological interpretation, SP logs were normalized to a target range of $X_{min_target} = 10$ mV and $X_{max_target} = 80$ mV. The selected limits were determined primarily from the observed distribution of SP responses within the available dataset, as the majority of the analyzed logs exhibited deflections within or close to this interval. In addition to this range also aligned with industry conventions for hydrocarbon-bearing formations and to preserve signal-to-noise ratio for subsequent interpretation (Asquith and Krygowski 2004).

Short Normal resistivity logs were rescaled to harmonize the electrical response of matrix and fluid content across multiple wells and logging campaigns. The target range was set to $X_{min_target} = 1$ Ω m and $X_{max_target} = 2.5$ Ω m based on typical electrical properties of hydrocarbon-bearing sandstone formations and field-wide statistical analysis.

The validation of the data conditioning is visualized in Fig. 4, which presents a comparative analysis of raw SP and SN (red) and standardized SP and SN (SSP and SSN, blue) logs for four representative wells: LH-1, LH-1 A, LH-3, and LH-11. A notable rectification was achieved in well LH-1 A. The raw SP acquisition for this well exhibited an anomalous polarity reversal, where permeable sandy formations incorrectly corresponded to positive deflections. The standardization workflow successfully identified and corrected this inversion, restoring the physical relationship between the log response and the lithology.

Figure 5 presents SP as a function of depth, comparing raw acquisition data (SP and SN, Fig. 5a) with the standardized outputs (SSP and SSN, Fig. 5b). The raw logs (SP and SN) exhibited substantial heterogeneity, including erratic baseline shifts, systematic instrument drift, and physically implausible negative values inconsistent with clay-free sandstone formations. In contrast, the standardized logs (SSP and SSN) show a coherent response, with a stable baseline and physically plausible values. By restricting the analysis to depths greater than 650 m, the processing workflow isolates the target stratigraphic intervals of primary geological interest and improves the signal-to-noise ratio for subsequent predictive modeling.

Rock physics modeling and lithology differentiation

To exploit the complementary sensitivities of the available wireline measurements, the SSP and SSN responses were cross-plotted to construct a two-dimensional feature space for lithology discrimination. A data-driven lithological classification was then performed using clustering that was guided by the independent lithological interpretation derived during drilling. Lithology boundaries were delineated by applying empirically defined decision polygons in the SSP–SSN domain. Multiple polygon configurations were systematically tested and iteratively refined through comparison with geological reports

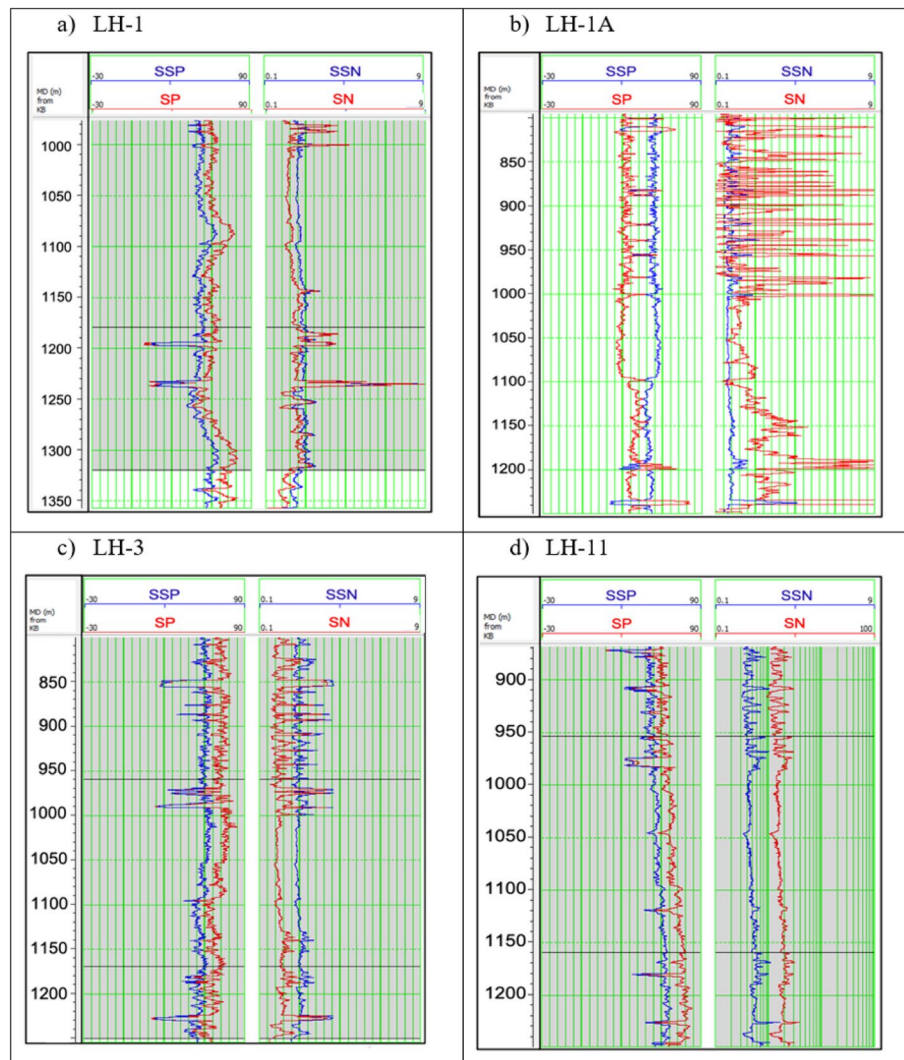


Fig. 4 A comparative analysis of the SN (mΩ) and SP (mV) logs before (SP, SN, red) and after (SSP, SSN, blue) standardization and normalization for **(a)** LH-1, **(b)** LH-1 A, **(c)** LH-3, and **(d)** LH-11, illustrating improved data consistency across wells. Note that these procedures do not alter the logs' response to lithology. (MD from KB = Measured Depth from Kelly Bushing)

and well-specific lithological interpretations. The final classification scheme was selected based on maximum agreement with the reference geological information across individual wells. On this basis, a binary lithology flag was assigned (0 = shale; 1 = sandstone), providing a consistent and geologically constrained framework for subsequent analysis. The binary lithology flag represents a simplified petrophysical classification rather than a complete geological facies model. Marl, shale, and other fine-grained intervals with similar electrical responses were grouped into the non-reservoir class, whereas reservoir-relevant sandstone intervals were grouped into the sandstone class.

ML for log prediction

This study applies supervised ML techniques to obtain NPHI logs in wells where only SSP and SSN data are available. To quantify the added value of geological constraints, two model configurations were evaluated. The baseline configuration (Model 1) utilized only the SSP and SSN logs as input features, representing a purely data-driven approach

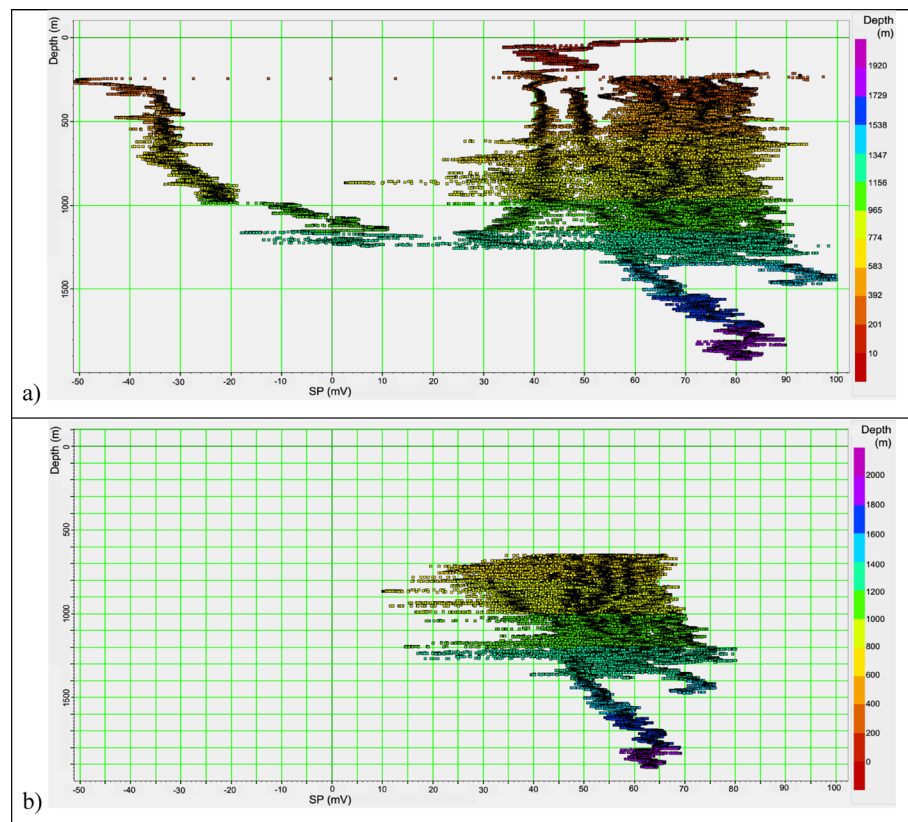


Fig. 5 SP (a) and SSP (b) versus depth of all wells of the Leopoldshafen field and the adjacent Stutensee-1 well in a depth below 650 m before (a) and after (b) standardization

based on electrical measurements. The enhanced configuration (Model 2) incorporated an additional categorical lithology indicator derived from a lithology-based clustering of SSP and SSN responses.

A suite of supervised ML algorithms, that represent widely used and complementary ML techniques for regression and classification tasks, is tested, including tree-based methods such as Random Forests (Breiman 2001), Gradient Boosting Machines (Friedman 2001), Extreme Gradient Boosting (Chen and Guestrin 2016), as well as the kernel-based Support Vector Regression (Smola and Schölkopf 2004), and the neural network-based Multi-Layer Perceptrons (Rumelhart et al. 1986).

To enhance predictive robustness and mitigate model-specific bias, a limited hyperparameter optimization strategy was applied across multiple supervised machine learning algorithms, including Random Forest (RF), Gradient Boosting (GB), Extreme Gradient Boosting (XGBoost), Support Vector Regression (SVR), and Multilayer Perceptron (MLP) neural networks. Hyperparameter tuning was conducted using a small-grid search framework, exploring a subset of model complexity, learning dynamics, and regularization parameters (Table 1). The relatively limited optimization strategy was adopted deliberately in consideration of the restricted size and distribution of the available dataset. In particular, the model development and validation were constrained by the availability of only one well (Stutensee-1) containing the required co-located input and target logs, while an independent well (LH-5 A) was reserved exclusively as a blind test dataset. Consequently, extensive hyperparameter optimization would provide limited additional

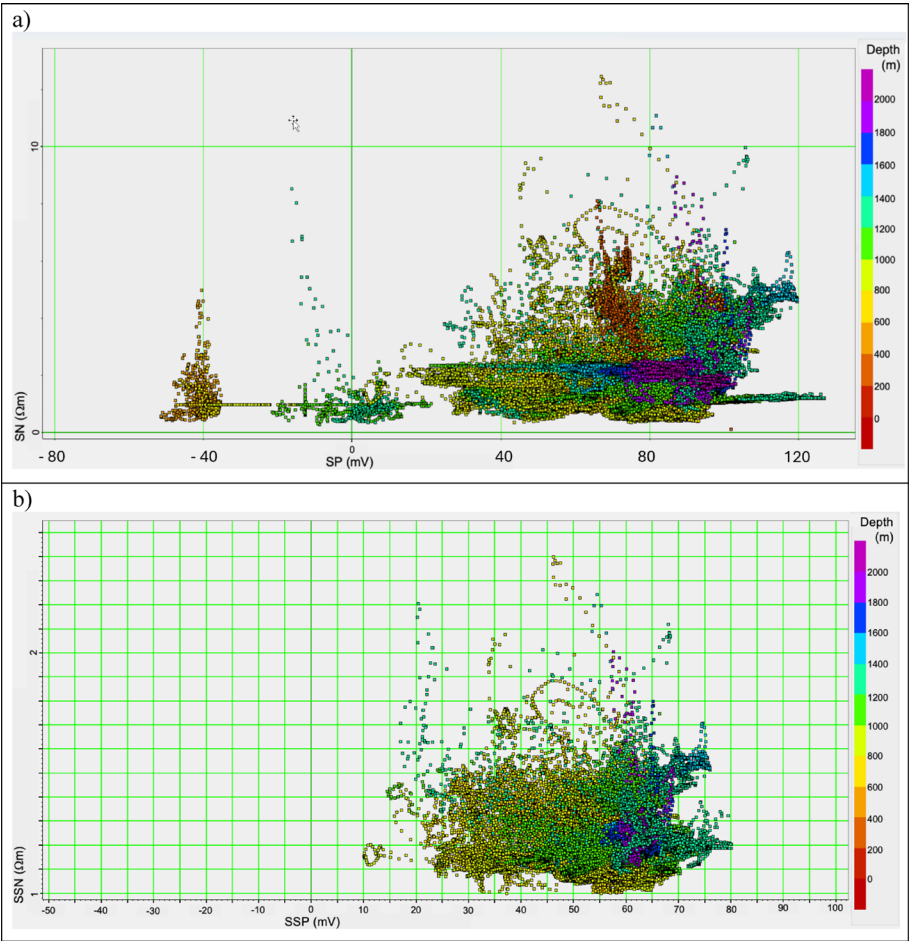


Fig. 6 SPversus SN **(a)** and SSP versus SSN **(b)** of all wells of the Leopoldshafen field and the adjacent Stutensee-1 well in a depth below 650m before **(a)** and after **(b)** standardization

Table 1 Hyperparameter grids used for model optimization

Model	Hyperparameter	Tested Values
Random Forest	n_estimators	100, 200, 300
	max_depth	None, 10, 20
	min_samples_split	2, 5, 10
Gradient Boosting	n_estimators	100, 200, 300
	learning_rate	0.01, 0.05, 0.1
	max_depth	3, 5, 7
XGBoost	n_estimators	100, 200, 300
	learning_rate	0.01, 0.05, 0.1
	max_depth	3, 5, 7
Support Vector Regressor (SVR)	C	0.1, 1, 10
	epsilon	0.01, 0.1, 0.2
	kernel	linear, rbf
MLP Neural Network	hidden_layer_sizes	(64,64), (128,128), (256,256)
	learning_rate_init	0.001, 0.01
	max_iter	2,000

statistical confidence and could increase the risk of overfitting to the characteristics of the single training well, particularly given the strong geological and data-related uncertainties inherent in legacy wireline datasets. The adopted strategy therefore represents a balance between model optimization and the preservation of independent predictive evaluation.

While more extensive optimization using automated frameworks such as AutoGluon (Erickson et al. 2020) or Bayesian hyperparameter optimization (Bergstra and Bengio 2012) could potentially identify marginally improved parameter configurations, such approaches were not considered appropriate as the primary optimization strategy for the present dataset because of the limited number of independent wells available for model development. Instead, the objective was to extract the maximum predictive value from the available legacy data while minimizing the risk of obtaining apparently high performance through excessive tuning of a model to a single well. The selected approach was sufficient to identify clear trends in algorithm performance and to evaluate the incremental contribution of lithological information while maintaining computational feasibility and methodological transparency. The limited search space should nevertheless be recognized as a methodological limitation, particularly for the MLP network, for which only three representative architectures were evaluated. More extensive optimization and validation using a larger number of wells would be desirable when additional data becomes available (Bergstra and Bengio 2012; Chen and Guestrin 2016; Hastie et al. 2009).

The model evaluation is based on the Root Mean Squared Error (RMSE) and the coefficient of determination (R^2), which are also standard metrics for continuous log prediction tasks (Bishop 2016; Hastie et al. 2009). Training and validation were conducted using the Stutensee-1 well, which provides the most complete logging suite in the study area. The dataset was partitioned into 80% training data and 20% blind test data, and k-fold cross-validation was applied to minimize overfitting and ensure model robustness (Wood 2020; Xie et al. 2018). Independent validation was performed using well LH-5 A.

The insufficient exposure of the ML model to gas-bearing sand responses during training was mitigated using a lithology-conditional correction in well LH-5 A. Specifically, sandstone intervals were isolated using the lithology flag, and a linear regression was fitted between the predicted and measured NPHI values within these intervals in Stutensee-1, yielding the correction in Eq. 2.

$$NPHI_{measured} = -0.3228 \cdot NPHI_{predicted} + 0.3098$$

Results

The data-driven lithological classification using clustering analysis guided by the independent lithological interpretation derived from drilling reports is shown in Figs. 7 and 8. Within this feature space, shale intervals are characterized by comparatively elevated SSP values associated with low SSN responses, whereas sandstone intervals preferentially cluster in domains defined by lower SSP values and higher SSN values.

Figure 9 illustrates the training (80%) and validation (20%) using SSP and SSN data of the Stutensee-1 well for ML algorithms with hyperparameter optimization which was performed for each algorithm to ensure optimal performance. A comparative evaluation of the RMSE and R^2 values for the applied ML methods and for both, the purely

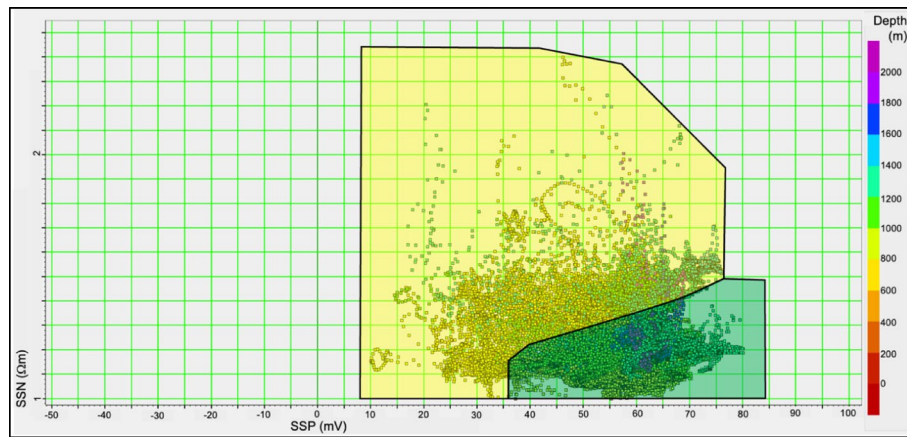


Fig. 7 SSP versus SSN of all wells of the Leopoldshafen field and the adjacent Stutensee-1 well in a depth below 650 m. Shale formations are identified within the green polygon, characterized by high SSP values and low SSN readings, while sand formations are delineated by the yellow polygon, corresponding to low SSP values and high SSN responses

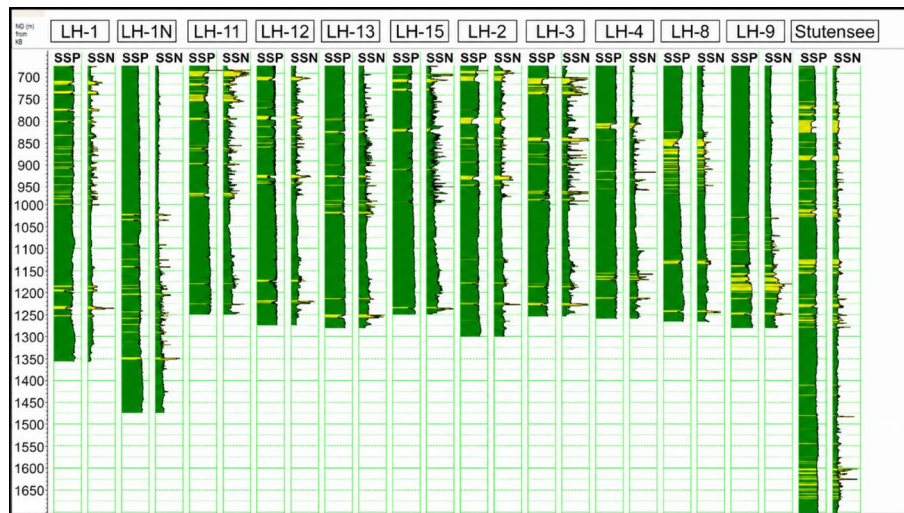


Fig. 8 Well correlation diagram illustrating lithological discrimination. For color coding see Fig. 7. The identified lithological units show a strong alignment with the sandstone layers identified during drilling

data-driven Model 1 (Baseline) and Model 2 (Enhanced) are shown in Figs. 10 and 11. These demonstrate the improvement of the predictive accuracy in Model 2 across all tested algorithms relative to the baseline configuration. While RF and MLP models exhibited comparable performance in the enhanced setup, the RF algorithm was selected for subsequent analyses due to its superior interpretability, inherent resistance to overfitting, and lower computational complexity compared to neural network architectures (Breiman 2001).

Independent validation of the enhanced ML framework (Model 2) was conducted using data from LH-5 A. As depicted in Fig. 12a, the predicted NPHI exhibits a modest correspondence with the measured log throughout the depth interval of 1,195–1,250 m, effectively reproducing both the absolute magnitude and the vertical variability of the signal. Quantitatively, the model achieves a coefficient of determination (R^2) of

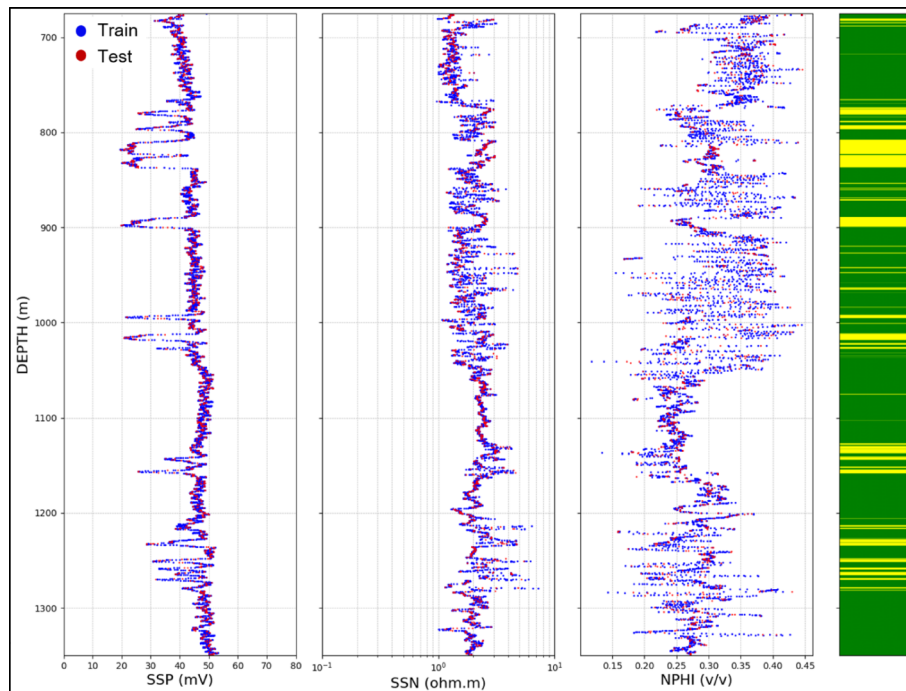


Fig. 9 Distribution of input and target log measurements with blue points representing the training dataset (80% of the samples) and red points the blind testing dataset (20%) for (a) SSP, (b) SSN, and (c) NPHI from the Stutensee-1 well used for ML model development and comparison to the litholog from Fig. 8 with green: shale and yellow: sand layers

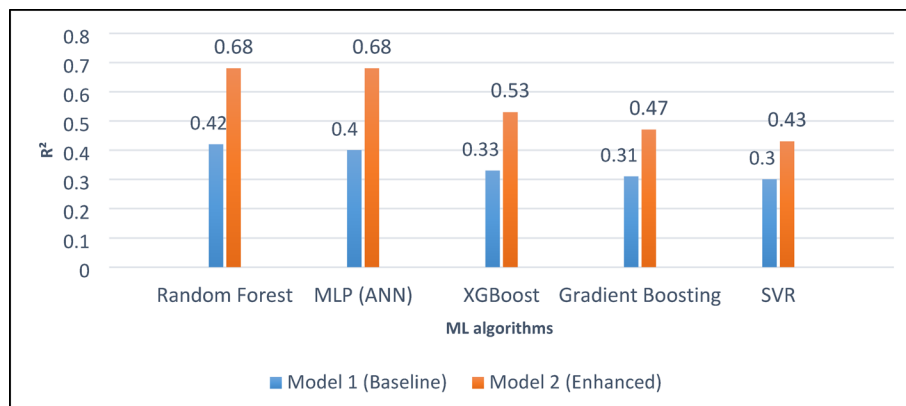


Fig. 10 Performance comparison of the applied supervised ML algorithms in predicting NPHI using R^2 through k-fold cross-validation for Model 1 (Baseline) that utilizes only SSP and SSN logs as inputs and Model 2 (Enhanced) that incorporates additional lithology flags derived from the rock-physics analysis

approximately 0.60, indicating a moderate yet moderate predictive capability in heterogeneous subsurface conditions.

A systematic deviation is, however, observed within the reservoir interval between 1,204 and 1,208 m, which is interpreted as a gas-saturated zone. In this interval, the model successfully captures the presence of a perturbation in NPHI but predicts it with an inverse trend relative to the measured response. This behavior can be attributed to the limited representativeness of the training dataset, which was derived exclusively from Stutensee-1. Specifically, the dataset is characterized by a pronounced class imbalance, with approximately 81% (5,813 samples) corresponding to shale lithologies and

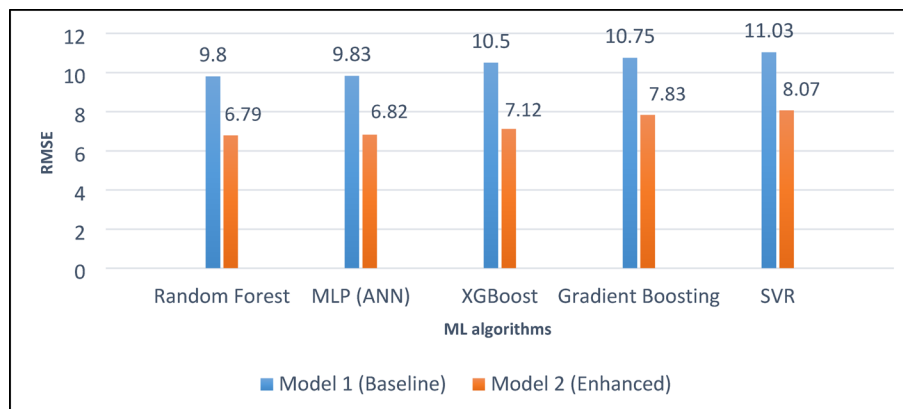


Fig. 11 Performance comparison of the applied supervised ML algorithms in predicting NPHI using RMSE through k-fold cross-validation for Model 1 (Baseline) that utilizes only SSP and SSN logs as inputs and Model 2 (Enhanced) that incorporates additional lithology flags derived from the rock-physics analysis

only 19% (1,112 samples) representing sandstone units (gas reservoir & non-reservoir). Consequently, the model exhibits a bias toward shale-dominated responses and lacks sufficient exposure to gas-bearing reservoir signatures, thereby limiting its generalization capability in such intervals.

To mitigate this limitation, a localized modelling strategy was implemented by constructing a targeted sub-model specifically trained on gas reservoir characteristics extracted from the training well. Application of this refined model to the reservoir interval in LH-5 A resulted in a marked improvement in predictive performance, with the coefficient of determination increasing to approximately 0.70. This outcome underscores the critical importance of lithology-balanced training datasets and reservoir-specific calibration in enhancing the reliability of machine learning predictions in subsurface characterization. The corrected NPHI log (Eq. 2) exhibits an improved agreement with the measured data across the entire depth interval, including the gas-bearing section (Fig. 12b).

The proposed workflow with all sandstone intervals correction has been applied to the remaining wells across the Leopoldshafen field (Fig. 13) and compared to porosity measurements from (David et al. 2026).

The predicted NPHI values in the calcareous sandstone layers and marls range from approximately 21% to >35% and from about 23% to 30%, respectively. Relative to typical values, these results indicate highly porous sandstones and comparatively tight marls. The greatest variability is observed in the Niederrödern marls of LH-2, where the well intersects the Leopoldshafen Fault. In general, the sandstone layers of the Niederrödern Formation exhibit the highest predicted NPHI values.

LH-13, the well closest to the planned DeepStor site, locally shows NPHI values exceeding 35% in the Karlsruhe Calcareous Sandstones. By contrast, effective porosity measured on plugs by helium pycnometry is systematically lower, ranging from 15.8% to 26.2% in the the plugs from LH-13 and ST-1. This difference likely reflects the fact that NPHI captures a more total pore-volume signal, whereas plug-derived helium porosity is intrinsically lower because plugs are taken from more competent parts of the well. The comparison indicates that the predicted NPHI values are within a realistic range for the Karlsruhe Calcareous Sandstones at LH-13.

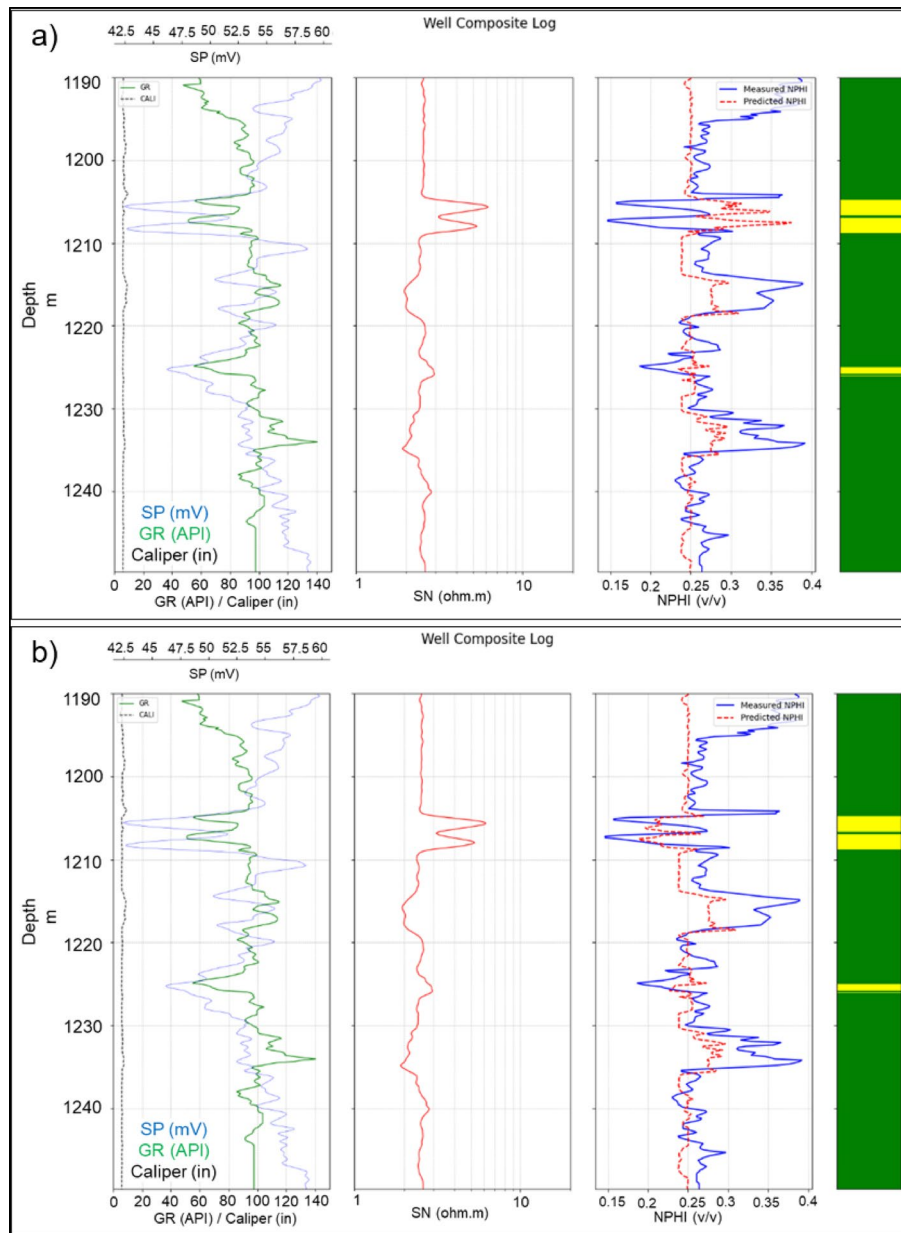


Fig. 12 Comparison between measured (blue) and predicted (red) NPHI logs in the LH-5 A well **(a)** without gas-correction and **(b)** with gas correction

Discussion

During the exploration of the Leopoldshafen oil field, SP and SN logs were ideally suited to distinguish hydrocarbon-bearing calcareous sandstones from the enclosing marls. From the perspective of heat storage, however, the near-exclusive availability of SP and SN logs, combined with the lack of NPHI logs, represents a major limitation for estimating storage capacity and delineating target zones for well placement. Our results show that ML algorithms can reliably predict porosity in legacy wells when trained on NPHI data from more recent wells.

The observed discrepancy between measured and predicted NPHI in gas-bearing intervals (Fig. 12) is not necessarily a drawback in this context. In gas-filled pores, neutron porosity measurements are known to be affected by the low hydrogen index of

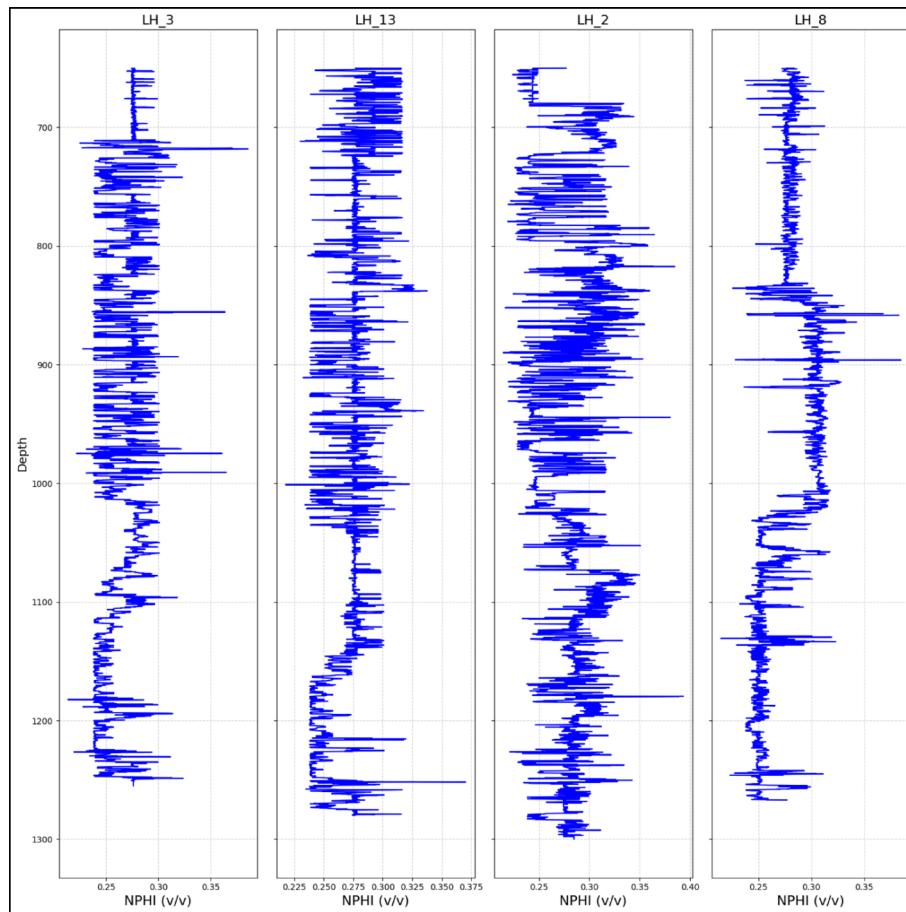


Fig. 13 Predicted NPHI logs from representative wells from NW to SE, namely LH-3, LH-13, LH-2, and LH-8, in the Leopoldshafen field

the gas phase, which causes NPHI to underestimate the true pore volume. As a result, intervals with gas saturation may appear to have lower porosity than they actually possess. For applications such as aquifer thermal energy storage, however, total porosity is the critical parameter because it governs the available pore volume for fluid circulation, heat exchange, and storage capacity. Consequently, the fact that NPHI may underpredict porosity in gas-bearing zones can be beneficial rather than problematic, because the ML-derived estimates may better approximate the effective total porosity relevant to reservoir utilization (Tas et al. 2025). This makes the workflow particularly valuable for identifying intervals with favorable storage potential in legacy wells where direct porosity measurements are sparse or unavailable.

However, uncertainty in the predicted NPHI remains considerable, as reflected by the relatively low R^2 and RMSE values (Figs. 10 and 11). Several factors may contribute to this uncertainty: (i) the indirect relationship between electrical logs and porosity, (ii) the analog origin and variable quality of the legacy data, (iii) the limited number of training wells, and (iv) the absence of density (RHOB) or sonic logs, which would provide additional constraints on volumetric and elastic properties. Comparable studies, for example on generating sonic logs from drilling parameter at the Utah FORGE site indicate that incorporating a larger number of individual datasets can reduce prediction uncertainty (Vivas and Salehi 2021), suggesting that the present dataset is close to the practical limit

of this approach. Accordingly, these uncertainties must be explicitly considered when using the predicted logs in quantitative reservoir simulations at the DeepStor site. In the context of HT-RTES, the reconstructed porosity should therefore be treated as an input for probabilistic modeling rather than as a deterministic parameter.

The superior performance of Model 2 relative to Model 1 across all tested algorithms (Figs. 10 and 11) supports the hypothesis that a larger set of input data improves prediction accuracy and that explicitly incorporating geological knowledge into the feature space constrains the solution to physically plausible regimes. Although SSP and SSN implicitly contain lithological information, the derived lithology indicator acts as a categorical prior that guides the ML algorithms to learn distinct relationships for sandstones and shales. This result is consistent with recent studies advocating physics-informed or physics-guided machine learning to improve generalization and interpretability in subsurface applications (Chakraborty et al. 2024; Suleymanov et al. 2023).

In addition to data sparsity, class imbalance is a major limitation in supervised ML applications to well-log prediction, lithofacies classification, and reservoir-property estimation, closely mirroring the challenges encountered in this study. (Hall 2016), for example, showed that lithology prediction from wireline logs is strongly biased when shale-dominated intervals outnumber sand-rich facies in the training dataset. In such cases, standard classifiers tend to overpredict the majority shale class, which can lead to systematic underestimation of porosity and permeability in reservoir-quality sandstones. To address this issue, we applied facies-specific weighting and stratified resampling, which improved prediction performance in underrepresented sandstone intervals without compromising the shale predictions.

By transforming legacy SP and SN logs into spatially continuous porosity estimates, this study reduces subsurface uncertainty and strengthens the feasibility assessment of repurposing mature hydrocarbon fields for thermal energy storage. More generally, the workflow illustrates how historical datasets, although often considered obsolete, can be re-used with modern data-science methods to support the energy transition. The workflow may be applicable to other mature basins with comparable legacy data availability, although its performance would require independent validation for each geological and logging context.

In the Leopoldshafen field, our results (Fig. 13) indicate both high- and low-porosity intervals within the calcareous sandstone layers of the two formations; however, no systematic vertical or lateral trend is evident. Given the circulation of highly saline thermal waters through these sandstones, with salinities of approximately 120 g/L (Banks et al. 2021), it is plausible that some low-porosity intervals have been affected by natural precipitation processes. Under such conditions, mineral precipitation from saline formation waters can promote pore-throat clogging and cement accumulation, thereby reducing effective porosity and locally increasing heterogeneity within otherwise reservoir-quality sandstones. This interpretation is consistent with the observed patchy distribution of predicted porosity values and may partly explain the contrast between high- and low-porosity sections.

At the same time, the pronounced variability in predicted NPHI values highlights the uncertainty associated with extrapolating to the planned DeepStor well. Without direct calibration data from the target location, it remains difficult to distinguish genuine spatial variations in pore structure from effects related to fluid composition, mineral

precipitation, or log-prediction uncertainty. Consequently, the DeepStor porosity estimates should be interpreted cautiously and ideally be refined using site-specific core, log, or fluid-chemistry data as they become available.

Conclusion

With the aim of repurposing the field for HT-RTES, this study presents an integrated rock-physics and ML workflow designed to address the absence of NPHI logs in the mature Leopoldshafen brownfield reservoir in the northern Upper Rhine Graben. The methodology exploits the only measurements consistently available across most wells, SP and SN logs, to predict porosity in wells lacking modern logging suites. This work is motivated by the need for reliable reservoir characterization to support the deployment of an HT-RTES at the DeepStor site.

A rigorous pre-processing strategy was implemented to maximize the informational value of analog well logs originally archived in paper format. This workflow included digitization, depth alignment, outlier removal, scale correction, and normalization, ensuring statistical consistency across multi-decadal datasets. A rock physics-based lithology discrimination approach was then applied using standardized SP–SN cross-plots and lithologs acquired during drilling, enabling the differentiation of sandstone and shale intervals and providing a geologically meaningful categorical input for ML.

The comparative evaluation of multiple supervised ML algorithms demonstrates that incorporating lithological information significantly improves NPHI prediction performance relative to purely data-driven models. Among the evaluated models, Random Forest achieved the best overall predictive performance in the present dataset while also offering established advantages in interpretability and resistance to overfitting. Blind validation against an independent well containing measured NPHI confirms that the proposed workflow “reproduces the overall trend and a substantial portion of the observed variability, although notable local discrepancies remain. The overall magnitude and vertical trends of porosity. Localized discrepancies are primarily associated with gas-bearing sandstone intervals, reflecting both the physical limitations of neutron logging tools and the under-representation of such conditions in the training dataset.

A lithology-conditional post-prediction calibration applied exclusively to sandstone intervals substantially reduced these discrepancies and yielded more geologically consistent porosity profiles. While this correction improves local accuracy, the reconstructed porosity logs remain associated with uncertainty and should be interpreted as first-order estimates suitable for probabilistic reservoir modeling rather than deterministic evaluation.

Overall, this study demonstrates that legacy SP and SN logs, often considered insufficient for quantitative petrophysical analysis, can be transformed into actionable subsurface information when combined with physics-guided feature engineering and ML. The proposed workflow significantly expands the usable information content of historical well data and provides a scalable and transferable solution for reservoir characterization in support of geothermal energy and HT-RTES development.

In the Leopoldshafen field, the calcareous sandstones exhibit strong local porosity variability without a clear systematic trend. This patchy distribution may reflect natural mineral precipitation from highly saline formation waters (~120 g/L), which can reduce effective porosity through pore clogging and cementation. However, the pronounced

uncertainty in the predicted NPHI values means that the DeepStor porosity estimates should be interpreted cautiously until they can be constrained by site-specific calibration data.

In conclusion, this study demonstrates that even highly uncertain and incomplete legacy datasets can be transformed into valuable subsurface information through careful preprocessing, physics-guided feature engineering, and ML. While uncertainties remain inherent and unavoidable, the proposed workflow significantly expands the usable information content of historical well data and provides a pragmatic and scalable solution for reservoir studies supporting geothermal energy development.

Acknowledgements

This study was conducted as part of the DeepStor project. The authors acknowledge the support of the Karlsruhe Institute of Technology (KIT) and the project partners. This study is part of “Geoenergy” in the research program “MTET—Materials and Technologies for the Energy Transition” of the Helmholtz Association. Eva Schill acknowledges the support by the U.S. Department of Energy, Hydrocarbon and Geothermal Technologies Office (HGEO), Office of Geothermal, under Award Number DE-AC02–05CH11231 with Lawrence Berkeley National Laboratory and within the funding framework of the VESTA project. We express our appreciation to Geosoft for providing access to the HampsonRussell software. Open Access funding was enabled and organized by Projekt DEAL.

Author contributions

A.T.T. conceived the study, developed the data science, preprocessing for the legacy well logs and machine learning workflows, conducted physics data analyses, and prepared the primary manuscript text, tables, and figures. - X.G. performed the analog-to-digital well data extraction, contributed to the geological setting section, and integrated the core analysis benchmarks. - E.S. provided research organization, enhanced the geological framework, refined the manuscript text and figures, and supplied critical operational data regarding the DeepStor infrastructure project. - F.B. served as the project manager, provided administrative oversight, and completed the final revision of the manuscript. - T.K. acted as the principal investigator and supervisor, steering the overall scientific trajectory of the research. All authors reviewed and approved the final version of the manuscript.

Data availability

The data that support the findings of this study are available from the state geological archives and project partners, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are, however, available from the authors upon reasonable request and with permission of the relevant third parties.

Declarations

Competing interests

This study is part of “Geoenergy” in the research program “MTET—Materials and Technologies for the Energy Transition” of the Helmholtz Association.

Received: 1 June 2026 / Accepted: 5 October 2026

Published online: 13 October 2026

References

- Akindipe D, McTigue J, Dobson P, Atkinson T, Witter E, Kumar R, Sonnenthal E, Umbro M, Lederhos J, Adams D, Zhu G. (2024). *Techno-Economic Analysis and Market Potential of Geological Thermal Energy Storage (GeoTES) Charged With Solar Thermal and Heat Pumps* (NREL/TP–5700–91225, 2474842, MainId:93003; p. NREL/TP–5700–91225, 2474842, MainId:93003). <https://doi.org/10.2172/2474842>
- Anifowose FA, Labadin J, Abdullaheem A. Ensemble machine learning: an untapped modeling paradigm for petroleum reservoir characterization. *J Pet Sci Eng*. 2017;151:480–7. <https://doi.org/10.1016/j.petrol.2017.01.024>.
- Archie GE. The Electrical Resistivity Log as an Aid in Determining Some Reservoir Characteristics. *Trans AIME*. 1942;146(01):54–62. <https://doi.org/10.2118/942054-G>.
- Asquith GB, Krygowski D. Basic well log analysis. 2. ed. AAPG; 2004.
- Banks J, Poulette S, Grimmer J, Bauer F, Schill E. Geochemical changes associated with high-temperature heat storage at intermediate depth: thermodynamic equilibrium models for the DeepStor site in the Upper Rhine Graben, Germany. *Energies*. 2021;14(19):6089. <https://doi.org/10.3390/en14196089>.
- Bauer F, Grimmer JC, Hought L, Hertweck T, Schill E. Development of intra-rift basins during stress field change in the central Upper Rhine Graben (SW Germany). *Tectonics*. 2025;44(4):e2024TC008721. <https://doi.org/10.1029/2024TC008721>.
- Bergstra J, Bengio Y. Random search for hyper-parameter optimization. *J Mach Learn Res* (Vol. 2012;13:281–305).
- Bishop CM. *Pattern Recognition and Machine Learning* (Softcover reprint of the original 1st edition 2006 (corrected at 8th printing 2009)). New York: Springer; 2016.
- Breiman L. Random Forests. *Mach Learn*. 2001;45(1):5–32. <https://doi.org/10.1023/A:1010933404324>.
- Chakraborty S, Datta Gupta S, Devi V, Yalamanchi P. Using rock physics analysis driven feature engineering in ML-based shear slowness prediction using logs of wells from different geological setup. *Acta Geophys*. 2024;72(5):3237–54. <https://doi.org/10.1007/s11600-023-01266-3>.

- Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. Proc 22nd ACM SIGKDD Int Conf Knowl Discovery Data Min. 2016;785–94. <https://doi.org/10.1145/2939672.2939785>.
- David C, Guillou M, Garipi X, Schill E, Gaucher E, Barnes C. Geomechanical properties of the Meletta sandstone—the high-temperature heat storage reservoir rock of DeepStor. Geomech Energy Environ. 2026;46:100821. <https://doi.org/10.1016/j.gete.2026.100821>.
- Erickson N, Mueller J, Shirkov A, Zhang H, Larroy P, Li M, Smola A. AutoGluon-Tabular: Robust and Accurate AutoML for Structured Data (Version 1). arXiv. 2020. <https://doi.org/10.48550/ARXIV.2003.06505>.
- Fleuchaus P, Schuppeler S, Bloemendal M, Guglielmetti L, Opel O, Blum P. Risk analysis of high-temperature aquifer thermal energy storage (HT-ATES). Renew Sustain Energy Rev. 2020;133:110153. <https://doi.org/10.1016/j.rser.2020.110153>.
- Friedman JH. Greedy function approximation: a gradient boosting machine. Ann Stat. 2001. <https://doi.org/10.1214/aos/1013203451>.
- Geyer M, Nittsch E, Simon T. (2023). *Geologie von Baden Württemberg*.
- Hall B. Facies classification using machine learning. Lead Edge. 2016;35(10):906–9. <https://doi.org/10.1190/tle35100906.1>.
- Han J, Kamber M, Pei J. Data Mining. Elsevier; 2012. <https://doi.org/10.1016/C2009-0-61819-5>.
- Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning. New York: Springer; 2009. <https://doi.org/10.1007/978-0-387-84858-7>.
- Heldt S, Beyer C, Bauer S. Uncertainty assessment of thermal recovery and subsurface temperature changes induced by high-temperature aquifer thermal energy storage (HT-ATES): a case study. Geothermics. 2024;122:103086. <https://doi.org/10.1016/j.geothermics.2024.103086>.
- Heldt S, Beyer C, Bauer S. Impact of subsurface heterogeneity and thermally induced buoyancy-driven flow on high-temperature aquifer thermal energy storage. Geothermics. 2026;138:103641. <https://doi.org/10.1016/j.geothermics.2026.103641>.
- Hinsken S, Ustaszewski K, Wetzel A. Graben width controlling syn-rift sedimentation: The Palaeogene southern Upper Rhine Graben as an example. Int J Earth Sci. 2007;96(6):979–1002. <https://doi.org/10.1007/s00531-006-0162-y>.
- Hinsken S, Schmalholz SM, Ziegler PA, Wetzel A. Thermo-Tectono-Stratigraphic Forward Modelling of the Upper Rhine Graben in reference to geometric balancing: Brittle crustal extension on a highly viscous mantle. Tectonophysics. 2011;509(1–2):1–13. <https://doi.org/10.1016/j.tecto.2010.12.006>.
- Honório BCZ, Drummond RD, Vidal AC, Sanchetta AC, Leite EP. Well log denoising and geological enhancement based on discrete wavelet transform and hybrid thresholding. Energy Explor Exploitation. 2012;30(3):417–33. <https://doi.org/10.1260/0144-5987.30.3.417>.
- Karniadakis GE, Kevrekidis IG, Lu L, Perdikaris P, Wang S, Yang L. Physics-informed machine learning. Nat Reviews Phys. 2021;3(6):422–40. <https://doi.org/10.1038/s42254-021-00314-5>.
- Lee KS. A review on concepts, applications, and models of aquifer thermal energy storage systems. Energies. 2010;3(6):1320–34. <https://doi.org/10.3390/en3061320>.
- LGRB. (2016). *Symbolschlüssel Geologie Baden Württemberg*.
- Rider MH. The geological interpretation of well logs. 2. ed., rev. reprinted. Rider-French Consulting; 2006.
- Rider MH, Kennedy M. The geological interpretation of well logs. 3 rd ed. Rider-French consulting; 2011.
- Rosenbrand E, Haugwitz C, Jacobsen PSM, Kjoller C, Fabricius IL. The effect of hot water injection on sandstone permeability. Geothermics. 2014;50:155–66. <https://doi.org/10.1016/j.geothermics.2013.09.006>.
- Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature. 1986;323(6088):533–6. <https://doi.org/10.1038/323533a0>.
- Schill E, KNAUTH R, GARIP X, BAUER F. (2024). *DeepStor—Heat Cycling in the Deep and Medium-Deep Subsurface*. Stanford Geothermal Workshop. <https://pangea.stanford.edu/ERE/db/GeoConf/papers/SGW/2024/Schill.pdf>
- Schumacher ME. Upper Rhine Graben: role of preexisting structures during rift evolution. Tectonics. 2002. <https://doi.org/10.1029/2001tc900022>.
- Smola AJ, Schölkopf B. A tutorial on support vector regression. Stat Comput. 2004;14(3):199–222. <https://doi.org/10.1023/B:STC.0.0000035301.49549.88>.
- Spichak V, Geiermann J, Zakharova O, Calcagno P, Genter A, Schill E. Estimating deep temperatures in the Soultz-sous-Forêts geothermal area (France) from magnetotelluric data. Near Surf Geophys. 2015;13(4):397–408. <https://doi.org/10.3997/1873-0604.2015014>.
- Stricker K, Grimmer JC, Egert R, Bremer J, Korzani MG, Schill E, et al. The potential of depleted oil reservoirs for high-temperature storage systems. Energies. 2020;13(24):6510. <https://doi.org/10.3390/en13246510>.
- Stricker K, Egert R, Schill E, Kohl T. Risk of surface movements and reservoir deformation for high-temperature aquifer thermal energy storage (HT-ATES). Geotherm Energy. 2024;12(1):4. <https://doi.org/10.1186/s40517-024-00283-9>.
- Stricker K, Grimmer J, Meixner J, Dashti A, Egert R, Gholamikorzani M, Schaetzler K, Schill E, Kohl T. (2020, March 23). *Utilization of abandoned hydrocarbon reservoirs for deep geothermal heat storage*. <https://doi.org/10.5194/egusphere-egu2020-20723>
- Suleymanov V, El-Husseiny A, Glatz G, Dvorkin J. Rock physics and machine learning comparison: elastic properties prediction and scale dependency. Front Earth Sci. 2023;11:1095252. <https://doi.org/10.3389/feart.2023.1095252>.
- Tas L, Hartog N, Bloemendal M, Simpson D, Robert T, Thibaut R, Zhang L, Hermans T. Efficiency and heat transport processes of low-temperature aquifer thermal energy storage systems: New insights from global sensitivity analyses. Geotherm Energy. 2025;13(1):2. <https://doi.org/10.1186/s40517-024-00326-1>.
- Tolba AT, Bauer F, Grimmer JC, Dashti A, Kohl T. Spectral decomposition reveals hidden fluvial channels: implications for geothermal potential in the Karlsruhe area, Upper Rhine Graben, Germany. J Appl Geophys. 2026;250:106266. <https://doi.org/10.1016/j.jappgeo.2026.106266>.
- Vivas C, Salehi S. (2021). *Machine learning approach to generate synthetic sonic logs in geothermal wells*. (Vol. 45), 3–6.
- Willard J, Jia X, Xu S, Steinbach M, Kumar V. Integrating scientific knowledge with machine learning for engineering and environmental systems. ACM Comput Surv. 2023;55(4):1–37. <https://doi.org/10.1145/3514228>.
- Witter E, Akindipe D, Zhu G, McTigue J. Geological thermal energy storage offers terawatt-hours of energy storage capacity. Npj Therm Sci Eng. 2026;1(1):8. <https://doi.org/10.1038/s44435-026-00008-3>.
- Wood DA. Predicting porosity, permeability and water saturation applying an optimized nearest-neighbour, machine-learning and data-mining network of well-log data. J Pet Sci Eng. 2020;184:106587. <https://doi.org/10.1016/j.petrol.2019.106587>.

- Xie Y, Zhu C, Zhou W, Li Z, Liu X, Tu M. Evaluation of machine learning methods for formation lithology identification: a comparison of tuning processes and model performances. *J Pet Sci Eng*. 2018;160:182–93. <https://doi.org/10.1016/j.petrol.2017.10.028>.
- Zhang J, Liu G, Wei Z, Li S, Zayier Y, Cheng Y. Machine Learning-Based Prediction of Well Logs Guided by Rock Physics and Its Interpretation. *Sensors*. 2025;25(3):836. <https://doi.org/10.3390/s25030836>.
- Zhang C, Qin Z, Jin Y, Huang W, He H, Ding Y. High temperature aquifer thermal energy storage system: Current status of research and future prospects. *Energy Ecol Environ*. 2026. <https://doi.org/10.1007/s40974-026-00417-8>.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.