

## The application of the k-means clustering method in the prediction of the primary and secondary mineral composition of sandstones: a case study from Lower and Middle Jurassic rocks in the Polish Basin

Sara WRÓBLEWSKA-JANC<sup>1,2\*</sup>, Kinga BOBEK<sup>3</sup>, Anna WYSOCKA<sup>4</sup> and Małgorzata KOZŁOWSKA<sup>1</sup>

- <sup>1</sup> University of Warsaw, Faculty of Geology, Żwirki i Wigury 93, 02-089 Warszawa, Poland; ORCID: 0000-0002-9919-1518 [S.W.-J.], 0000-0002-1904-6098 [M.K.]
- <sup>2</sup> ORLEN S.A. PGNiG Exploration and Production Branch in Warsaw, Marcina Kasprzaka 25A, 01-224, Warszawa, Poland
- <sup>3</sup> Polish Geological Institute – National Research Institute, Rakowiecka 4, 00-975 Warszawa, Poland; ORCID: 0000-0002-6137-9945
- <sup>4</sup> Institute of Geological Sciences, Polish Academy of Sciences, Twarda 51/55, 00-818 Warszawa, Poland; ORCID: 0000-0003-1015-5033



Wróblewska-Janc, S., Bobek, K., Wysocka, A., Kozłowska, M., 2024. The application of the k-means clustering method in the prediction of the primary and secondary mineral composition of sandstones: a case study from Lower and Middle Jurassic rocks in the Polish Basin. *Geological Quarterly*, 68, 30; <https://doi.org/10.7306/gq.1758>

Associate editor: Piotr Krzywiec

Comprehensive knowledge of both the primary and secondary mineral composition of rocks, as well as the diagenetic processes occurring within them, stands as a fundamental requirement for accurately estimating reservoir rock parameters across various geological sectors, such as petroleum geology, geothermal energy, and carbon capture and storage. In the era of widespread automatization, machine learning methods are increasingly being used for interpretation, which, with the support of appropriate datasets and the experience of interpreters, make it possible to draw a variety of conclusions about geological processes occurring within selected geological formations, as well as entire sedimentary basins. We describe application of the k-means clustering method for the rapid prediction of primary and diagenetic mineral composition using the example of Lower and Middle Jurassic sandstones in the Polish Basin – one of the main aquifers and a potential reservoir formation. The model was based on the correlation between neutron porosity (NPHI), bulk density (RHOB), interval transit time (DT), deep resistivity (LLD), total natural gamma-ray (GR) and spectral gamma-ray values (K, Th, U), correlated with the results of petrographic, petrophysical and qualitative geochemical analysis. This correlation was the basis for distinguishing 5 different sandy petrofacies with variable primary and diagenetic characteristics typical of Jurassic sandstones in the Polish Basin.

Key words: cluster analysis, Jurassic sandstones, well-logging, petrofacies, machine learning.

### INTRODUCTION

In recent years unsupervised machine-learning methods have gained significant popularity across various sectors of the exploration industry (Hussein et al., 2021; Bhattacharya, 2022; Mohammadian et al., 2022; Wang et al., 2022; Ali et al., 2023). While geological data interpretation should still be controlled by the experienced professional mind, unsupervised machine learning can be applied to the well logs to obtain clusters, that can be correlated with the lithology of the borehole (Chang et al., 2021; Ippolito et al., 2021; Singh and Ojha, 2022). However,

more detailed characterization of particular lithologies can be performed to determine various lithotypes that exhibit diverse physical properties, usually related to the primary and/or secondary mineral composition of the rock. This type of analysis is called petrofacies analysis and is crucial especially to proper reservoir characterization, providing a better understanding of pore size generation and distribution. The classification of reservoir petrofacies can be based on various criteria, categorizing reservoir rocks into distinct units that were deposited under similar geological conditions and have undergone analogous diagenetic alterations (Ingersoll, 1990; De Ros and Goldberg, 2007; Cui et al., 2017; Lis-Śledzionka, 2019; Wróblewska and Kozłowska, 2019; Wróblewska, 2022).

The application of petrofacies analysis is particularly crucial in complex geophysical datasets when core samples are limited or unavailable. This type of research on Lower and Middle Jurassic sandstones in the Polish Basin was previously con-

\* Corresponding author: e-mail: sara.wroblewska@student.uw.edu.pl

ducted by [Wróblewska \(2022\)](#) using archival, recalculated wireline log datasets along with geochemical, petrophysical, and petrographic results from core samples. The Lower and Middle Jurassic sandy rocks of the Mid-Polish Trough in the Polish Basin provide an excellent case for testing various methods of determining sandy petrofacies. The substantial petrographic variation in the primary and diagenetic mineral composition of these sandy rocks poses challenges for geophysical interpretation and the standard calculation of their reservoir properties. The Jurassic clastic rocks belong to the arenites group, including variable amounts of K-feldspars and mica grains ([Leonowicz, 2005](#); [Kozłowska and Maliszewska, 2015](#)). Even a small admixture of these radiogenic grains can result in anomalous gamma-ray measurements, potentially leading to a misinterpretation of shale volume and therefore effective porosity calculations. Conversely, minor diagenetic changes such as non-radioactive calcareous and quartz cements or the kaolinization of unstable grains in Jurassic sandstones can alter pore space connectivity and consequently reduce permeability. [Wróblewska \(2022\)](#) proposed a novel method known as the “shale difference” approach to facilitate the recalculation of archival well data measured in non-standard units. This method relies on quantifying the difference between the shale volume parameter calculated from spontaneous potential and that obtained from the gamma-ray log, forming the foundation for distinguishing four distinct sandy petrofacies. The primary objective of this method is to interpret sandy rock successions that contain abundant, less radioactive clay minerals, particularly kaolinite. Kaolinite is a product of the alteration of K-feldspars and lithic grains and is less clearly detectable by the total gamma-ray log than other clay minerals. Its presence is of particular significance due to the creation of secondary microporosity between kaolinite grains. Moreover, the application of deep resistivity and sonic measurement was the basis for distinguishing carbonate-cemented intervals in the Z-GN4 borehole.

The primary aim of this study is to demonstrate how the application of unsupervised machine learning can contribute to the more objective and rapid determination of petrofacies in uncored intervals within generally well-recognized reservoir formations on the example of clastic Jurassic rocks from the B-2 borehole in the Polish Lowlands, which further can be correlated with particular primary and diagenetic features of the lithotypes distinguished. This simple method may also be applied in further reservoir formation evaluation for oil and gas exploration and CCUS (carbon capture, utilisation and storage technologies)/geothermal aquifer assessment.

## GEOLOGICAL SETTING

The B-2 borehole is situated within the Warsaw segment of the marginal western part of the Kościerzyn-Puławy synclinorium ([Fig. 1](#)). This geological structure, along with the adjacent Mid-Polish Swell, originated as a result of Late Cretaceous to Paleogene regional uplift and the inversion of a tectonic unit known as the Mid-Polish Trough ([Dadlez, 1997, 1998](#); [Gutowski et al., 2003](#); [Krzywiec, 2006](#); [Żelaźniewicz et al., 2011](#)). The Jurassic clastic deposits found in the Polish Basin, which are the focus of this research, are part of the Permian-Mesozoic Mid-European Epicontinental Basin ([Dadlez, 1998](#)). The development of Mesozoic deposits in this region was influenced by regional thermal subsidence and salt tectonics, associated with the Upper Permian (Zechstein) evaporites, which began in the Early Triassic (e.g., [Dadlez et al., 1995, 1998](#); [Marek and Pajchłowa, 1997](#); [Dadlez, 1998](#); [Karnkowski, 1999](#); [Krzywiec et al., 2003](#); [Stephenson et al., 2003](#); [Krzywiec,](#)

[2006](#)). The main source of the terrigenous material in the Jurassic basin was probably surrounding regions of the Pre-Carpathian land to the south, Belorussian High and Ukrainian Shield to the east, the Bohemian Massif to the southwest, and the Fennoscandian Shield to the north ([Dadlez and Marek, 1969](#); [Ziegler, 1990](#); [Feldman-Olszewska, 1997a](#); [Marek and Pajchłowa, 1997](#); [Pieńkowski, 2004](#); [Fig. 2](#)).

The Lower Jurassic strata consist predominantly of sandstones and mudstones with thin intercalations of coal, calcareous sandstone, and siderite of shallow marine and/or continental origin deposited in fluvial, deltaic, estuarine, lacustrine, and/or lagoonal environments ([Feldman-Olszewska, 1997b, 1998](#); [Pieńkowski, 2004](#); [Fig. 2](#)). The Middle Jurassic is characterized by marine and fluvial sandstones, as well as fine-grained anoxic shelf deposits. These deposits formed during successive ([Dayczak-Calikowska, 1964](#); [Dayczak-Calikowska and Moryc, 1988](#); [Feldman-Olszewska, 1997b, 2005](#)) transgressions and brief regressive periods in the Early/Late Bajocian and Early Callovian ([Dayczak-Calikowska and Moryc, 1988](#); [Dayczak-Calikowska, 1997](#); [Kopik, 1998](#)). Shallow marine carbonate ramps and starved shelf deposits are also common in the upper part of the Middle Jurassic profile. The gradual dominance of carbonate over clastic deposits from the Early Oxfordian was caused by the basin deepening and widening and the slow subsidence of the source area in the Callovian and the beginning of the Oxfordian. The maximum distribution of Jurassic deposits in the Polish Basin, during the Middle Callovian and Early Oxfordian, was driven by a relative sea-level rise, likely due to extensional tectonic movements associated with the opening of the Tethys Ocean ([Lewandowski et al., 2005](#); [Matyja and Wierzbowski, 2006](#)). In the Middle Callovian, geochemical anomalies and stratigraphically condensed, nodular layers were caused by dissolution processes that hindered Callovian carbonate deposition. These “nodular beds” are represented by marly and/or sandy limestones, calcareous sandstones with common marine fauna, and chamosite and/or glauconite ([Premik, 1933](#); [Znosko, 1957, 1968](#); [Feldman-Olszewska, 1997b](#)). These sediments dominated until the Early Oxfordian when carbonate sedimentation accelerated.

Lower Jurassic sandy rocks are mainly represented by quartz, sublithic, and subarkosic arenites ([Teofilak, 1960, 1962](#); [Modliński, 1974](#); [Marek and Pajchłowa, 1997](#); [Krystkiewicz, 1999](#); [Kozłowska and Kuberska, 2014](#)). Clay matrix is rare and consists of kaolinite, illite, quartz, organic matter, and iron hydroxides. Among authigenic minerals, quartz overgrowths, kaolinite and various early and late diagenetic carbonates dominate. The diagenetic process of Lower Jurassic sandstones in neighbouring areas occurred in three stages ([Krystkiewicz, 1999](#); [Kozłowska and Kuberska, 2014](#)). The first stage of diagenesis was dominated by early mechanical compaction and dissolution ([Fig. 3](#)). Meanwhile, the generation of kaolinite, quartz, and siderite was initiated. The next generations of calcite, ankerite, siderite, and illite started to appear during the late diagenesis stage. At the same time, the dissolution of unstable grains still continued. The main diagenetic processes dominating in Lower Jurassic sandstones are cementation, compaction and dissolution. Quartz cementation dominates among the other authigenic mineralization types, but clay and carbonates are also common. The presence of rigid quartz overgrowths in the early stages of burial resulted in the preservation of primary porosity; the kaolinization process, however, was the main reason for secondary microporosity generation. Typically, the porosity of the Lower Jurassic sandstones exceeds 20%, accompanied by permeability surpassing 1000 mD ([Kozłowska and Kuberska, 2014](#)).

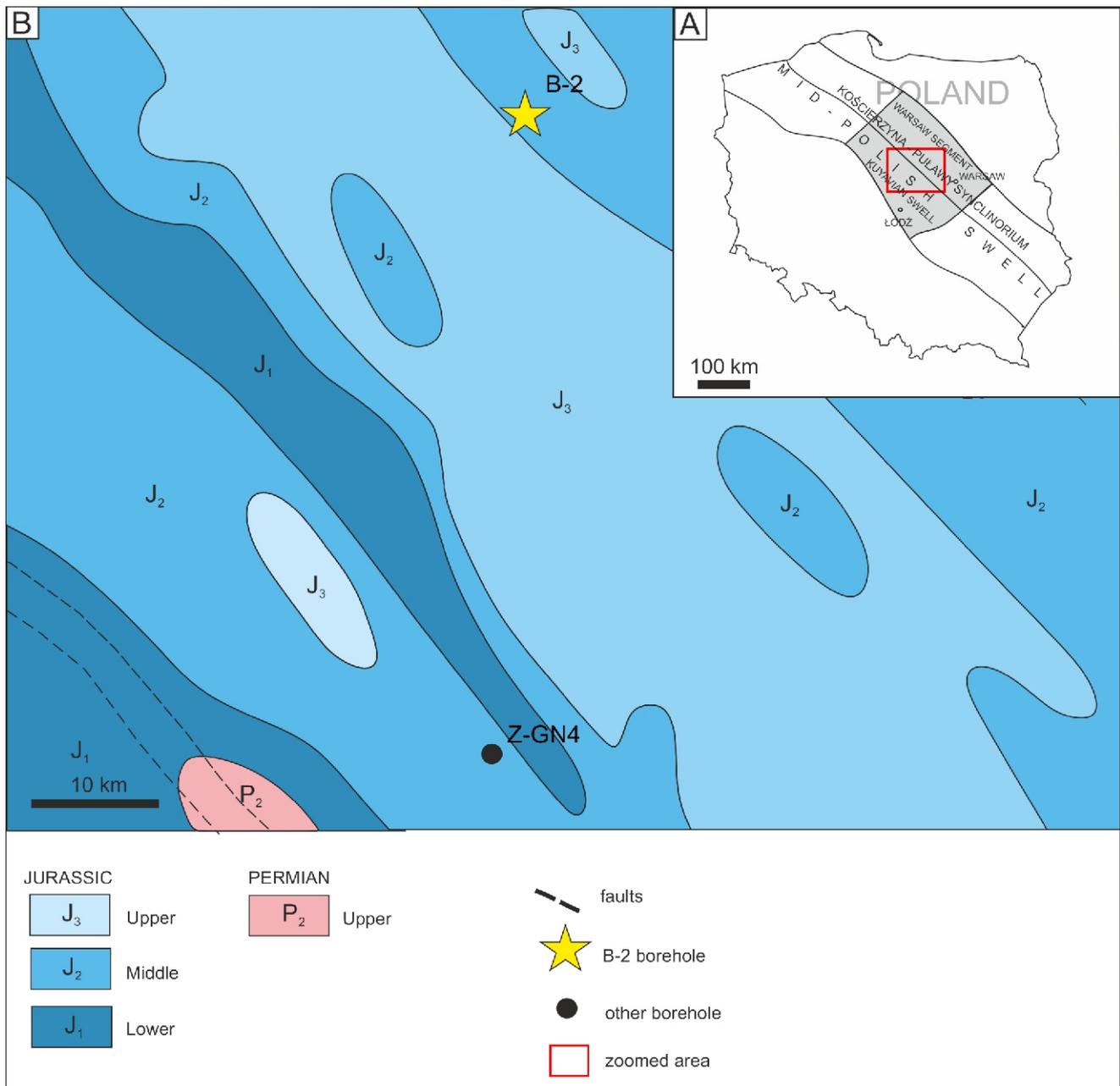


Fig. 1A – location of the research area in Poland (based on Narkiewicz and Dadlez, 2008); B – geological map (2000 m deep) of the Kuyavian segment of the Mid-Polish Swell and Warsaw Synclinorium (based on Kotański, 1997)

The Middle Jurassic sandy rocks consist mainly of sandstones and heterolithic deposits with siderite and calcareous sandstone intercalations (Maliszewska, 1999). The main component of the grain framework is quartz with a visible admixture of feldspars and lithic grains. Biotite, muscovite, glauconite, pyrite and heavy minerals are less abundant. Kaolinite and chlorite are dominant components of the clay matrix (Teofilak, 1962; Maliszewska, 1968). The most common diagenetic processes creating the sandy rock structure are compaction, dissolution, cementation alteration, and replacement. Siliceous cementation is common in the lower part of the Middle Jurassic succession. Dolomite, Mn/Fe calcite, siderite, and ankerite dominate in the Bathonian and Callovian deposits (Maliszewska, 1968, 1999). Most of the diagenetic processes took place during eo- and mesodiagenesis (Fig. 3). Authigenic mineraliza-

tion and recrystallization processes such as feldspar overgrowths, carbonate, and hematite were probably caused by the migration of highly mineralized Zechstein fluids (Znosko, 1957; Maliszewska, 1968). The formation of microcrystalline siderite and calcite cement occurred during the early burial stage when the main source of calcite was dissolved and recrystallized calcareous bioclasts. Later, the calcite cement was replaced and/or recrystallized by ankerite and/or dolomite. The infiltration of meteoric water during regional Cretaceous/Paleogene uplift enhanced the processes of kaolinitization of feldspars (Maliszewska, 1999). Secondary dissolution and recrystallization processes contributed to good reservoir properties, with sandstone porosity typically exceeding 10% and permeability reaching up to 700 mD.

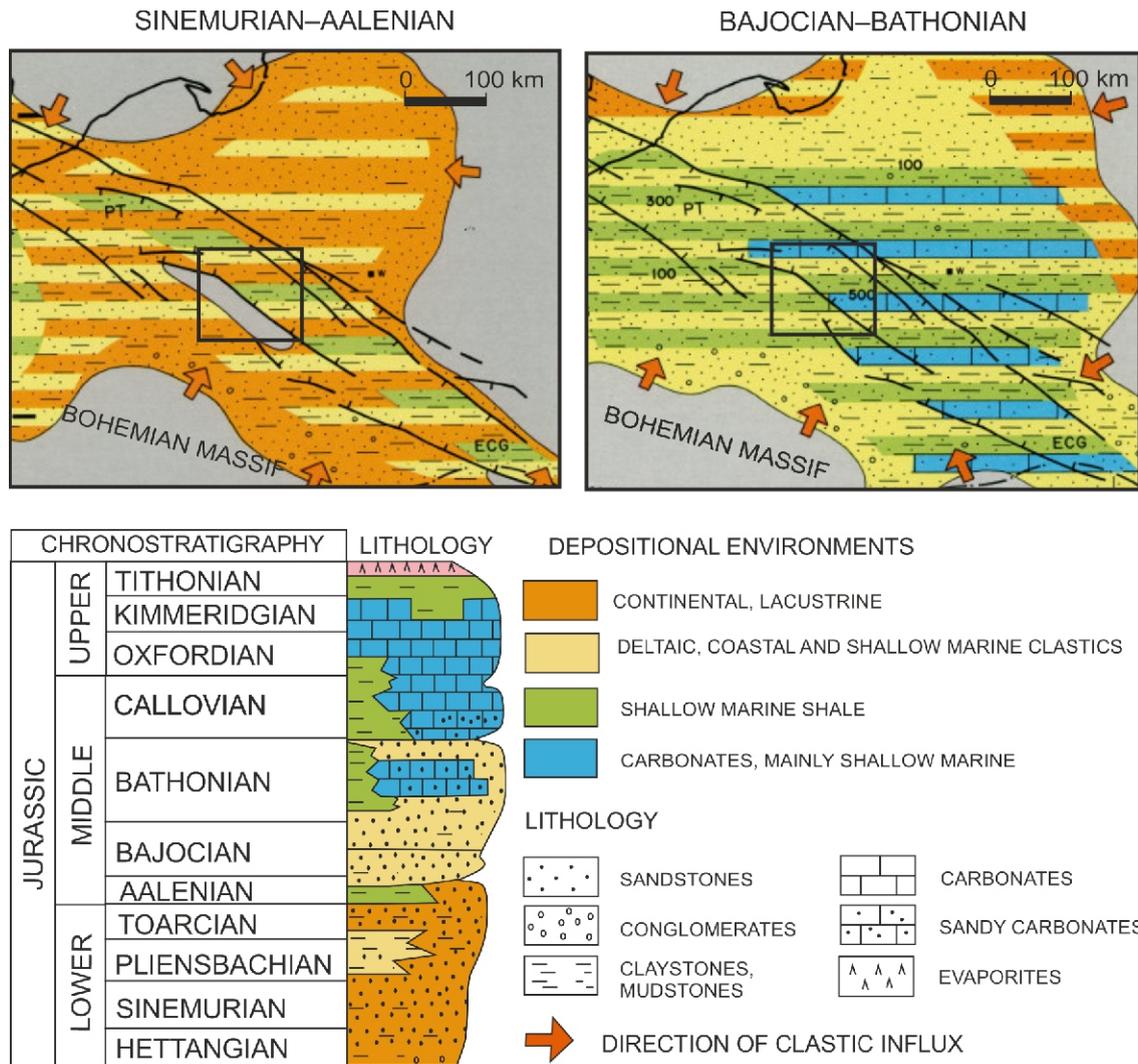


Fig. 2. Location of the research area on the palaeogeographical map of the Lower and Middle Jurassic in Poland (Ziegler, 1982; modified)

## MATERIALS AND METHODS

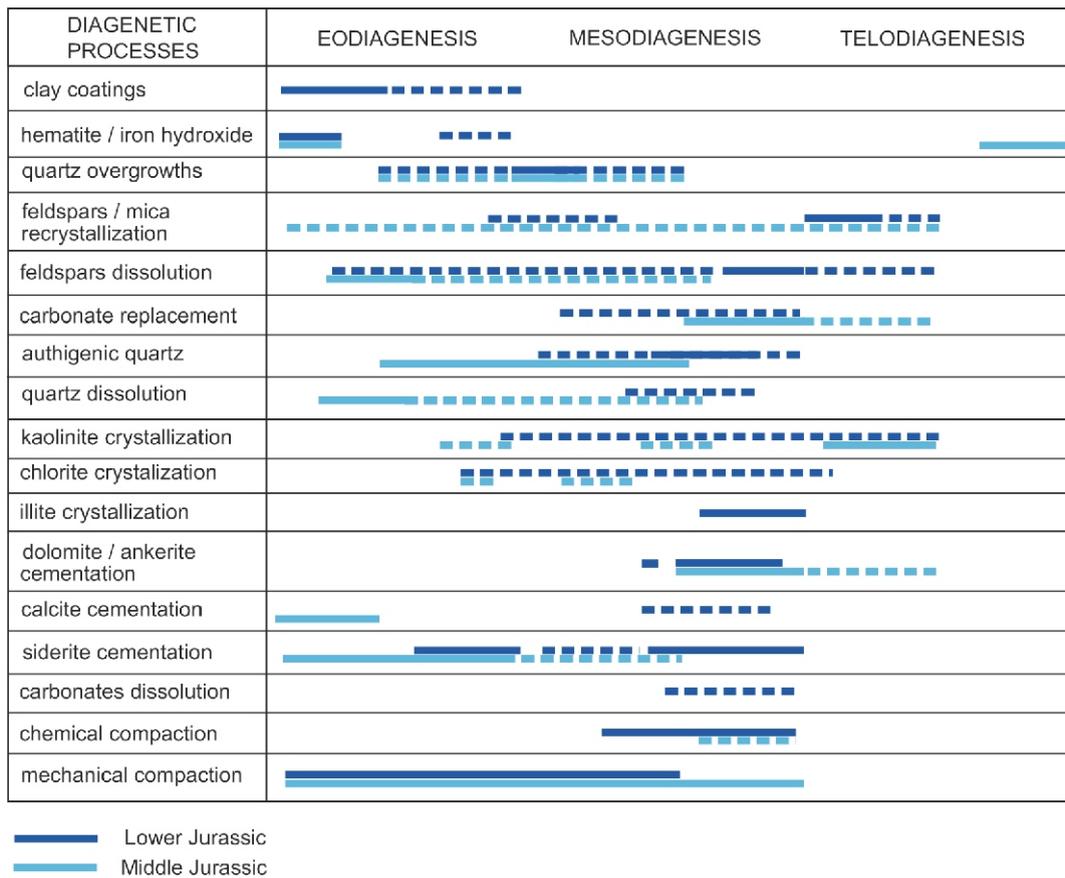
The main components of this study included an analysis of well logs, petrographic examination of thin sections, and macroscopic inspection of drilling core samples. The B-2 section was selected due to its substantial thickness (1472.5 m) of the Jurassic clastic succession, the availability of core samples, lab data and calibrated well logs.

Macroscopic analysis of 37.3 m of core was conducted to properly correlate the well data. Lithology, grain size of clastic rocks, colour, sedimentary structures, HCl reaction, presence of charred plant detritus, and visible mineral components of sandy rocks were considered. Most of these, such as muscovite, glauconite, kaolinite, and potassium feldspars, have a significant influence on petrophysical parameters and can cause disturbances in well-log measurements (Leonowicz, 2005; Kozłowska and Maliszewska, 2015).

This was followed by a petrographic analysis of 24 thin sections of clastic rocks. The description focused on the characteristics of the original sandy rocks, such as grain size, mineral composition, matrix type and content, texture, and diagenetic features such as cements, grain contacts, and products of mineral transformations, which are essential for determining the reservoir parameters of sandstones. To classify the rocks, the

percentages of quartz, feldspar and lithic grains were determined by planimetry based on 300 points per sample. Each sample was described according to the Pettijohn et al. (1974) sandstone classification.

Basic geophysical well log analysis was carried out using TechLog software provided by Schlumberger. Additionally, scripts in the Python programming language were prepared to perform data analysis, outlier detection, and k-means clustering, using code packages such as Pandas (McKinney, 2010), NumPy (Harris et al., 2020), SciPy (Virtanen et al., 2020) and Scikit-Learn (Pedregosa et al., 2011). For this study, the measurements of borehole diameter (caliper log), natural gamma ray (GR), deep resistivity (LLD), interval transit time (DT), bulk density (RHOB), sandstone-scale neutron porosity (NPHIs), and spectral gamma ray (potassium (K), uranium (U), and thorium (Th)) were used as the most valuable means of distinguishing petrofacies. In the first stage, the quality of the measurements was controlled manually. The caliper log was used to distinguish depth intervals with elongation of the borehole walls (bad hole), which was determined when the caliper was more than 50 mm larger than the nominal bit size. Since the poor condition of the borehole wall greatly affects the overall accuracy of geophysical well logging, zones detected with such significant increase in the borehole diameter were removed



**Fig. 3. Diagenetic changes in Lower and Middle Jurassic in neighbouring boreholes (Wróblewska, 2022 based on Maliszewska, 1999; Krystkiewicz, 1999; Kozłowska and Kuberska, 2014)**

Dotted line – the uncertain extent of diagenetic transformations

from further analysis, reducing the number of data points examined from 9660 to 8163. After removing bad hole data points, the remaining dataset was standardized by removing the mean and scaling to unit variance. This procedure allows logs with different value ranges to be used, since the centering and scaling are done independently on each feature by computing the appropriate statistics on the samples (Pedregosa et al., 2011). After standardization, a Mahalanobis distance (De Maesschalck et al., 2000) was applied to detect and remove the extreme values in the remaining dataset. The Mahalanobis distance allows the detection of outliers in a multivariate data set by measuring the correlation between the data. A distance between points, which expresses the number of standard deviations an observation is from the mean of the distribution, is calculated by the formula:

$$D(X, \mu) = \sqrt{(X - \mu)^T \Sigma^{-1} (X - \mu)} \quad [1]$$

where:  $X$  – stands for the data matrix,  $\mu$  – is the mean of the variable and  $\Sigma$  – denotes a covariance matrix of the dataset (De Maesschalck et al., 2000; Ghorbani, 2019).

To visually represent a multidimensional dataset of eight geophysical well logs in 2D space, Principal Component Analysis (PCA) was used. PCA is an algorithm that reduces data dimensions while preserving the most variation in the dataset. This goal is achieved by identifying directions of orthogonal variables along which the variation in the data is maximal (Jolliffe, 2005; Ringnér, 2008). The dataset analyzed is expressed in the

form of a data table (matrix  $X$ ) containing  $I$  observations described by  $J$  variables. To reduce the number of  $J$  variables to  $K$  principal components, the calculation requires several computational steps: (1) calculation of the mean of each variable  $J$ , (2) computation of the covariance matrix ( $A = XTX$ ), (3) computation of eigenvectors and eigenvalues (Andrew, 1973) from the covariance matrix ( $A$ ) and sorting the obtained eigenvectors by decreasing eigenvalues, and finally (4) constructing a  $J \times K$  dimensional eigenvector matrix  $W$  and transformation of observations into the new subspace by transposing the matrix  $W$  (Abdi and Williams, 2010; Jolliffe and Cadima, 2016). By using this technique, the eight available geophysical well logs (GR, LLD, DT, RHOB, NPHIs, K, Th, and U) were reduced to just two components (shown as "Principal Component 1" and "Principal Component 2").

The final stage of research was a petrofacies analysis of the clastic rocks represented mostly by sandstones and mudstones. To categorize the standardized measurements from the Lower and Middle Jurassic rocks from the B-2 borehole, a k-means clustering method was applied using multidimensional input data, which were the geophysical well logs noted above. The k-means algorithm was used to group the input data into predefined clusters of similar clastic lithotypes, particularly sandy lithotypes, combined with the results of petrographic and petrophysical studies supplemented by qualitative geochemical results. This approach allowed us to distinguish sandstone petrofacies with specific petrographic and petrophysical characteristics.

## K-MEANS CLUSTERING METHOD IN LITHOLOGY IDENTIFICATION

Distinction of clusters using a k-means algorithm was the basis of the petrofacies analysis performed on borehole B-2. The algorithm is used to divide input data observations with  $n$  features into separate k-groups ( $k$ ) of equal variance by minimizing the sum of the squared Euclidean distances between each data point ( $x_i$ ) and the centroid ( $\mu_j$ ) (centre of the cluster) (Hartigan and Wong, 1979; Eq. 2):

$$\sum_{i=0}^n \min_k (\|x_i - \mu_j\|^2) \quad [2]$$

Calculations are performed in two separate phases. In the first stage, the location of centres of a predefined number of  $k$  clusters is selected by the 'k-means++' method, which selects initial cluster centroids using sampling based on an empirical probability distribution of the points. Then, each input data point is assigned to the nearest cluster, and its distance to the centre is computed. This stage is completed when all input data points are included in one of the clusters. In the next phase, the average of all data points within a particular cluster is recalculated, creating new centroids. After the initial allocation, all data objects are compared to the new centroids and moved to the cluster whose centroid is the closest one. This iterative process continues until the Euclidean distance between the points within the clusters is minimized (Likas et al., 2003; Steinley, 2006; Sinaga and Yang, 2020). The main problem with using the k-means algorithm is an effective selection of the number of clusters, which must be predefined. Since the number of target lithological units that can be distinguished from available geophysical well logs is unknown, control of the number of clusters within this study was performed based on the 'silhouette' method (Rousseeuw, 1987; Pedregosa et al., 2011). Silhouette analysis is a graphical representation of the separation distance between the resulting clusters. The construction of silhouettes is based on the full pairwise distance matrix over all data points, normalized to the number of data points within the particular cluster. The value of the average silhouette score ranges from -1 to 1 and is a reflection of the dissimilarity between points grouped within the same cluster. A value close to 1 implies the data points included within the same cluster have the highest similarity and have been correctly assigned. On the other hand, a value close to -1 indicates that data points were assigned to the wrong cluster because their similarity to points assigned to the same cluster is much lower than to the points grouped in another cluster (Rousseeuw, 1987; Wang et al., 2017).

## PETROGRAPHY OF SANDY ROCKS

Quartz is the main component of the sandy rocks analyzed from the B-2 borehole. These are mainly quartz, subarkosic, and sublithic arenites with subordinate sandy limestones and heterolithic strat. Observed rock fragments are usually polycrystalline quartzite and sandstone grains. Alkaline feldspars are a minor sandstone component, though in some parts of the section their content notably rises to 5–25%. Iron ooids are a common component of sandy limestones in the upper part of the Jurassic profile. Heavy minerals, organic matter, and muscovite grains are rarely visible in the grain framework.

In the Lower and Middle Jurassic rocks of the B-2 section, dominant diagenetic processes include cementation, kaolinization, and dissolution. Carbonate cementation, predominantly represented by ankerite, siderite and calcite, is most conspicuous in the upper section, as indicated by the XRD results found in the borehole report (Karelski and Modzelewski, 1996). In

some samples, carbonate cement exceeds 50% of the rock in the form of poikilotopic cement (Fig. 4A) or rhombohedral crystals (Fig. 4B). Occasionally, carbonates replace unstable, detrital grains. In addition to carbonates, many of the samples analyzed show authigenic quartz (Fig. 4C, D) and feldspar (Fig. 4E) overgrowths. The intensity of quartz cementation has both favourable and detrimental effects on porosity. Early diagenetic cement can mitigate compaction processes and prevent the reduction of primary porosity (Fig. 4C). Conversely, it can fill remaining pores during later diagenetic stages (Fig. 4D). Other observed cements in the lower part of the Jurassic rocks include anhydrite and gypsum (Fig. 4F, G), likely sourced from the underlying salt dome's fluid migration.

Secondary porosity is mostly created by the kaolinization of feldspars (Fig. 5A, B), glauconite and lithic clasts. It can also develop within dissolved iron ooids (Fig. 5C). Microporosity between kaolinite crystals and void spaces inside dissolved grains can significantly enhance the petrophysical parameters of sandstones. In the deeper buried part of the section, compaction and illitization processes are visibly more intense (Fig. 5D). The most common is illitization of muscovite and feldspar grains. All of these primary and secondary sandstone components and alterations can change the geophysical properties of the rocks analyzed.

The porosity of the Jurassic sandstones is notably high, reaching up to 23.26% in the B-2 borehole. This porosity primarily exists as primary porosity, enclosed within isolated pore spaces between authigenically altered quartz grains. In some cases, primary clay coatings in less compacted sandstones help preserve this porosity from quartz overgrowths. Moreover, secondary microporosity caused by the dissolution and kaolinization process is present. Only in some parts of the section does compaction and/or cementation completely infill the pore spaces.

The intensity of these diagenetic changes in the sandstones and the content of some components were determined on a 4-point scale based on microscopic observations (Tables 1 and 2).

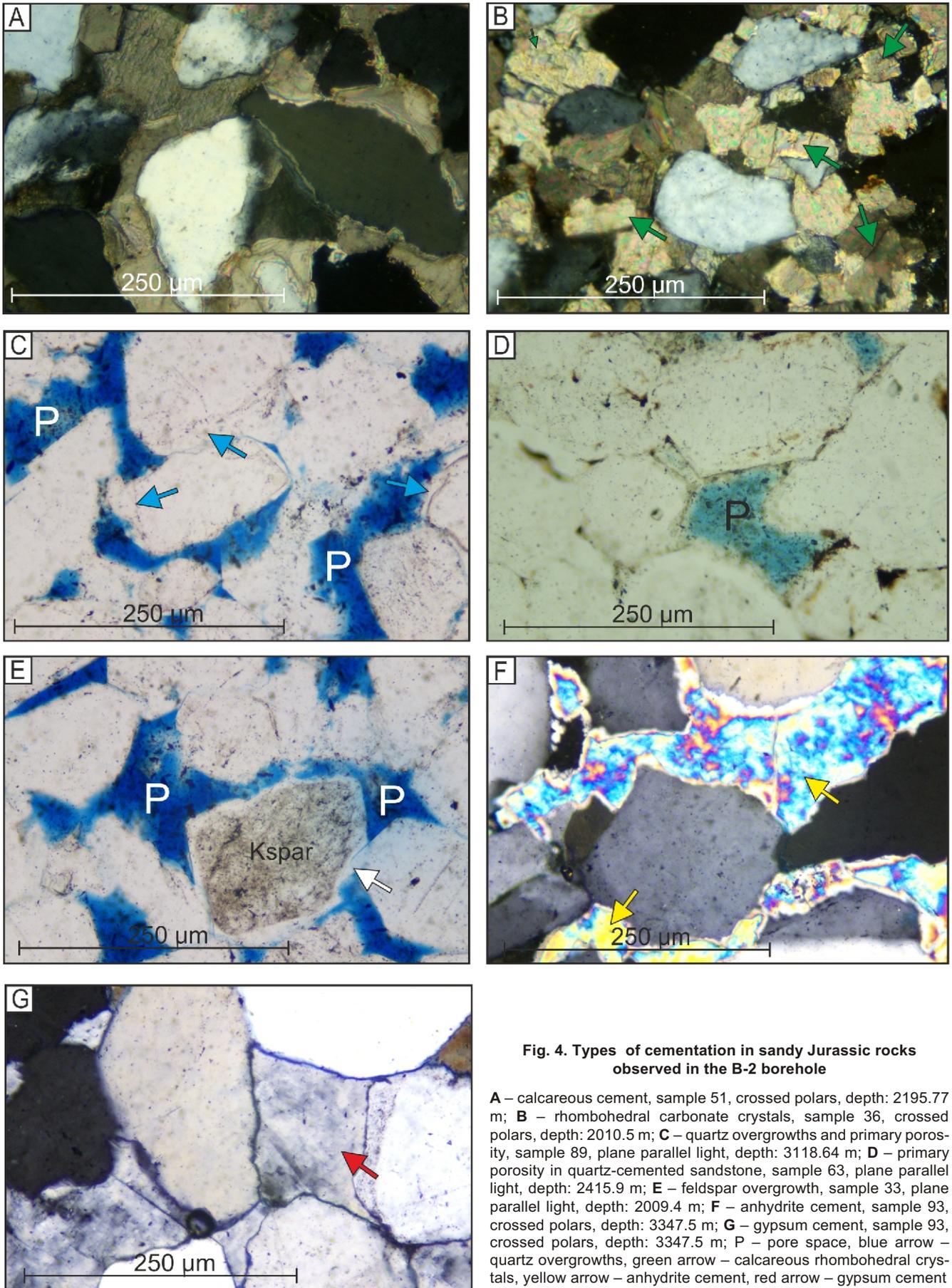
## RESULTS

## INTERPRETATION OF GEOPHYSICAL WELL-LOG DATA

Examining the general condition of the borehole and having the Mahalanobis distance values allowed us to distinguish several factors controlling the occurrence of extreme values. Based on the classical log view (plot of each log vs. depth), we decided that any point with a Mahalanobis distance greater than 1.5 of its standard deviation above the mean should be treated as an outlier (Fig. 6). Marking points as outliers revealed several sources of error, namely from "bad hole" (diameter enlargement), measurement gaps, the presence of claystone, the presence of high resistivity rocks, as well as high-density rocks and coal/carbonate shale (Fig. 7). As a result, an additional 497 of the 8163 points were removed from further analysis, leaving only a set of points representing fine- and coarse-grained clastic rocks.

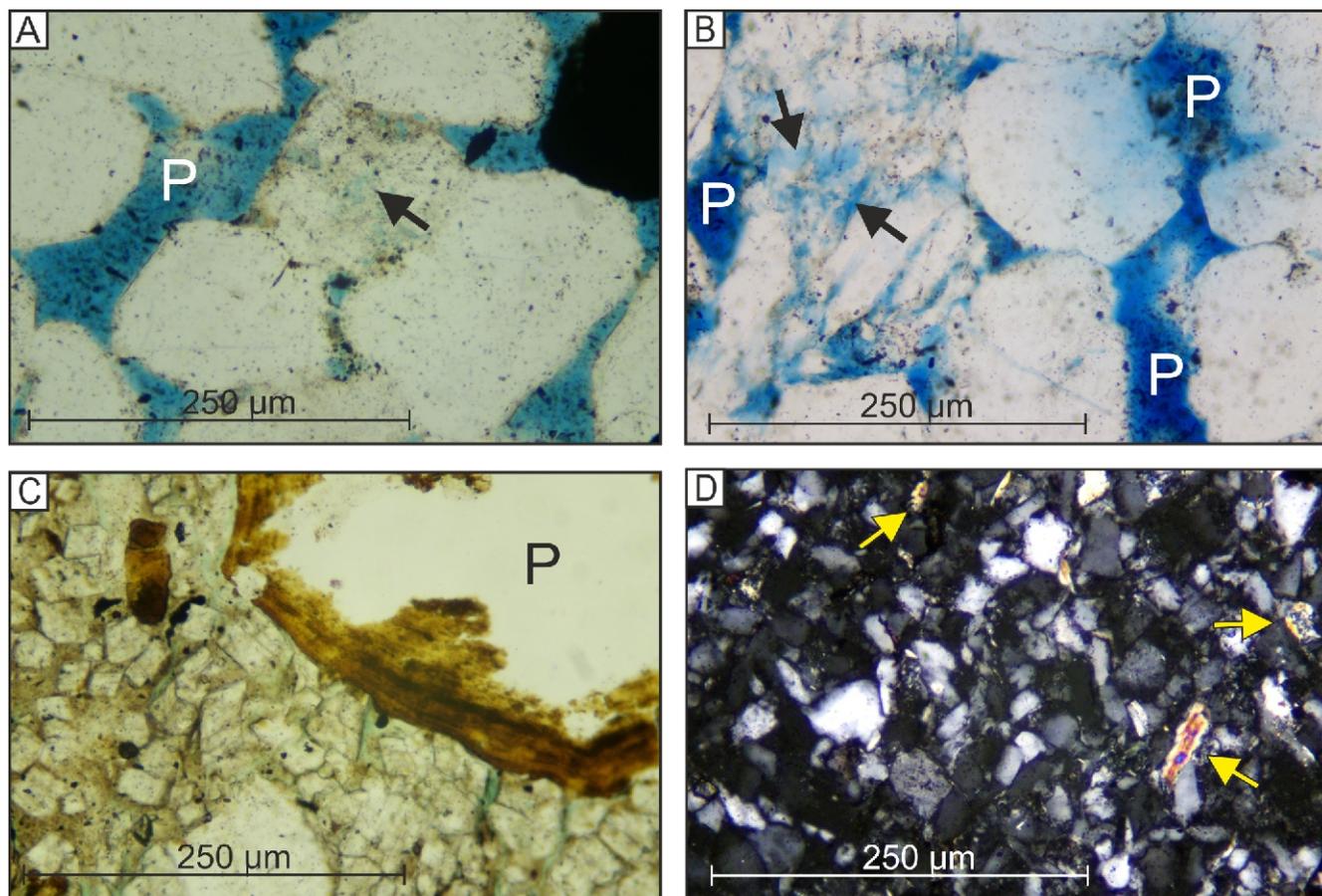
## CLUSTER ANALYSIS

After the evaluation of a selected number of clusters, the k-means algorithm was applied to selected geophysical well logs (GR, DT, LLD, NPHIs, RHOB, K, Th, U) in several steps. In the first iteration, the standardized input dataset was divided into 2 clusters (Cluster 1 and 2), which were pointed by the high-



**Fig. 4. Types of cementation in sandy Jurassic rocks observed in the B-2 borehole**

**A** – calcareous cement, sample 51, crossed polars, depth: 2195.77 m; **B** – rhombohedral carbonate crystals, sample 36, crossed polars, depth: 2010.5 m; **C** – quartz overgrowths and primary porosity, sample 89, plane parallel light, depth: 3118.64 m; **D** – primary porosity in quartz-cemented sandstone, sample 63, plane parallel light, depth: 2415.9 m; **E** – feldspar overgrowth, sample 33, plane parallel light, depth: 2009.4 m; **F** – anhydrite cement, sample 93, crossed polars, depth: 3347.5 m; **G** – gypsum cement, sample 93, crossed polars, depth: 3347.5 m; P – pore space, blue arrow – quartz overgrowths, green arrow – calcareous rhombohedral crystals, yellow arrow – anhydrite cement, red arrow – gypsum cement



**Fig. 5. Types of porosity and diagenetic processes in the sandy Jurassic rocks of the B-2 borehole**

**A, B** – kaolinization process and secondary microporosity, sample 48, crossed polars (C), plane parallel light (D), depth: 2193.68 m; **C** – secondary porosity in iron ooid, sample 33, crossed polars (A), plane parallel light (B), depth: 2009.4 m; **D** – illitization of muscovite grains in highly compacted sandstone, sample 82, depth: 3013.03 m; Qtz – monocristalline quartz, P – pore space, Kspar – potassium feldspars, yellow arrow – illitized muscovite grain

Table 1

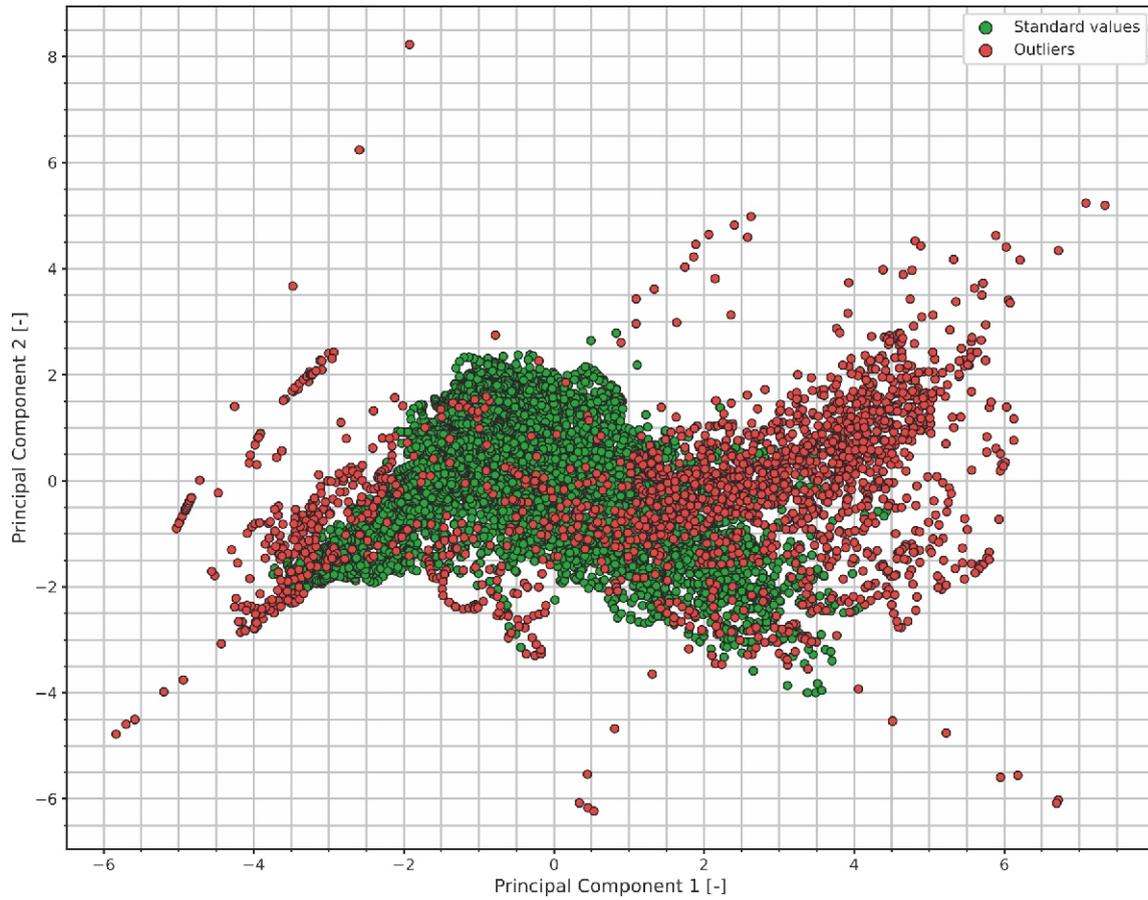
**4-point scale of diagenetic processes intensity and particle content in sandy Jurassic rocks in the B-2 borehole**

Level	Diagenetic process intensity / content of particle					
	Calcareous cementation	Authigenic quartz overgrowths	Clay content (diagenetic/primary)	Compaction intensity	Feldspar content	Thin section porosity
3	Pores completely filled with cement	Thick quartz overgrowths visible on every grain, destroying the primary porosity	>20%	Concavo-convex /sutured grain contacts	>15%	15–20%
2	10–25% of pores filled with cement	Thin quartz overgrowths on every grain	10–20%	Long and concavo-convex	5–15%	10–15%
1	<10% of pores filled with cement	Thin quartz overgrowths on certain grains	5–10%	Long and point	0–5%	5–10%
0	No cementation	No cementation	0–5%	Point	0%	0–5%

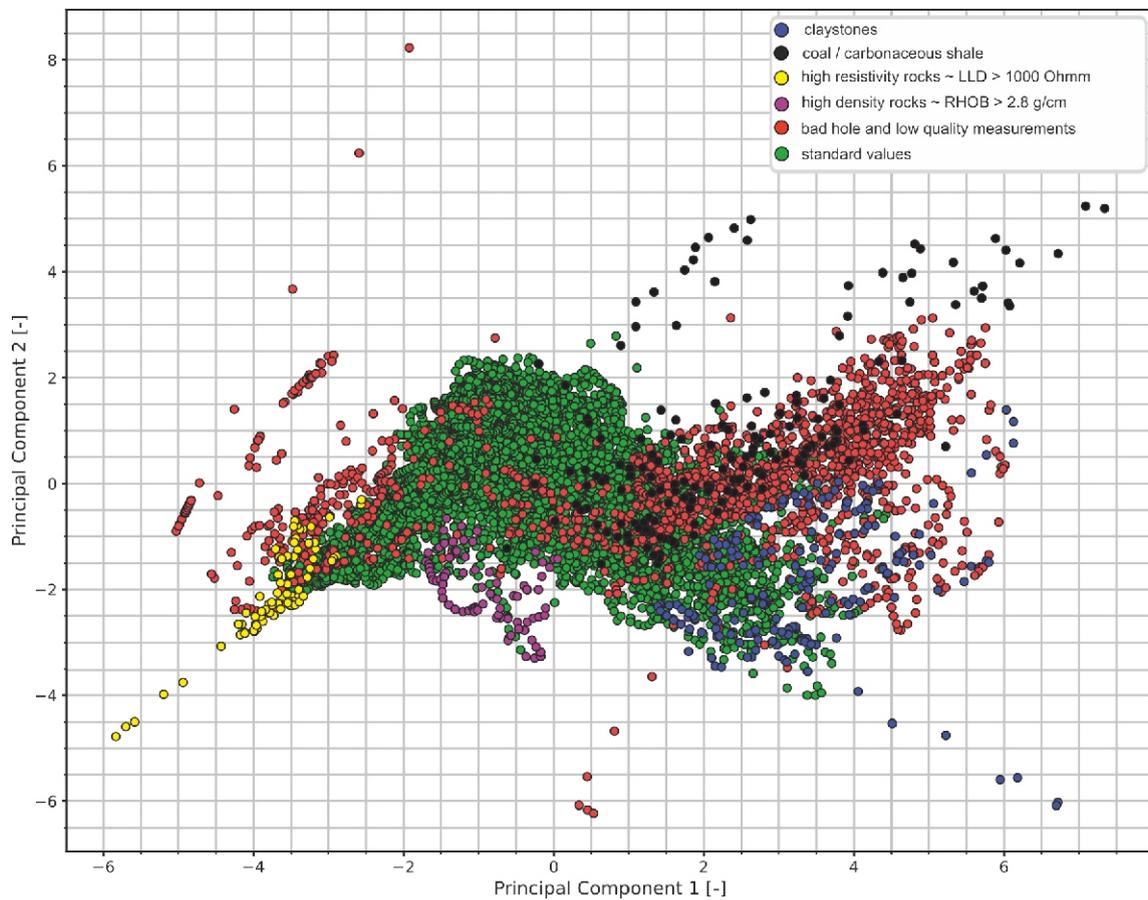
Table 2

## Results of petrographic analysis of sandy rocks from the B-2 core section

Sample nr	md	Sandstone classification (Pettijohn et al., 1974)	Quartz cementation intensity (0–3)	Carbonate cementation intensity (0–3)	Clay content (0–3)	Feldspar content	Thin section porosity (0–3)	Compaction intensity (0–3)	Other features and components
33	2009.4	Calcareous lithic arenite	0	3	0	0	1	0	Iron ooids, chlorite
36	2010.5	Calcareous lithic arenite	0	3	0	0	0	0	
40	2011.7	Calcareous sublith arenite	0	3	0	1	1	0	
43	2012.7	Calcareous sublith arenite	0	3	0	1	1	1	
46	2121.5	Quartz arenite	1	0	1	0	3	1	
48	2193.68	Sublith arenite	2	0	1	1	3	1	Heavy minerals
51	2195.77	Calcareous sublith arenite	0	3	0	0	0	1	
53	2196.77	Quartz arenite	0	0	3	1	1	2	
55	2198.07	Quartz arenite	2	0	1	0	2	2	
56	2411.4	Sublith arenite	1	0	3	0	1	1	
59	2413.2	Sublith arenite	2	0	1	1	3	1	
61	2415.0	Quartz arenite	0	0	1	1	3	1	
63	2415.9	Quartz arenite	3	0	1	0	2	2	
65	2417.1	Quartz arenite	3	0	1	0	2	2	
66	2499.63	Quartz arenite	2	0	1	0	3	1	
67	2500.13	Quartz arenite	1	0	2	1	2	2	
68	2501.03	Quartz arenite	1	1	1	1	2	2	
80	3011.57	Subarkosic arenite	1	0	2	2	0	3	Illitization of mica
82	3013.03	Subarkosic arenite	1	0	2	2	0	3	Illitization of mica, chloritization
84	3014.03	Heterolith	1	0	3	2	0	3	Illitization of mica
89	3118.64	Subarkosic arenite	3	0	0	1	3	2	
89A	3118.99	Quartz arenite	1	0	1	1	1		
91	3262.05	Subarkosic arenite	1	0	2	2	0	3	Illitization of mica
93	3347.5	Sublith arenite	0	0	0	0	1	1	Anhydrite and gypsum cement



**Fig. 6.** The visualization of data points from the Lower and Middle Jurassic interval with highlighted outliers detected by borehole elongation and Mahalanobis distance



**Fig. 7.** The visualization of data points from the Lower and Middle Jurassic interval with highlighted source of detected extreme values

Table 3

Maximum, minimum and mean values (in parenthesis) of geophysical well-logs for 1st and 2nd cluster

Cluster	GR [API]	DT [us/ft]	LLD [ohmm]	NPHIs [v/v]	RHOB [g/cm <sup>3</sup> ]	K [%]	Th [ppm]	U [ppm]
1	18.1–63.0 (30.9)	53.1–96.3 (71.9)	0.5–960.0 (36.1)	0.02–0.30 (0.16)	2.0–2.73 (2.41)	0.5–2.1 (1.0)	0.1–6.9 (1.81)	0.1–3.0 (1.26)
2	41.5–127.9 (74.3)	57.1–98.6 (75.1)	1.05–170.6 (9.0)	0.08–0.40 (0.22)	2.07–2.75 (2.46)	0.7–2.6 (1.66)	0.7–12.3 (5.57)	0.8–4.4 (2.30)

est silhouette score equal to 0.4. The following three iterations concerned the division of the remaining dataset into 2 (Cluster A and B) with a score of 0.45. In the last stage both clusters A and B were divided into 3 and 2 clusters (A1, A2, A3 and B1, B2) with a score 0.27 and 0.33, respectively.

In the following step, cluster analysis of the geophysical dataset was carried out, including successively:

1. The division of the input dataset into 2 clusters (Cluster 1 and Cluster 2).
2. The division of clusters interpreted as coarse-grained clastic rocks (Cluster 1) into 2 clusters (Cluster A and B).
3. The division of clusters interpreted as slightly diagenetically altered (Cluster A), and strongly diagenetically altered (B) into 3 and 2 clusters (A1, A2, A3 and 1B, 2B) respectively.

The first cluster analysis was performed to separate the types of clastic lithotypes. The best match was obtained for 2 clusters (Fig. 8), whose average geophysical properties are given in Table 3.

Clusters 1 and 2 show distinct geophysical characteristics that serve as the basis for their classification as coarse-grained and fine-grained clastic rocks, respectively. Cluster 2 displays notably higher natural gamma-ray and neutron porosity values compared to Cluster 1. Simultaneously, these rocks demonstrate lower deep resistivity values, leading to their classification as fine-grained clastic rocks. The elevated gamma-ray values can be attributed to the presence of potassium and thorium in

the clay mineral structure (Table 3). In certain cases, the presence of uranium, either associated with stratigraphic condensation in fine-grained rocks or with heavy minerals such as monazite or zircon, may contribute to the observed increase. Additionally, clay minerals are responsible for the higher neutron porosity values and the corresponding decrease in resistivity, primarily due to the presence of bound water absorbed by the clay surface within the electrical double layer.

In the next step, coarse-grained rocks (Cluster 1) were subdivided to differentiate sandstones with favourable reservoir properties. The best fit was obtained for 2 clusters, A and B (Fig. 9), with the typical geophysical values shown in Table 4.

The most significant differences between clusters A and B are primarily evident in the deep resistivity, interval transit time, neutron porosity, and bulk density logs. A detailed analysis of geophysical well log responses facilitated their categorization into slightly diagenetically altered (Cluster A) and strongly diagenetically altered (Cluster B) sandstones, which can be further classified as porous and slightly/non-porous. Strongly diagenetically altered sandstones (B) are characterized by strong carbonate, sulphate, and/or silica cementation, primarily in the form of authigenic syntaxial overgrowths on detrital quartz grains. These rocks, regardless of the type of cementation and the level of compaction, show low natural gamma radioactivity, which is attributed to the predominance of non-radioactive components in their mineral composition and minimal clay mineral content. They also display low porosity, resulting from diage-

### Silhouette analysis for k-means clustering on sample data with n clusters = 2

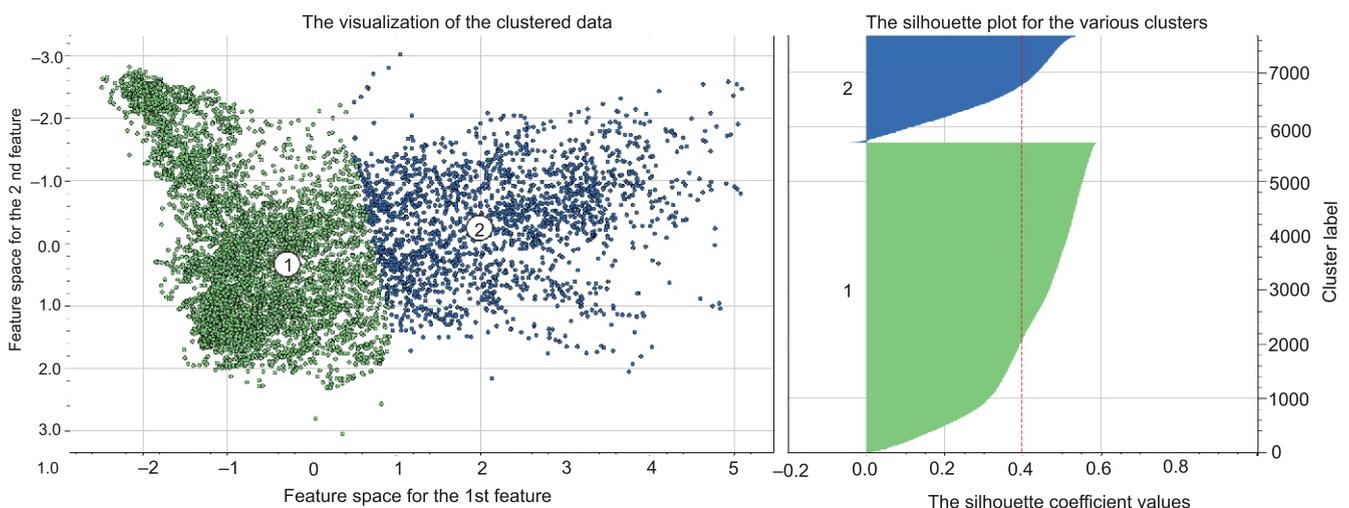


Fig. 8. The visualization of clustered data and silhouette plot for 2 clusters (label 1 and 2) of coarse- and fine-grained clastic rocks

Cluster centres are marked on the scatter plot by white circles with cluster labels; red dashed line on silhouette plot is an average score, which is 0.4

Table 4

Maximum, minimum and mean values of geophysical well-logs for cluster A and B of coarse-grained clastics

Cluster	GR [API]	DT [us/ft]	LLD [ohmm]	NPHIs [v/v]	RHOB [g/cm3]	K [%]	Th [ppm]	U [ppm]
A	18.1-61.2 (32.2)	62.0-96.3 (76.1)	0.49-186.3 (2.15)	0.08-0.31 (0.20)	2.01-2.61 (2.35)	0.58-2.10 (1.06)	0.05-6.91 (1.82)	0.16-2.72 (1.27)
B	19.7-63.0 (27.5)	53.1-73.4 (60.4)	1.26-960.0 (130.7)	0.02-0.21 (0.02)	2.31-2.73 (2.60)	0.5-1.9 (0.91)	0.3-6.9 (1.8)	0.3-3.0 (1.3)

Silhouette analysis for k-means clustering on sample data with n clusters = 2

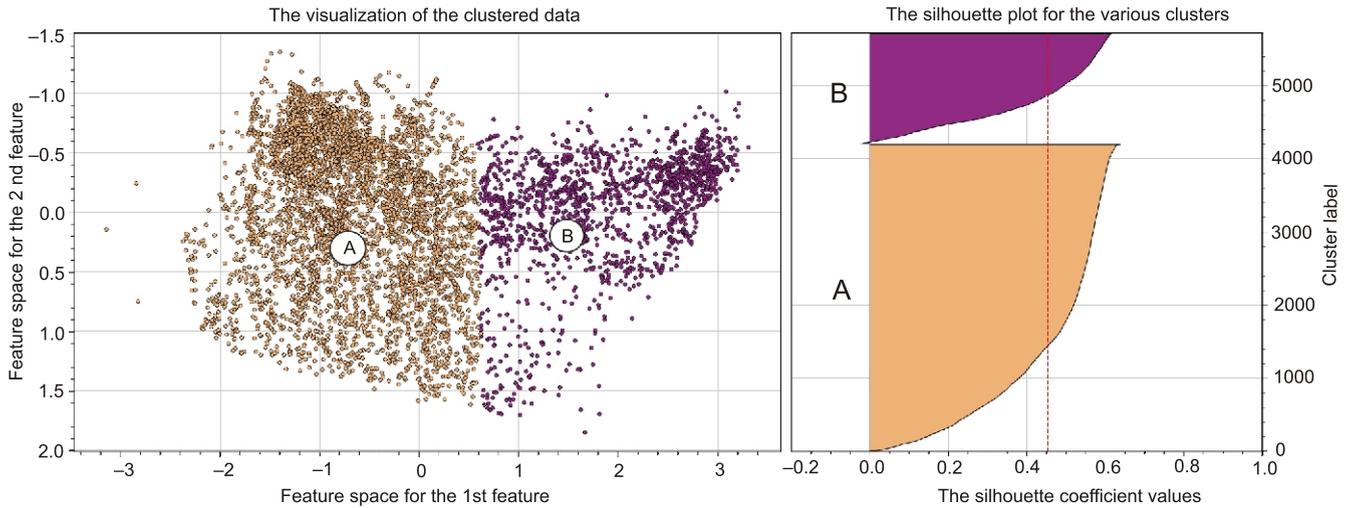


Fig. 9. Visualization of clustered data and a silhouette plot for A and B clusters of coarse-grained clastic rocks

Cluster centres are marked on the scatter plot by white circles with cluster labels;  
red line on silhouette plot is an average score, which is 0.45

Silhouette analysis for k-means clustering on sample data with n clusters = 3

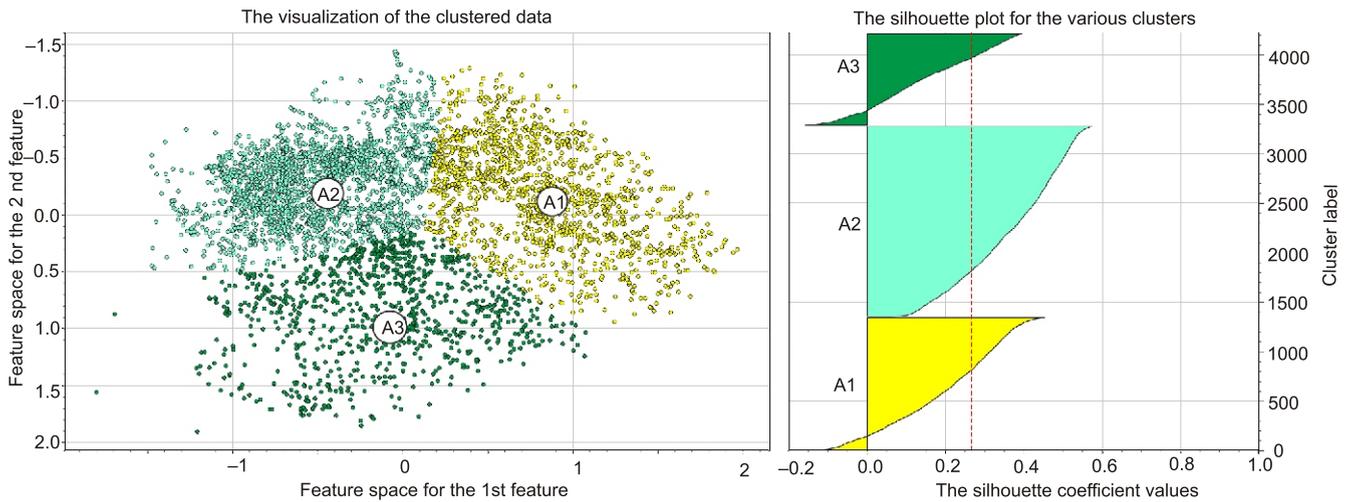
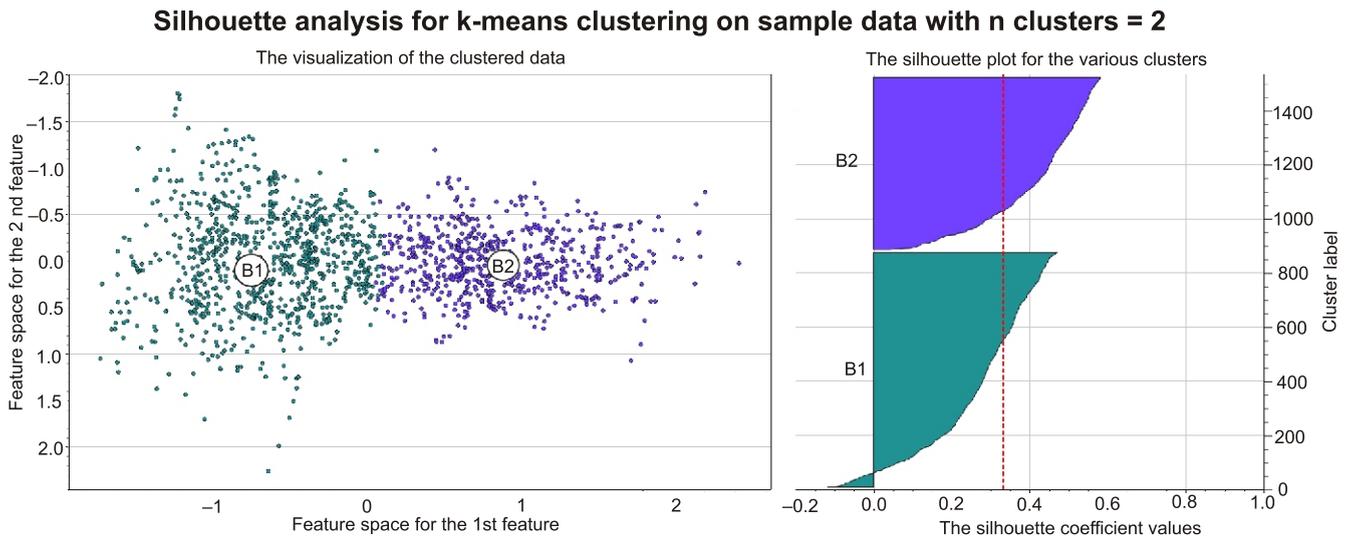


Fig. 10. Visualization of clustered data and silhouette plot for 3 clusters of slightly diagenetically altered sandstones of the Lower and Middle Jurassic in the B-2 borehole

Cluster centres are marked on the scatter plot by white circles with cluster labels;  
red line on silhouette plot is an average score, which is 0.27



**Fig. 11. Visualization of clustered data and silhouette plot for 2 clusters of strongly diagenetically altered sandstones of the Lower and Middle Jurassic in the B-2 borehole**

Cluster centers are marked on the scatter plot by white circles with cluster labels; red line on silhouette plot is an average score, which is 0.33

netic mineralization or limited secondary porosity within dissolved grains. Due to their predominantly compacted structure, they feature low interval transit time (DT) and NPHIs values, along with high resistivity (LLD). Additionally, cemented sandstones may have elevated average bulk densities (RHOB) due to the frequent presence of dense minerals such as carbonates or sulfates.

The final step in the analysis was to differentiate the five sandstone petrofacies within the two groups mentioned above. One of our goals was to identify the petrofacies with the best petrophysical parameters, so at this point we focus more on the less diagenetically altered sandstones represented by one of

the identified clusters. These less altered rocks show very discrete differences in geophysical measurements that may have been missed in the process of clustering all the sandstones in one run. Therefore, we decided to treat the two clusters as different groups and performed further clustering on them separately to detect their internal division. This process resulted in the identification of 3 and 2 clusters for each sandstone group, respectively (see Figs. 10 and 11). The final interpretation of the well log responses for each cluster served as the basis for distinguishing 5 distinct sandy petrofacies (Figs. 10–13 and Table 5), referred to as P1-P5 in the following sections.

Table 5

**Maximum, minimum and mean values of geophysical well-logs for the A1-3 and B1-2 clusters with petrofacies (P1–P5) interpretation results of the Lower and Middle Jurassic sandstones in the B-2 borehole**

Cluster	GR [API]	DT [us/ft]	LLD [ohmm]	NPHIs [v/v]	RHOB [g/cm3]	K [%]	Th [ppm]	U [ppm]	Petrofacies
A1	18.1–44.2 (25.3)	68.8–87.5 (76.9)	0.49–137.7 (1.2)	0.14–0.28 (0.21)	2.12–2.49 (2.31)	0.6–1.4 (0.9)	0.1–3.2 (1.2)	0.2–2.1 (1.1)	P1– quartz arenites with preserved high primary porosity
A2	20.0–61.2 (37.1)	62.0–81.2 (71.0)	0.7–186.3 (3.9)	0.08–0.26 (0.16)	2.21–2.61 (2.42)	0.6–2.0 (1.2)	0.3–6.9 (2.4)	0.2–2.7 (1.4)	P2– quartz, subarkosic and sublithic arenites with kaolinite/clay coatings and secondary porosity
A3	23.7–60.3 (39.7)	72.5–96.3 (81.8)	0.54–13.4 (1.6)	0.16–0.30 (0.22)	2.0–2.54 (2.3)	0.7–2.1 (1.28)	0.4–5.2 (2.3)	0.4–2.5 (1.4)	P3 – subarkosic and sublithic arenites with high porosity
B1	19.8–63.0 (30.2)	54.6–73.4 (63.4)	1.3–330.7 (17.0)	0.02–0.21 (0.09)	2.31–2.72 (2.54)	0.5–1.9 (0.9)	0.4–7.0 (2.1)	0.3–3.0 (1.3)	P4 – cemented quartz, subarkosic and sublithic arenites with slight porosity
B2	20.2–34.5 (23.9)	53.1–64.4 (56.3)	12.8–960.0 (285.3)	0.02–0.1 (0.03)	2.33–2.73 (2.65)	0.5–1.3 (0.9)	0.3–2.8 (1.4)	0.3–1.8 (1.12)	P5 – non-porous, strongly cemented and/or compacted quartz arenites

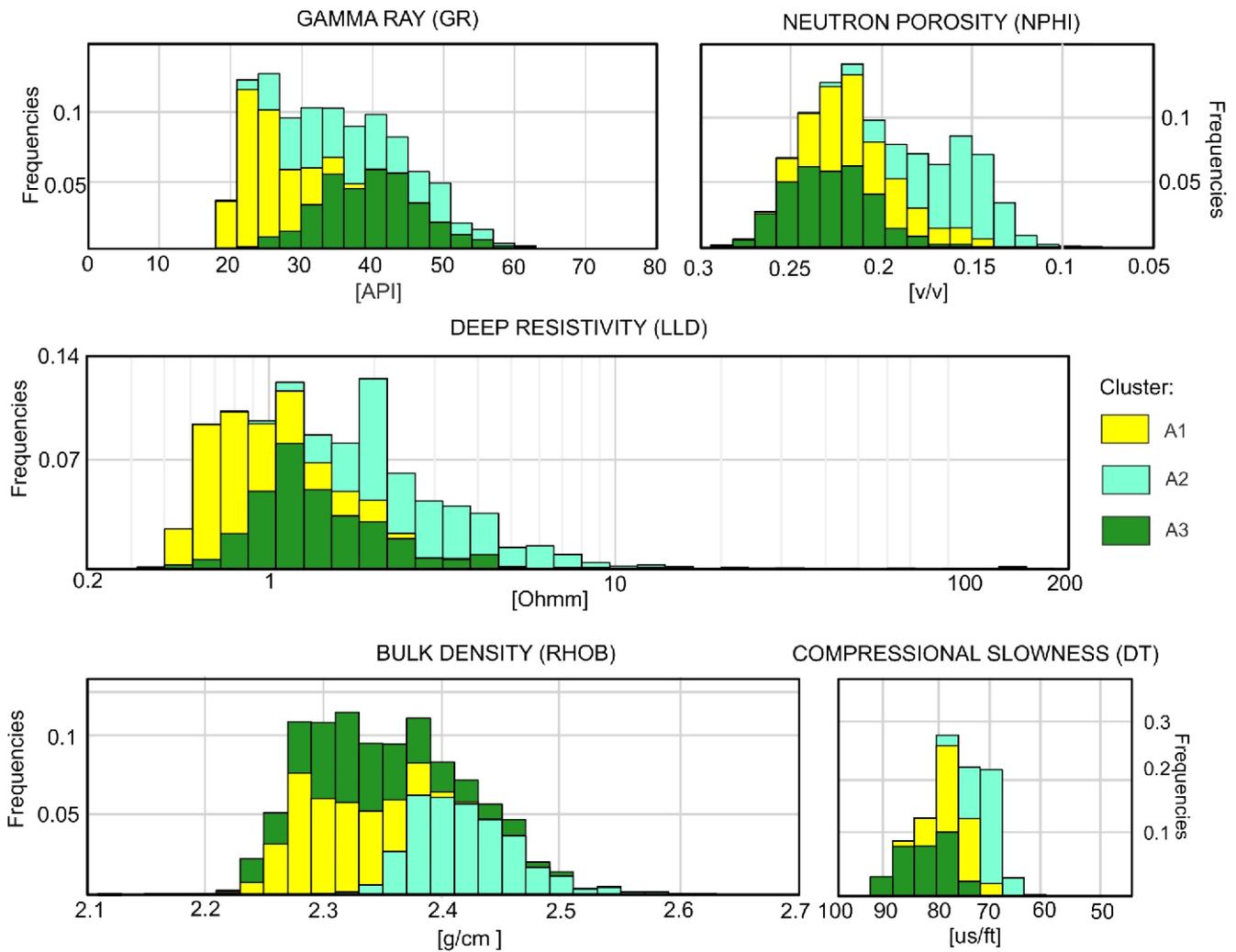


Fig. 12. Histograms of geophysical properties distribution in the A1–A3 clusters

INTERPRETATION OF GEOPHYSICAL AND RESERVOIR PARAMETERS  
OF P1–P5 PETROFACIES

QUARTZ ARENITES WITH PRESERVED HIGH PRIMARY POROSITY – P1

Petrofacies P1 shows the best reservoir properties within the entire Jurassic profile with a mean value of total calculated porosity of 20.5%. The pore throat radius exhibits the highest values, though is the most varied among all the sandy lithotypes with a clear dominance of large (>15µm) macropores with insignificant participation of micropores, most probably in more intensely cemented and/or compacted quartz arenites (Fig. 14). The high porosity of the P1 petrofacies is characterized by the lowest natural gamma radioactivity and resistivity (Table 5). Simultaneously, it demonstrates very low bulk density and one of the highest NPHI values (Fig. 12). These characteristics are typical of porous quartz arenites with little or no quartz cementation allowing for the preservation of primary porosity even in deeply buried strata. Based on the analysis conducted, this lithological type predominates in the depth interval from 2117.0 m to 2795.0 m (Fig. 15), corroborating the predominantly quartz composition of these rocks (Fig. 16). In thin sections, they show slight diagenetic alteration with varying proportions of quartz mineralization (Fig. 17). Similar porous quartz

arenites with preserved primary porosity were previously described by Wróblewska (2022) as the S2 petrofacies in the Z-GN4 borehole.

QUARTZ ARENITES WITH KAOLINITE/CLAY COATINGS  
AND SECONDARY POROSITY – P2

Kaolinization is commonly observed in the Jurassic sandstones of the B-2 core section. Kaolinite is a product of the alteration of feldspar, glauconite and lithic grains (Churchman and Lowe, 2012). This mineral does not produce as much natural radioactivity as the other clay minerals, so the total gamma ray value of the kaolinized sandstone is visibly lower than the other clayey sandy petrofacies. However, residual feldspar, glauconite fragments and/or clay coatings may slightly increase their radioactivity due to their potassium and thorium content. Petrofacies P2 exhibits moderately high porosity, as indicated by low resistivity and elevated neutron porosity values with a dominance of 4–8 µm macropores (Table 5 and Fig. 14). This porosity type is probably mostly secondary with the primary porosity saved by the presence of clay coatings that prevented the generation of quartz overgrowths. The increased content of primary and diagenetic clay from dissolution of feldspar grains weakens the sandstone grain framework, resulting in greater

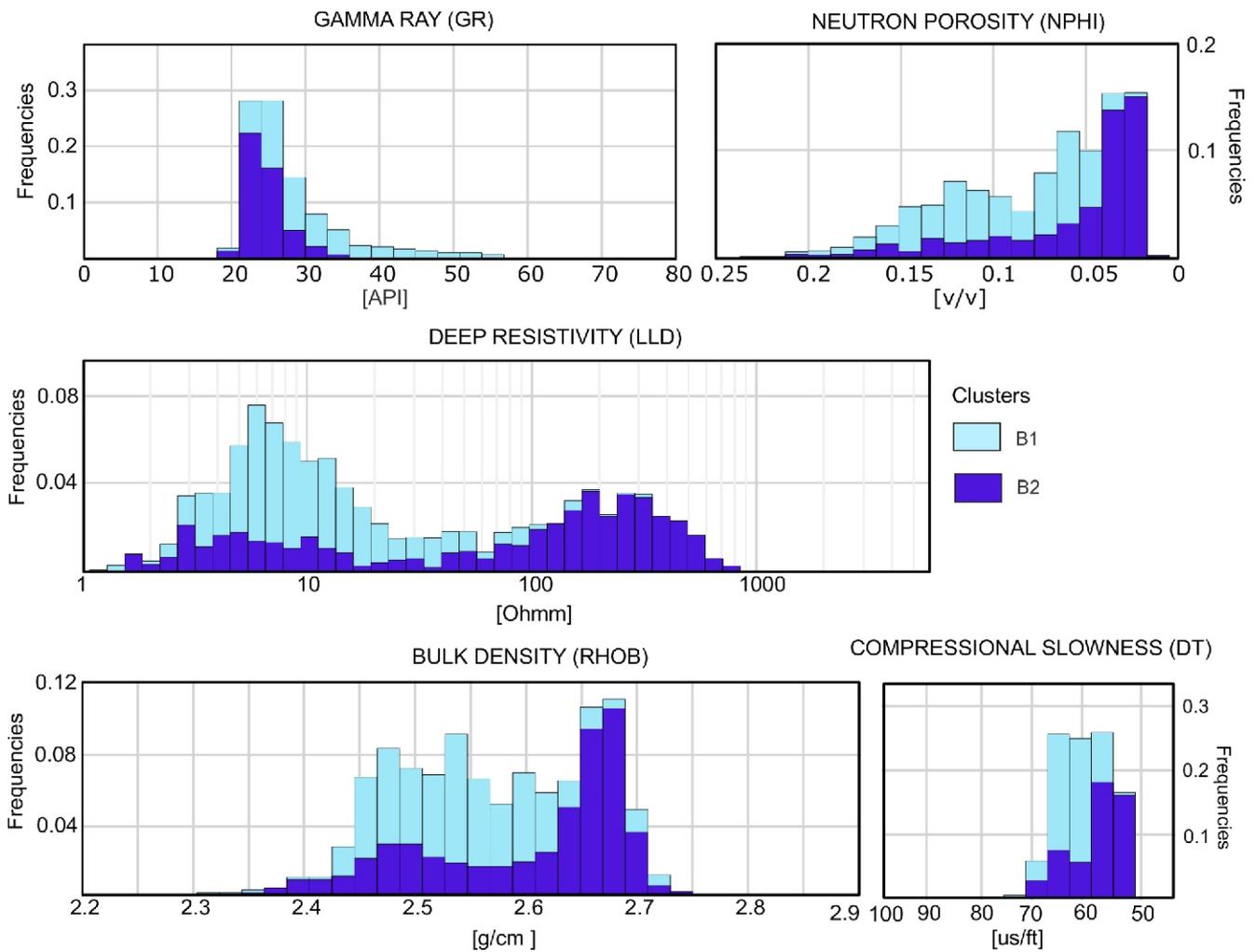


Fig. 13. Histograms of geophysical properties distribution in the B1 and B2 clusters

compaction and consequently in the reduction of porosity (Fig. 14). Based on the petrofacies analysis, this lithological type is predominant in the depth interval ranging from 2795.0 to 3148.0 m (Fig. 15). This observation is supported by increased potassium feldspar and kaolinite contents, as indicated by archival qualitative XRD and petrographic analysis results (Figs. 16–18). Petrofacies P2 can be compared to the S1 sandstone lithotype with non-radioactive clay minerals and secondary porosity distinguished by Wróblewska (2022) in the Z-GN4 borehole.

#### SUBARKOSIC AND SUBLITHIC ARENITES WITH HIGH POROSITY – P3

Subarkosic and sublithic arenites are common rock types among the Jurassic sandstones. The increased content of potassium-rich minerals, such as residual fragments of potassium feldspars and muscovite, together with heavy minerals and clay derived from the alteration of these particles, make these petrofacies slightly more radioactive than the P1 and P2 petrofacies. This rock type ranks second in terms of porosity after the P1 petrofacies with a mean porosity of 19% with dominating pore throat radius exceeding 20  $\mu\text{m}$  (Fig. 14). This characteristic is evident in the noticeable decrease in bulk density and resistivity curves, coupled with an increase in neutron po-

rosity and interval transit time (Table 5). The pore space mostly appears to be secondary, taking the form of microporosity between kaolinite grains, accompanied by relatively high intergranular primary porosity. The primary distinguishing feature between the P2 and P3 petrofacies likely lies in the degree of grain dissolution and the volume of primary porosity, with P2 being the most diagenetically altered and dissolved among them. Petrofacies analysis has revealed that this petrofacies is subordinate to the P1 petrofacies in the 2008.0–2799.0 m interval while dominating mainly in the Middle Jurassic profile (Figs. 15–17). According to the results obtained by Wróblewska (2022) based on archival well data together with the “clay difference” method, petrofacies P3 should probably be directly compared to the S4 petrofacies of subarkosic and sublithic sandstones and part of the S3 petrofacies representing sandstones with preserved primary porosity or dissolved radioactive particles excluding cemented intervals.

#### CEMENTED QUARTZ, SUBARKOSIC AND SUBLITHIC ARENITES WITH SLIGHT POROSITY – P4

Sandstones within the P4 petrofacies show a wider range of natural gamma radioactivity compared to the P5 petrofacies. This variation is likely due to the presence of variable contents

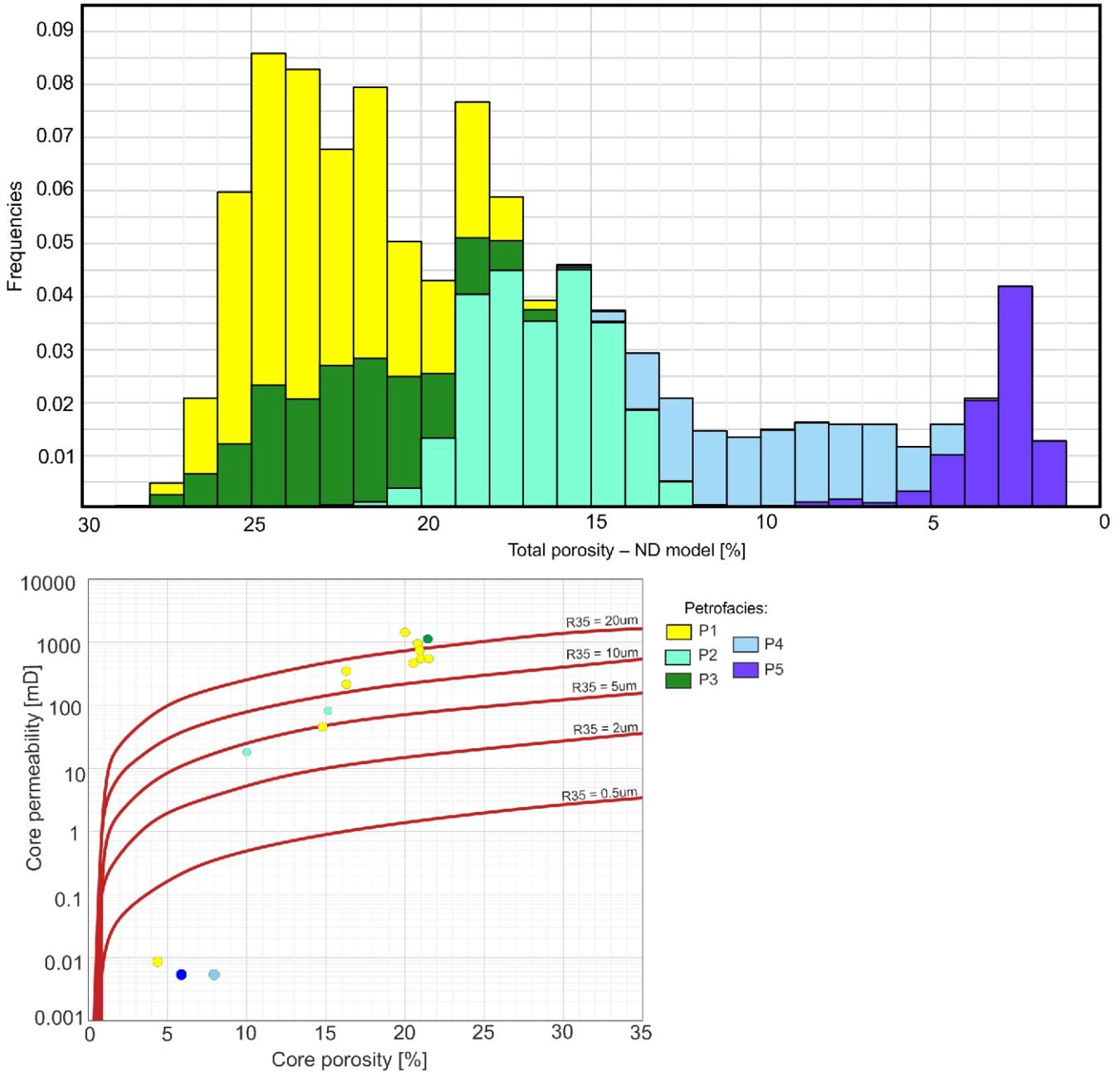


Fig. 14. Results of porosity measurements of core samples compared to the NPHIs – RHOB porosity modelling

of residual feldspar fragments and lithic grains within a grain framework that is cemented by non-radiogenic quartz. Some local carbonate cementation may also be present. The porosity of this lithotype is variable, averaging around 6.5% with a dominance of impermeable micropores (Fig. 14), which is significantly lower than that of the P2 and P3 petrofacies. This lower porosity is likely related to the dissolution of carbonate cement or the presence of residual, closed pore spaces between the secondary, regenerated surfaces of quartz grains. The P4 petrofacies is sporadically distributed throughout the section, with a notably increased proportion in the interval from 3069.0 to 3413.0 m and 2008.0–2138.0 m (see Figs. 15 and 19), adjacent to the P5 and P3 lithotypes respectively. These characteristics are similar to the cemented intervals in the S3 and S4 petrofacies of calcareous subarcosic and sublithic sandstones described by Wróblewska (2022).

NON-POROUS, STRONGLY CEMENTED AND/OR COMPACTED QUARTZ ARENITES – P5

Sandstones of the P5 petrofacies represent sandy rocks with the worst reservoir properties (Fig. 14). The low intensity of natural gamma-ray, increased bulk density of these rocks together with the lowest neutron porosity values indicate a high content of quartz cementation and/or strong compaction. Locally, where density rises to 2.75 g/cm<sup>3</sup>, carbonate or sulphate mineralization is most likely present. The cementation and/or compaction intensity of these sandstones is evident by the highest resistivity indications and lowest DT values among all the sandy petrofacies. The P5 petrofacies predominates in the interval from 3168.0 to 3413.0 m and is characterized by increased carbonate and anhydrite/gypsum cementation (Figs. 15 and 19). This can be compared to the fully cemented and/or compacted S1 (sandstones with non-radioactive clay

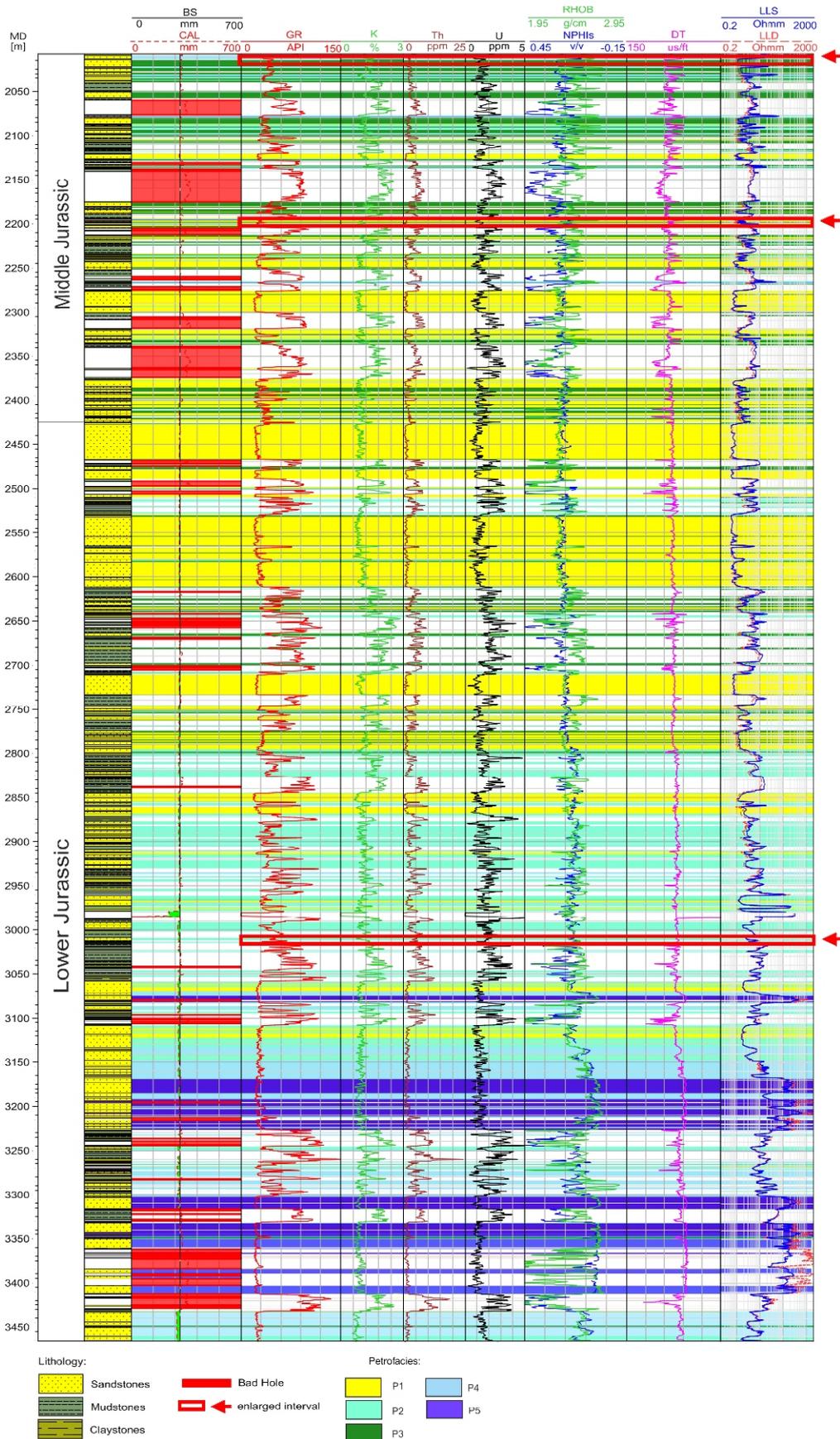


Fig. 15. Results of the petrofacies analysis of the Lower and Middle Jurassic in the B-2 borehole

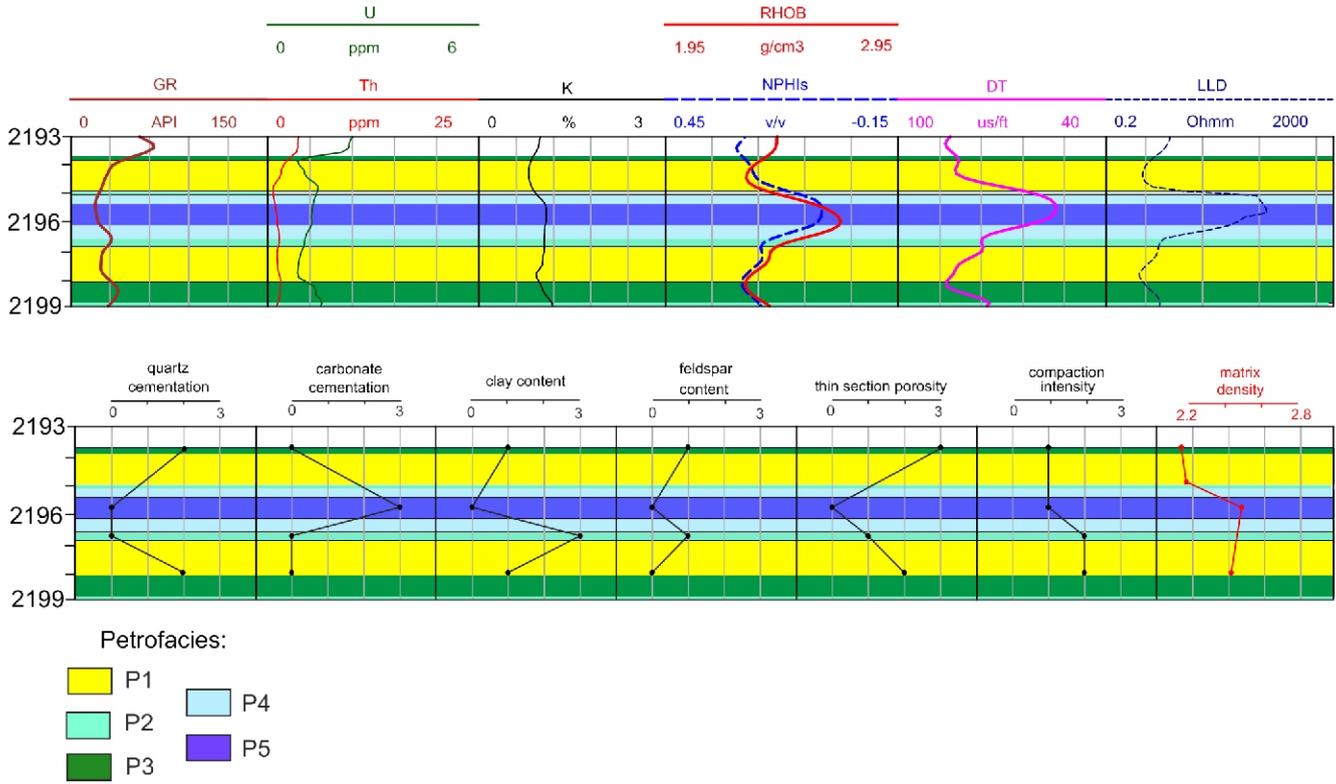


Fig. 16. Log responses compared with the petrographic analysis results in the 2193.0 2199.0 (MD) interval

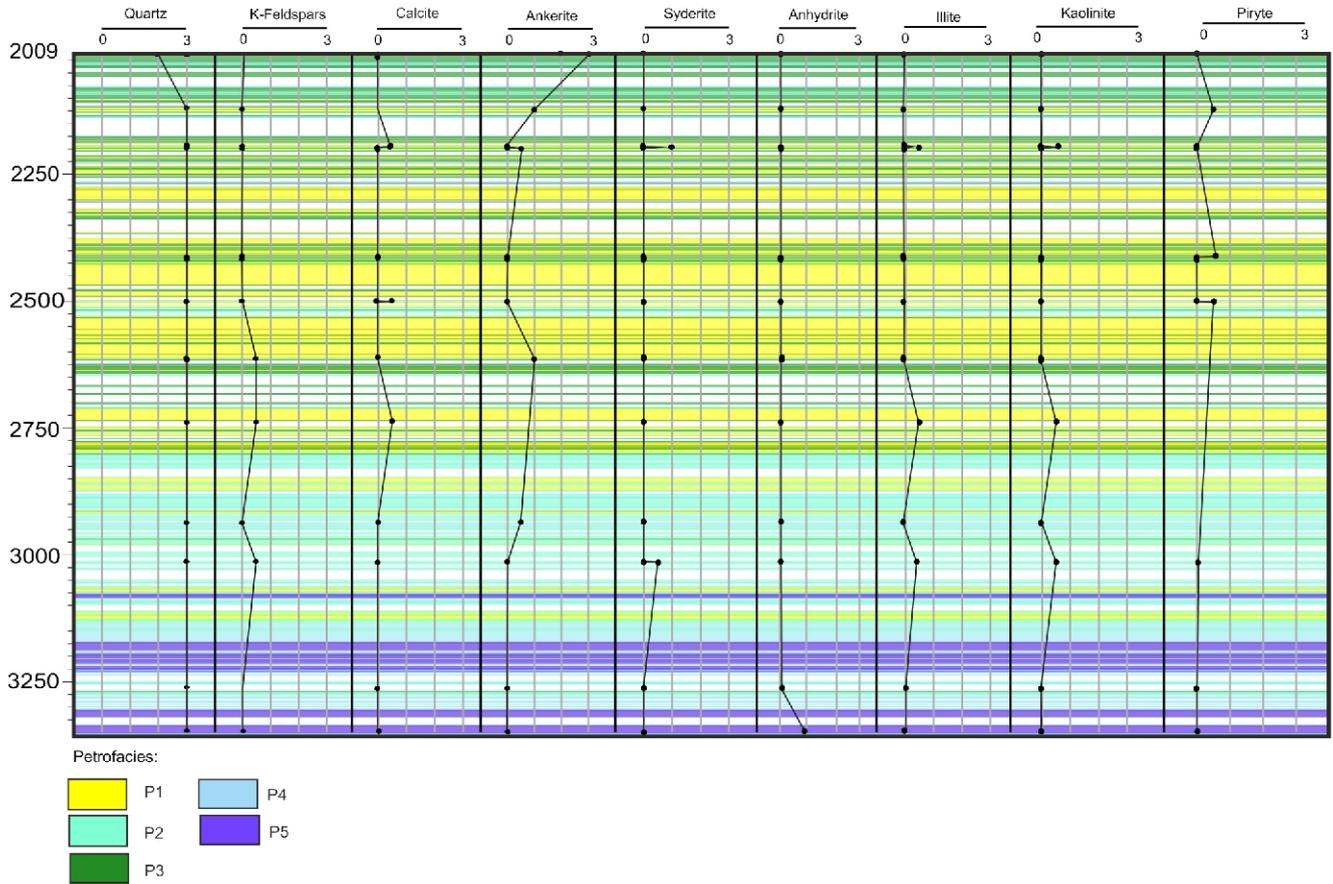


Fig. 17. Archival XRD analysis results

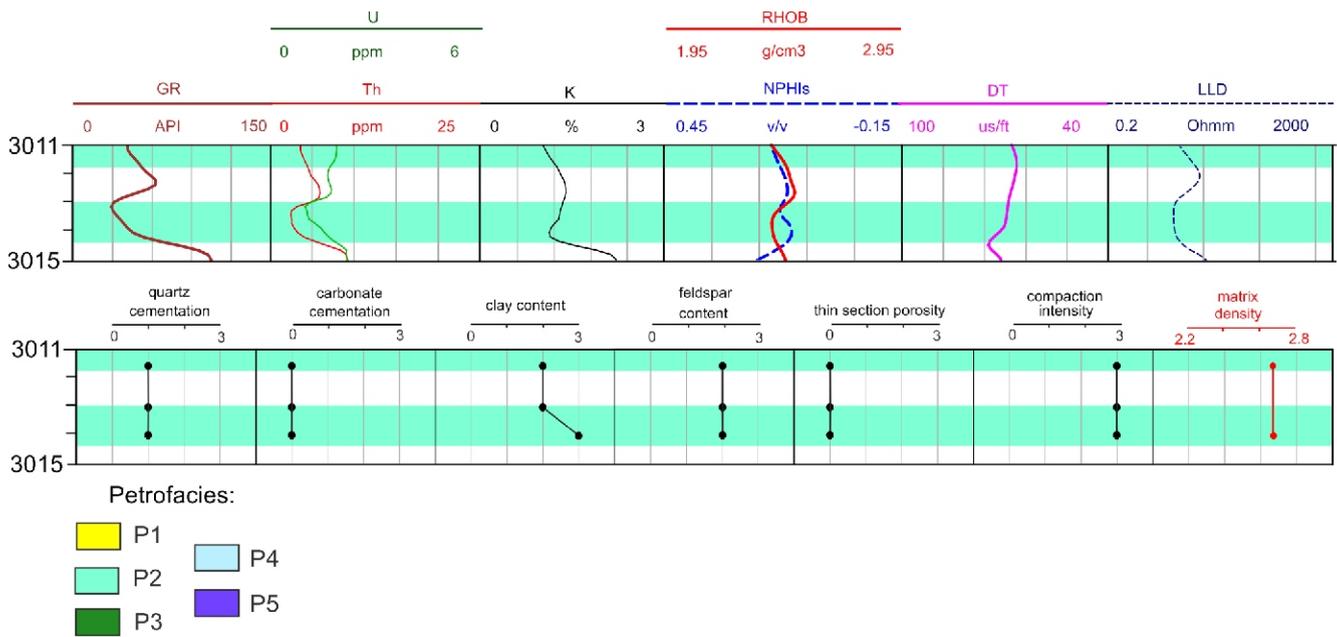


Fig. 18. Log responses compared with the petrographic analysis results in the 3011.0–3015.0 (MD) interval

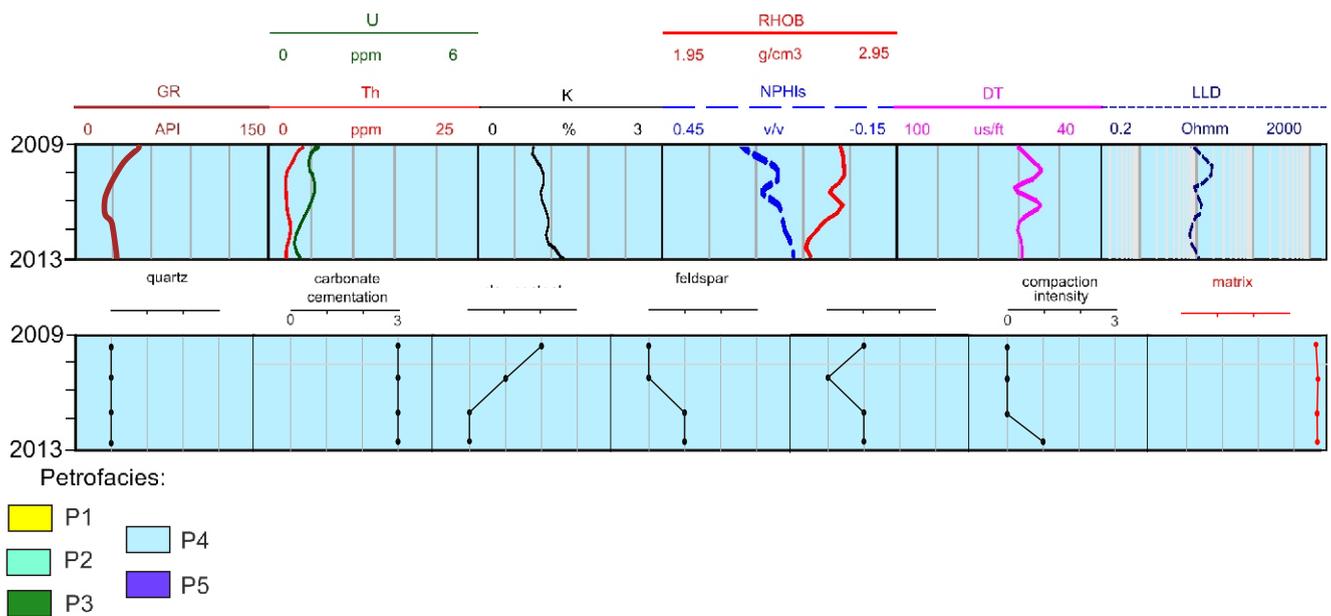


Fig. 19. Log responses compared with petrographic analysis results in the 2009–2013 m b.g.l. interval

minerals), S2 (sandstones with clay coatings and/or quartz overgrowths) and for slightly higher potassium values S3 and S4 (subarkosic and sublithic sandstones) petrofacies recognized in the Z-GN4 borehole by Wróblewska (2022).

PETROPHYSICAL PARAMETERS

The type of petrofacies with their primary and secondary features is directly reflected in their petrophysical properties. Among the other Jurassic rocks, the P1–P3 petrofacies are characterized by significant porosity. The best reservoir parameters show petrofacies P1 represented by dominant quartz arenites. Its highest porosity is the effect of preservation of primary pore spaces by slight authigenic quartz cementation in the

form of grain overgrowths. In the P3 petrofacies the significant secondary pore space evolved as a result of strong dissolution of K-feldspars, glauconite and K-rich lithic grains. The elevated volume of diagenetic kaolinite seems to corroborate this process. The P3 petrofacies is more radiogenic than the P2 petrofacies because of a higher admixture of residual radiogenic grains, mainly K-feldspars and muscovite, which probably have not yet been fully dissolved. The occurrence of a higher admixture of rigid radiogenic grains is probably the main reason for the higher porosity in comparison to the P2 sandstones, which was preserved by a lower influence of compaction. The worst petrophysical properties among the most porous petrofacies are observed in the P2 sandstones, probably due to the highest admixture of clay minerals (especially non-radio-

genic kaolinite and illitized grains) or slight carbonate cementation. The poorest reservoir potential is predominantly seen in the strongly altered petrofacies P4 and P5.

In the cemented, clayey sublithic, and subarkosic arenites of the P4 petrofacies, the reservoir properties are generally unfavourable. Only in some intervals does porosity slightly increase due to calcareous cement or intraclast dissolution. The generally poor reservoir parameters result from low primary porosity, which is additionally partly closed within dissolved grains or by diagenetic carbonate, clay and/or quartz cementation. Moreover, in some cases, compaction can further reduce porosity. Where the dissolution of carbonate cements is more advanced, porosity may rise, by up to 7%.

The P5 sandstones have the lowest porosity due to strong diagenetic cementation of quartz, carbonate and/or anhydrite. The effect of the abundance of high-density cement is visible on the RHOB log. There is also a visible increase in deep resistivity and a decrease in the interval transit time on the DT log (Figs. 15 and 16). In some parts, the lack of porosity results from strong compaction, especially in the bottom part of the section.

This analysis of diagenetic parameters, with particular emphasis on cementation processes and the type of clay minerals present in the rocks, allows for a more precise adjustment of petrophysical modeling criteria. Particularly important is the selection of appropriate parameters for the calculation of shale volume, later used for the proper modeling of effective porosity and permeability (Serra, 1984; Rider, 2002; Asquith and Krygowski, 2004; Wróblewska, 2022). Illite, the main diagenetic clay mineral, leads to the closure of connections between pore spaces in sandy rocks, resulting in reduced effective porosity and permeability. It can be easily detected based on gamma-ray measurements because it is a mineral carrier of radioactive potassium isotopes and is an important indicator of the shaliness of sandy rocks. Some detrital grains, as well as K-feldspars and micas, may also increase the intensity of gamma radiation in sandstones due to their radioactive potassium content, but their presence has no effect on effective porosity. In subarkosic sandstones, the presence of these radioactive elements does not reduce effective porosity and permeability. Moreover, diagenetic kaolinization of K-feldspars often leads to an increase in the porosity of these rocks, as documented during petrofacies analysis. Shale volume can also be calculated based on spontaneous potential measurements, which can distinguish highly cemented or compacted impermeable intervals (Rider, 2002; Adeoti et al., 2009; Szabó, 2011; Willis et al., 2017; Wróblewska, 2022). This measurement is becoming less common in the exploration industry, but the application of the clay difference method in vintage boreholes, together with machine learning results, can significantly improve the reliability of well interpretation and correlation procedures.

#### PETROFACIES DISTRIBUTION

The distribution of specific lithotypes illustrates a relationship representative of the primary and secondary Lower and Middle Jurassic strata in the Polish Lowlands. It is also possible to correlate the features observed in the B-2 core section with those documented by Wróblewska (2022) in the Z-GN4 borehole. Within the Lower Jurassic rocks, a high proportion of quartz in the grain framework of sandy rocks is evident. The dominant rocks belong to petrofacies P1, P2, P4 and P5. The distribution of petrofacies suggests that the textural maturity of the Lower Jurassic sandstones increases upwards in the profile. A gradual transition can be observed from rocks with an in-

creased content of unstable grains, such as feldspar or muscovite (P3, P4), through a series of sandstones with a high amount of diagenetic kaolinite formed by the dissolution of feldspar grains (P2), to the pure quartz sandstones of the P1 petrofacies at the top of the Lower Jurassic profile (Figs. 15 and 17). The increasing proportion of river and deltaic facies towards the top of – at the expense of terrestrial facies dominating at the bottom – seems to corroborate this trend in the Lower Jurassic depositional sequence in the Polish Lowlands. In addition to the original sedimentary features, significant compaction and carbonate/sulphate cementation are also evident, especially at the bottom of the profile, where the P5 petrofacies dominates. Apart from compaction, the anhydrite and/or gypsum cementation probably originates from an underlying Zechstein salt pillow. A high degree of compaction and/or cementation is also observable in thin sections. A similar relationship in textural maturity and compaction was observed by Wróblewska (2022) in the Z-GN4 borehole, although the research conducted there had lower accuracy due to the nature of the archival data. Within the Middle Jurassic rocks, the opposite tendency is observable. The influence of carbonate cementation, and the proportion of granular components other than quartz, increase towards the top of the profile. The P1 petrofacies, dominant at the base of the Middle Jurassic, gradually changes to the P3 petrofacies with a subordinate contribution of P4 (Fig. 15). The decrease in textural maturity and an increasing proportion of carbonate cementation and clay content reflect the rising proportion of shallow-marine facies up the Middle Jurassic profile of the Polish Lowlands, including carbonates providing a source of calcium carbonate in the later diagenesis process. The trend in the B-2 profile is comparable to that in the Z-GN4 Middle Jurassic succession, where the content of feldspars, lithic clasts and calcareous cement visibly increases towards the top of the section (Wróblewska, 2022).

#### CONCLUSIONS

The k-means clustering method, even though with a relatively poor silhouette score and lack of evident cluster separation, is a fast, valuable technique that objectively distinguishes sandy petrofacies based on their geophysical parameters, reflecting both primary and secondary features. Applying this method facilitates the proper categorization of various lithotypes, particularly in cases of petrographically diverse sandstone sequences. The effectiveness of this method has been tested in this study on petrographically and diagenetically varied Lower and Middle Jurassic sandstones in the Polish Basin. The results of this analysis can be applied across various fields, including local and regional palaeogeographic studies, detailed sedimentological analyses, as well as reservoir formation evaluation and modeling. Based on the chosen calculation criteria estimated for particular petrofacies, the k-means clustering method may be used for proper shale volume and porosity modeling in complex, mixed-layer reservoirs (Serra, 1984; Rider, 2002; Asquith and Krygowski, 2004). Furthermore, this method has potential for broader basin analysis and could prove valuable in identifying reservoir and seal rocks within petroleum systems.

In the Polish Basin, the porous and permeable Jurassic clastic rocks have significant potential as natural hydrocarbon reservoirs, especially if they are sealed by cemented and/or compacted rocks. However, in the absence of exploration prospects, these deep, isolated Jurassic aquifers could also serve as candidates for geothermal exploration or for carbon capture

and storage. Consequently, the application of this rapid detection method may find widespread use in the future, not only within the exploration industry but also as part of the global energy transition process.

**Acknowledgements.** The authors would like to express sincere thanks to the Polish Oil and Gas Company PGNiG Orlen S.A. for providing the geological data and to the anonymous reviewers for their valuable remarks and comments,

which significantly increased the quality of the methodological process. The realization of this paper was possible thanks to the cooperation with the Polish Geological Institute – National Research Institute, where part of the research work was carried out under research grant no. 61.3203.2200.00.0, headed by Sara Wróblewska.

## REFERENCES

- Abdi, H., Williams, L.J., 2010.** Principal component analysis. *WIREs Computational Statistics*, **2**: 33–459.
- Adeoti, L., Ayolabi, E.A., James, P.L., 2009.** An integrated approach to volume of shale analysis: Niger Delta example, Offshore Field. *World Applied Sciences Journal*, **7**: 448–452.
- Ali, N., Chen, J., Fu, X., Hussain, W., Ali, M., Iqbal, S.M., Anees, A., Hussain, M., Rashid, M., Thanh, H.V., 2023.** Classification of reservoir quality using unsupervised machine learning and cluster analysis: Example from Kadanwari gas field, SE Pakistan. *Geosystems and Geoenvironment*, **2**, 100123; <https://doi.org/10.1016/j.geogeo.2022.100123>
- Andrew, A.L., 1973.** Eigenvectors of certain matrices. *Linear Algebra and its Applications*, **7**: 151–162.
- Asquith, G., Krygowski, D., 2004.** *Basic Well Log Analysis*, 2nd edn. The American Association of Petroleum Geologists, Tulsa, Oklahoma.
- Bhattacharya, S., 2022.** Unsupervised time series clustering, class-based ensemble machine learning, and petrophysical modeling for predicting shear sonic wave slowness in heterogeneous rocks. *Geophysics*, **87**: D161–D174; <https://doi.org/10.1190/geo2021-0478.1>
- Chang, J., Li, J., Kang, Y., Lv, W., Xu, T., Li, Z., Xing Zheng, W., Han, H., Liu, H., 2021.** Unsupervised domain adaptation using maximum mean discrepancy optimization for lithology identification. *Geophysics*, **86**: ID19–ID30; <https://doi.org/10.1190/geo2020-0391.1>
- Churchman, G.J., Lowe, D.J., 2012.** Alteration, formation, and occurrence of minerals in soils. In: *Handbook of Soil Sciences*. 2nd edition., **1**: Properties and Processes (eds. P.M. Huang, Y. Li and M.E. Sumner): 1–72. CRC Press (Taylor & Francis), Boca Raton, FL.
- Cui, Y., Wang, G., Jones, S.J., Zhou, Z., Ran, Y., Lai, J., Li, R., Deng, L., 2017.** Prediction of diagenetic facies using well logs – A case study from the upper Triassic Yanchang Formation, Ordos Basin, China. *Marine and Petroleum Geology*, **81**: 50–65; <https://doi.org/10.2516/ogst/2014060>
- Dadlez, R., 1997.** Epicontinental basins in Poland: Devonian to Cretaceous – relationships between the crystalline basement and sedimentary infill. *Geological Quarterly*, **41** (4): 419–432.
- Dadlez, R., 1998.** Devonian to Cretaceous epicontinental basins in Poland: relationship between their development and structure (in Polish with English summary). *Prace Państwowego Instytutu Geologicznego*, **165**: 17–30.
- Dadlez, R., Marek, S., 1969.** Structural style of the Zechstein-Mesozoic complex in some areas of the Polish Lowland (in Polish with English summary). *Geological Quarterly*, **13** (3): 543–565.
- Dadlez, R., Marek, S., Pokorski, J., 1998.** Palaeogeographical Atlas of the Epicontinental Permian and Mesozoic in Poland. Państwowy Instytut Geologiczny, Warszawa.
- Dadlez, R., Narkiewicz, M., Stephenson, R.A., Visser, M.T.M., Van Wees, J.D., 1995.** Tectonic evolution of the Mid-Polish Trough: modelling implications and significance for central European geology. *Tectonophysics*, **252**: 179–195; [https://doi.org/10.1016/0040-1951\(95\)00104-2](https://doi.org/10.1016/0040-1951(95)00104-2)
- Dayczak-Calikowska, K., 1964.** Atlas geologiczny Polski – Zagadnienia stratygraficzno-facjalne, z. 9 Jura, cz. II Jura środkowa (in Polish). Wyd. Geol., Warszawa.
- Dayczak-Calikowska, K., 1997.** Middle Jurassic. Sedimentation, paleogeography and paleotectonics (in Polish with English summary). *Prace Państwowego Instytutu Geologicznego*, **153**: 269–282.
- Dayczak-Calikowska, K., Moryc, W., 1988.** Evolution of sedimentary basin and palaeotectonics of the Middle Jurassic in Poland (in Polish with English summary). *Geological Quarterly*, **32** (1): 117–136.
- De Maesschalck, R., Jouan-Rimbaud, D., Massart, D.L., 2000.** The Mahalanobis distance. *Chemometrics and Intelligent Laboratory Systems*, **50**: 1–18; [https://doi.org/10.1016/S0169-7439\(99\)00047-7](https://doi.org/10.1016/S0169-7439(99)00047-7)
- De Ros, L., Fernando, Goldberg, K., 2007.** Reservoir petrofacies: a tool for quality characterization and prediction. AAPG Annual Convention and Exhibition, Long Beach, CA.
- Feldman-Olszewska, A., 1997a.** Depositional systems and cyclicity in the intracratonic Early Jurassic basin in Poland. *Geological Quarterly*, **41** (4): 475–490.
- Feldman-Olszewska, A., 1997b.** Depositional architecture of the Polish epicontinental Middle Jurassic basin. *Geological Quarterly*, **41** (4): 491–508.
- Feldman-Olszewska, A., 1998.** Lower and Middle Jurassic. In: *Palaeogeographical Atlas of the Epicontinental Permian and Mesozoic in Poland* (eds. R. Dadlez, S. Marek and J. Pokorski). Państwowy Instytut Geologiczny, Warszawa.
- Feldman-Olszewska, A., 2005.** Sedimentary environments of the Middle Jurassic epicontinental deposits from the central part of the Polish Basin (Kuyavian Region). *Volumina Jurassica*, **3**: 130–131.
- Ghorbani, H., 2019.** Mahalanobis distance and its application for detecting multivariate outliers. *Facta Universitatis, Series: Mathematics and Informatics*, **34**: 583–595; <https://doi.org/10.22190/FUMI1903583G>
- Gutowski, J., Krzywiec, P., Walaszczyk, I., Pożaryski, W., 2003.** Od ekstensji do inwersji – zapis aktywności północno-wschodniej brzeżnej strefy uskokowej świętokrzyskiego segmentu bruzdy wśród polskiej w osadach jury górnej i kredy na podstawie interpretacji danych sejsmiki refleksyjnej (in Polish). *Volumina Jurassica*, **1**: 124–125.
- Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M.H., Brett, M., Haldane, A., del Rio, J.F., Wiebe, M., Peterson, P., Gerard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., Oliphant, T.E., 2020.** Array programming with NumPy. *Nature*, **585**: 357–362; <https://doi.org/10.1038/s41586-020-2649-2>
- Hartigan, J.A., Wong, M.A., 1979.** A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **28**: 100–108.

- Hussein, M., Stewart, R.R., Sacrey, D., Wu, J., Athale, R., 2021. Unsupervised machine learning using 3D seismic data applied to reservoir evaluation and rock type identification. *Interpretation*, **9**: T549–T568; <https://doi.org/10.1190/INT-2020-0108.1>
- Ingersoll, R.V., 1990. Actualistic sandstone petrofacies: discriminating modern and ancient source rocks. *Geology*, **18**: 733–736; [https://doi.org/10.1130/0091-7613\(1990\)018<0733:ASPDMA>2.3.CO;2](https://doi.org/10.1130/0091-7613(1990)018<0733:ASPDMA>2.3.CO;2)
- Ippolito, M., Ferguson, J., Jenson, F., 2021. Improving facies prediction by combining supervised and unsupervised learning methods. *Journal of Petroleum Science and Engineering*, **200**, 108300; <https://doi.org/10.1016/j.petrol.2020.108300>
- Jolliffe, I., 2005. Principal Component Analysis. In: *Encyclopedia of Statistics in Behavioral Science* (eds. B.S. Everitt and D. Howell). John Wiley & Sons, Ltd., New York.
- Jolliffe, I.T., Cadima, J., 2016. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **374**, 20150202; <https://doi.org/10.1098/rsta.2015.0202>
- Karelus, U., Modzelewski, R., 1996. Dokumentacja wyników odwiertu poszukiwawczego Bielsk 2 (in Polish). *Narodowe Archiwum Geologiczne* (nr 133804), Warszawa.
- Karnkowski, P.H., 1999. Origin and evolution of the Polish Rotliegendes basin. *Polish Geological Institute Special Papers*, **3**: 1–93.
- Kopik, J., 1998. Lower and Middle Jurassic of the north-eastern margin of the Upper Silesian Coal Basin (in Polish with English summary). *Biuletyn Państwowego Instytutu Geologicznego*, **378**: 67–129.
- Kotański, Z., 1997. Geological Atlas of Poland: Geological Maps of Horizontal Cutting 1:750 000. Państwowy Instytut Geologiczny, Warszawa.
- Kozłowska, A., Kuberska, M., 2014. Diagenesis and porosity of the Lower Jurassic sandstones in the Polish Lowlands (in Polish with English summary). *Biuletyn Państwowego Instytutu Geologicznego*, **458**: 39–60.
- Kozłowska, A., Maliszewska, A., 2015. Berthierine in the Middle Jurassic sideritic rocks from southern Poland. *Geological Quarterly*, **59** (4): 551–564; <https://doi.org/10.7306/gq.1231>
- Krystkiewicz, E., 1999. Lower Jurassic (in Polish with English summary). *Prace Państwowego Instytutu Geologicznego*, **167**: 64–77.
- Krzywiec, P., 2006. Structural inversion of the Pomeranian and Kuivian segments of the Mid-Polish Trough – lateral variations in timing and structural style. *Geological Quarterly*, **50** (1): 151–168.
- Krzywiec, P., Kramarska, R., Zientara, P., 2003. Strike-slip tectonics within the SW Baltic Sea and its relationship to the inversion of the Mid-Polish Trough – evidence from high-resolution seismic data. *Tectonophysics*, **373**: 93–105; [https://doi.org/10.1016/S0040-1951\(03\)00286-5](https://doi.org/10.1016/S0040-1951(03)00286-5)
- Leonowicz, P., 2005. The Ciechocinek Formation (Lower Jurassic) of SW Poland: petrology of green clastic rocks. *Geological Quarterly*, **49** (3): 317–330.
- Lewandowski, M., Krobicki, M., Matyja, B.A., Wierzbowski, A., 2005. Palaeogeographic evolution of the Pieniny Klippen Basin using stratigraphic and palaeomagnetic data from the Veliky Kamenets section (Carpathians, Ukraine). *Palaeogeography, Palaeoclimatology, Palaeoecology*, **216**: 53–72; <https://doi.org/10.1016/j.palaeo.2004.10.002>
- Likas, A., Vlassis, N., J. Verbeek, J., 2003. The global k-means clustering algorithm. *Pattern Recognition*, **36**: 451–461.
- Lis-Śledziona, A., 2019. Petrophysical rock typing and permeability prediction in tight sandstone reservoir. *Acta Geophysica*, **67**: 1895–1911; <https://doi.org/10.1007/s11600-019-00348-5>
- Loutit, T.S., Hardenbol, J., Vail, P.R., Baum, G.R., 1988. Condensed sections: The key to age determination and correlation of continental margin sequences. *SEPM Special Publication*, **42**: 39–45; <https://doi.org/10.2110/pec.88.01.0183>
- Maliszewska, A., 1999. Middle Jurassic (in Polish with English summary). *Prace Państwowego Instytutu Geologicznego*, **167**: 78–93.
- Maliszewska, A., 1968. The mineralization of Dogger deposits based on the example of Głogowiec borehole (in Polish with English summary). *Geological Quarterly*, **12** (1): 105–116.
- Marek, S., Pajchłowa, M., eds., 1997. The epicontinental Permian and Mesozoic in Poland (in Polish with English summary). *Prace Państwowego Instytutu Geologicznego*, **153**.
- Matyja, B.A., Wierzbowski, A., 2006. The oceanic “Metis Geotectonic Event” (Callovian/ Oxfordian) and its implications for the Peri-Tethyan area of Poland. *Volumina Jurassica*, **4**: 60–61.
- McKinney, W., 2010. Data Structures for Statistical Computing in Python. *Python in Science Conference*, 56–61; <https://doi.org/10.25080/Majora-92bf1922-00a>
- Modliński, Z., ed., 1974. Bartoszyce IG-1, Gołdap IG-1 (in Polish). *Profile głębokich otworów wiertniczych Instytutu Geologicznego*, **14**.
- Mohammadian, E., Kheirollahi, M., Liu, B., Ostadhassan, M., Sabet, M., 2022. A case study of petrophysical rock typing and permeability prediction using machine learning in a heterogeneous carbonate reservoir in Iran. *Scientific Reports*, **12**, 4505; <https://doi.org/10.1038/s41598-022-08575-5>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: machine learning in Python. *Journal of Machine Learning Research*, **12**: 2825–2830.
- Pettijohn, F.J., Potter, P.E., Siever, R., 2012. *Sand and Sandstone*. Springer Science & Business Media, New York.
- Pieńkowski, G., 2004. The epicontinental Lower Jurassic in Poland. *Polish Geological Institute Special Papers*, **12**: 5–154.
- Premik, J., 1933. *Budowa i dzieje geologiczne okolic Częstochowy* (in Polish). Drukarnia Łazarskiego, Warszawa.
- Rider, M.H., 2002. *The Geological Interpretation of Well Logs*, 2nd edn. Rider-French Consulting, Sutherland.
- Ringné, M., 2008. What is principal component analysis? *Nature Biotechnology*, **26**: 303–304.
- Rousseeuw, P.J., 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, **20**: 53–65.
- Serra, O., 1984. *Fundamentals of Well-log Interpretation: The interpretation of logging data*. Elsevier, Amsterdam.
- Sinaga, K.P., Yang, M.S., 2020. Unsupervised k-means clustering algorithm. *IEEE Access*, **8**: 80716–80727; <https://doi.org/10.1109/ACCESS.2020.2988796>
- Singh, A., Ojha, M., 2022. Machine learning in the classification of lithology using downhole NMR data of the NGHP-02 expedition in the Krishna-Godavari offshore Basin, India. *Marine and Petroleum Geology*, **135**: 105443; <https://doi.org/10.1016/j.marpetgeo.2021.105443>
- Steinley, D., 2006. K-means clustering: a half-century synthesis. *The British Journal of Mathematical and Statistical Psychology*, **59**: 1–34.
- Stephenson, R.A., Narkiewicz, M., Dadlez, R., Van Wees, J.D., Andriessen, P., 2003. Tectonic subsidence modelling of the Polish Basin in the light of new data on crustal structure and magnitude of inversion. *Sedimentary Geology*, **156**: 59–70; [https://doi.org/10.1016/S0037-0738\(02\)00282-8](https://doi.org/10.1016/S0037-0738(02)00282-8)
- Szabó, N.P., 2011. Shale volume estimation based on the factor analysis of well-logging data. *Acta Geophysica*, **59**: 935–953.
- Teofilak, A., 1960. *Petrografia liasu i doggeru w wierceniu Gorzów Wielkopolski IG-1* (in Polish). *Narodowe Archiwum Geologiczne* (nr 54429), Warszawa.
- Teofilak, A., 1962. *Jura dolna – Charakterystyka petrograficzna głównych typów skał; Dogger – Charakterystyka petrograficzna głównych typów skał* (in Polish). In: *Budowa geologiczna niżu polskiego* (ed. M. Pożaryski). Wyd. Geol., Warszawa.

- Vail, P.R., Hardenbol, J., Todd, R.G., 1984. Jurassic unconformities, chronostratigraphy, and sea-level changes from seismic stratigraphy and biostratigraphy. *AAPG Memoir*, **36**: 129–144.
- Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, Í., Feng, Y., Moore, E.W., Vander Plas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors, 2020. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, **17**: 261–272.
- Wang, F., Franco-Penya, H.-H., Kelleher, J.D., Pugh, J., Ross, R., 2017. An analysis of the application of simplified silhouette to the evaluation of k-means clustering validity. In: *Machine Learning and Data Mining in Pattern Recognition* (ed. P. Perner): 291–305. Springer International Publishing, Cham.
- Wang, H., Lu, S., Qiao, L., Chen, F., He, X., Gao, Y., Mei, J., 2022. Unsupervised contrastive learning for few-shot TOC prediction and application. *International Journal of Coal Geology*, **259**, 104046; <https://doi.org/10.1016/j.coal.2022.104046>
- Willis, J.J., McIntosh, D.S., Jr, Zwennes, J.W., Ferguson, G.J., 2017. Quick-look technique for quantifying shale distribution types using total porosity versus shale volume crossplots. *Gulf Coast Association of Geological Societies Transactions*, **67**: 539–549.
- Wróblewska, S.M., 2022. The application of archival borehole data to recognizing the primary mineral composition and diagenetic changes in Jurassic sandy rocks of the Polish Basin. *Geological Quarterly*, **66**, 24; <https://doi.org/10.7306/gq.1656>
- Wróblewska, S.M., Kozłowska, M., 2019. Influence of primary composition and diagenetic mineralization on natural gamma-ray and density logs of Carboniferous sandstones of the Dęblin Formation. *Geological Quarterly*, **63** (4): 741–756; <https://doi.org/10.7306/gq.1498>
- Zheng, A., Casari, A., 2018. *Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists*. O'Reilly Media, Inc.
- Ziegler, P.A., 1990. *Geological Atlas of Western and Central Europe*. Shell Internationale Petroleum Maatschappij B.V.
- Znosko, J., 1957. Upward movements of Kłodawa salt diapir and its influence on genesis of sideritic coquinaes (in Polish with English summary). *Geological Quarterly*, **1** (1): 90–104.
- Znosko, J., 1968. Transgressive oscillations of the Dogger sea between Gorzów Wielkopolski and Zakrzewo (in Polish with English summary). *Geological Quarterly*, **12** (2): 308–315.
- Żelaźniewicz, A., Aleksandrowski, P., A., Buła, Z., Karnkowski, P., Konon, A., Słaczka, A., Żaba, J., Żytko, K., 2011. *Polska bez pokrywy kenozoicznej* (in Polish). In: *Regionalizacja tektoniczna Polski*: 25–28. Komitet Nauk Geologicznych PAN, Wrocław.