ORIGINAL STUDY

Using geophysical log data to predict the fracture density in a claystone host rock for storing high-level nuclear waste

Emese Tóth[1] · Ervin Hrabovszki[1] · Tivadar M. Tóth[1]

Abstract

Previously drilled boreholes of a host rock for a potential nuclear waste repository in Hungary revealed a highly fractured claystone rock body. A crucial step for characterizing the hydrodynamic behavior of such a fractured reservoir is fracture identification and accurate calculation of the fracture density. Although acoustic borehole televiewers provide a reliable base for determining the fracture density, older boreholes usually lack such data. However, conventional borehole geophysical measurements are often accessible in such cases. The aim of this study was to identify any correlations between well log data and fracture density. Multiple linear regression analysis was performed on data from two boreholes penetrating the Boda Claystone Formation in southwest Hungary. The upper section of the BAF-4 borehole was used for training, where the fracture density was estimated with a fit of $R^2 = 0.767$. The computed regression function predicted the fracture density with high accuracy in both boreholes for all intervals with typical lithological features. However, in some sections where anomalous well log data indicated changes in the lithology, the prediction accuracy decreased. For example, the function underestimated the fracture density in sandy intervals.

Article highlights

- Fracture density of a potential nuclear waste repository predicted by using regression analysis on geophysical logs.
- Fracture density of a claystone body influenced by resistivity and density.
- Prediction accuracy may be influenced by grain size, bedding type, and presence of reductive layers.

✉ Emese Tóth
  tothemese@geo.u-szeged.hu

1   Department of Mineralogy, Geochemistry and Petrology, University of Szeged, Szeged 6722, Hungary

1 Introduction

A host rock of a waste repository that contains fractures may be considered a fractured reservoir. The viability of a fractured reservoir as a waste repository depends on how fluid and contaminants can be retained in the formation. Claystone formations are often targeted as potential repository sites, along with salt, tuff, and granitic rock formations (Ahn and Apted 2010). Claystone generally has very low matrix porosity and permeability, which allows fluid to move exclusively along fracture planes (Anders et al. 2014). Therefore, the fracture network geometry, fracture aperture, and fracture density are among the key parameters that affect the fluid flow in fractured reservoirs. The detection of fractures and precise estimation of the fracture density are important for characterizing the hydrodynamic behavior of a fractured reservoir. Such information can reflect fracture development and connectivity, which are useful for optimizing waste disposal (Nelson 2001; Ge et al. 2020).

Fractures in boreholes can be detected both directly and indirectly. Direct methods include core analysis, impression packers, and borehole cameras. Indirect methods include packer borehole tests and well log data (Ja'fari et al. 2012). The most accurate method is through direct observation and core analysis, but this approach is costly and time-consuming, and cores are frequently unavailable. In addition, highly fractured zones are often lost during core recovery, and mechanical fractures are frequently induced, which significantly alters the fracture density of multiple sections (Laubach et al. 1988). Human error and non-constant resolution limits may also affect the observed fracture density in direct observation of cores. Downhole cameras are small devices that can capture images and video inside a borehole. These tools collect high-resolution data and offer a quick and efficient approach to detecting discontinuities. Unfortunately, image logs are not available for hundreds of boreholes drilled before such technology was introduced in the 1980s (Tokhmchi et al. 2010; Ja'fari et al. 2012). One of the most popular methods for localizing fractures along boreholes is the acoustic borehole televiewer (BHTV), which is rapid, cost-effective, and accurate (MartinezTorres 2002). A BHTV provides a continuous and oriented image of the borehole wall which reflects the orientation, dip angle and thickness of fractures as a function of depth. BHTV logs are beneficial for evaluating boreholes without oriented cores (Massiot et al. 2017). Images are produced based on the travel time of the ultrasonic signal reflected from the borehole wall in this technique (Zemanek et al. 1970; Pöppelreiter et al. 2010). Potential fluid routes may develop along any discontinuity in a formation. Thus, every single discontinuity is worth defining as a fracture when evaluating a fracture network.

Discrete fracture network (DFN) modeling is a valuable technique for determining the fracture network geometry in a fractured reservoir if the geometric characteristics of individual fractures are known (Witherspoon et al. 1980; Neuzil and Tracy 1981). By modeling the fracture networks of numerous boreholes, it becomes possible to analyze the hydrogeological behavior over a wider area and the impact of various processes on the fracture network characteristics. One of the most critical parameters for correlating boreholes is the fracture density. Although BHTV datasets provide a firm basis for computing fracture density logs, this information is frequently unavailable for older boreholes. In such cases, however, standard geophysical well logs are typically available. Incorporating such old boreholes in spatial correlation studies can help improve understanding of fractured rock bodies, the locations and sizes of communicating fracture clusters, and the positions of significant structural boundaries.

The fracture density is a significant characteristic that is the primary basis for correlating data from boreholes in the Boda Claystone Formation (BCF). P10 is the most widely used parameter for quantifying the fracture density, and it is defined as the number of fractures per meter, which can be derived from BHTV data for any borehole segment. Downhole geophysical logging can be used to examine the petrophysical characteristics of a rock formation. Logging equipment responds to various features in the borehole environment (Shalaby and 98 Islam 2017). Even for older boreholes, conventional well logs frequently include geophysical data such as natural gamma rays, spontaneous potential, sonic transit time, bulk density, neutron porosity, caliper data, and resistivity. In contrast, unconventional well logs such as the Formation MicroScanner (FMS) and Formation MicroImager (FMI) are too specialized, expensive, or recently invented to be utilized in every borehole (Gamal et al. 2022).

Several earlier studies have attempted to estimate the fracture density by using conventional well logs. Most of these studies were related to hydrocarbon production, so they typically targeted formations such as fractured carbonate and sandstone reservoirs. Tokhmchi et al. (2010) used well log data (caliper data, density, neutron porosity, sonic transit time, resistivity, natural gamma rays) to calculate the fracture density in naturally fractured reservoirs. Ja'fari et al. (2012) used an adaptive neuro-fuzzy inference method on well log data (sonic transit time, deep resistivity, neutron porosity, bulk density) to calculate the fracture density. Zazoun (2013) used artificial neural networks (ANNs) on standard well log data (gamma rays, sonic transit time, caliper data, neutron porosity, bulk density) as well as core data to predict the fracture density. Aghli et al. (2016) used well log data (bulk density, neutron porosity, gamma rays, sonic transit time, photoelectric absorption, caliper data) to determine fracture zones in a fractured carbonate host rock. Taherdangkoo and Abdideh (2016) used the wavelet transformation approach on conventional well log data (gamma rays, bulk density, sonic transit time, photoelectric absorption) to estimate the locations of fractured zones and calculate the number of fractures in each zone of a reservoir. Dong et al. (2020) used a semi-supervised learning system on conventional well log data (gamma rays, caliper data, spontaneous potential, neutron porosity, acoustic impedance, density, resistivity) to predict fracture zones in tight sandstone. Owing to a lack of directly obtained advanced logging data, Gamal et al. (2022) integrated conventional well logs, thin sections, and other available data to detect fractures in a carbonate rock body. However, no previous study has attempted to use conventional well log data to estimate the fracture density in a claystone formation.

The aim of this study was to quantify how different geophysical features describe the fracture density so that the fracture density of boreholes can be determined without BHTV data. The relationship between geophysical characteristics recorded in well logs and the fracture density of a claystone formation was explored by linear regression analysis. The developed approach was tested on two boreholes of the Boda Claystone Formation, which is a potential site for a high-level nuclear waste depository in Hungary.

2 Methods

Multiple linear regression (MLR) analysis is a statistical technique to model the linear relationship between a dependent variable and a set of independent variables.The dependent variable is referred to as the predictand or response, while the independent variables are

referred to as predictors. MLR model establishes a simultaneous statistical relationship between the single continuous outcome y and the predictor variables x_k ($k = 1, 2, …, p-1$):

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + … + \beta_{p-1} x_{i,p-1} + \epsilon_i \tag{1}$$

where β_0 represents the intercept, also called the constant (the mean of Y when all $X_k = 0$), and each β_k represents a slope with respect to X_k, ε_i is the ith error. The β_k are called partial regression coefficients (Eberly 2007; Olive 2017; Tranmer et al. 2020).

The multiple regression analysis procedure consists of the following steps (Tranmer et al. 2020). In order to build a multiple linear regression model, additional assumptions must be verified. It is important to confirm that a linear relationship is likely to exist for each predictor. If the relationship is not linear, a transformation of the variables is required. The normality assumption assumes that the residuals have a normal distribution with a mean of zero.

The predictors should be independent of each other. High correlations between two or more independent variables are a sign of multicolinearity, which creates redundant information that distorts the results of a regression model. The model can be simplified by removing the highly correlated variables. The variance inflation factor (VIF) can show how multicollinearity has increased the variance of the regression estimates. Multicollinearity can be a problem if the VIF is higher than 10 (Alin 2010):

$$VIF_i = \frac{1}{1 - R_i^2} \tag{2}$$

Where VIF_i is the variance inflation factor of the ith predictor and R^2_i is the multiple coefficient of determination in a regression of the ith predictor on all other predictors (Alin 2010).

The Durbin-Watson (DW) statistic tests for autocorrelation in the residuals of the regression analysis and determines whether there is a significant correlation based on the order in which they appear in the data file:

$$d = \frac{\sum_{i=1}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2} \tag{3}$$

where n is the number of observation, $\varepsilon_i = y_i - \bar{y}_i$ (y_i is observed values, $\bar{y}_i$ is the predicted values). DW value is always between 0 and 4, if there is no autocorrelation the DW value should be between 1.5 and 2.5 (Vinod 1973).

Subsequently, the estimation of the linear regression model and the diagnostic evaluation is performed. The model should then be reduced by removing non-significant predictors after making inferences about the regression coefficient. Finally, the model performance (whether the model provides an adequate fit to the data) has to be checked.

This study used multiple regression analysis with the backward elimination method. Backward stepwise regression is a stepwise regression method that starts with a full model and gradually removes variables to obtain a reduced model that best fits the data. The process of backward elimination is terminated once all remaining variables meet the requirements to remain in the model (Olive 2017). The dependent variable was the fracture density measured as P10. The set of independent variables was the geophysical well log data.

Fig. 1 (a) Location of the Mecsek Mountains in Hungary. **(b)** geological map of the Mecsek Mountains showing the distribution of the Boda Claystone Formation (BCF) modified after Konrád and Sebe (2010). Legend: (1) Neogene sediments, (2) Jurassic and Cretaceous sediments and Cretaceous volcanic rocks, (3) Triassic sediments (sandstones, carbonates, and evaporites), (4) Upper Permian–Triassic Kővágószőlős Sandstone Formation, (5) Upper Permian BCF, (6) Palaeozoic, (7) fault, (8) strike-slip fault, (9) thrust fault, (10) syncline and anticline, (11) borehole sites
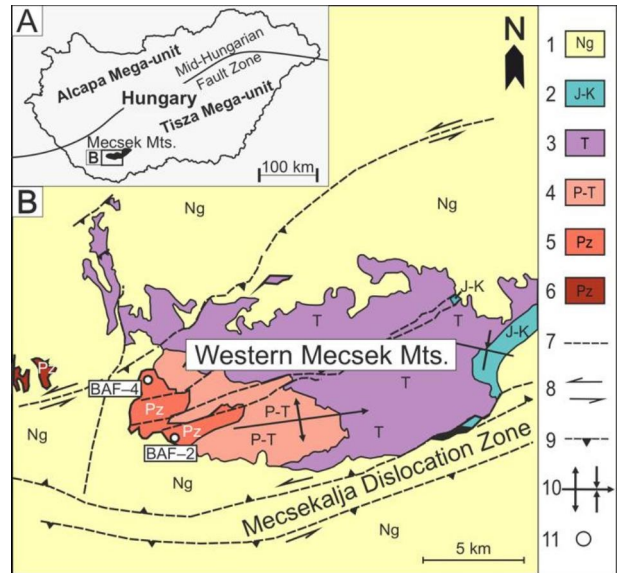
Different goodness-of-fit measures can be used for linear regression models. The R^2 value ranges from 0 to 1 and provides a relative measure of the portion of the variation in the dependent variable that is explained by the model (Bowerman et al. 2005). The standard error of estimate (S_e) measures how far away data points typically are from the regression line. Mean absolute error (MAE) indicates the average value of residuals.

3 Geological setting

Several 1000-m-deep boreholes with nearly 100% core recovery were previously drilled in the BCF to investigate its potential as a final storage location for high-level radioactive waste in Hungary (Konrád and Hámos 2006). These boreholes offer a rare opportunity to study and model the fracture network of a claystone formation as precisely as possible. The BCF may be appropriate for storing nuclear waste in Hungary because of its thickness, low permeability and porosity. The average thickness of the BCF is 900 m and is a member of the 4000–5000 m thick Palaeozoic–Triassic sedimentary sequence of the Western Mecsek Mountains in southwest Hungary (Fig. 1). This Late Permian formation primarily comprises well-compacted reddish-brown claystone and siltstone with layers of fine sandstone and dolomite. The rock-forming minerals are quartz, albite, illite-muscovite, chlorite, calcite, dolomite, and hematite (Árkai et al. 2000). The BCF was deposited in a playa mudflat in an alkaline lake environment during the Late Permian (Árkai et al. 2000; Varga et al. 2005, 2006; Varga 2009; Konrád et al. 2010).

Several successive tectonic stages characterize the structural evolution of the region. Significant NE–SW shortening occurred in the Late Cretaceous (Benkovics et al. 1997) followed by Neogene events related to the formation of the Pannonian Basin. Early Miocene deformation caused by tensional forces was followed by Late Miocene (Sarmatian) compression and thermal subsidence of the basin (Bergerat and Csontos 1988; Csontos

and Bergerat 1992; Fodor et al. 1999; Csontos et al. 2002; Maros et al. 2004). The ongoing tectonic inversion of the basin is the most recent significant event (Konrád and Sebe 2010). Previous drillings have revealed a rock body with a significant number of mineral veins and fractures. Examining these structural elements is crucial for evaluating the retention qualities of the claystone. BAF-2 is a borehole that has been extensively investigated, and numerous studies have focused on its veins and fractures (Hrabovszki et al. 2017, 2020, 2022; Tóth et al. 2020, 2022a, b). Four distinct vein generations have been identified: straight veins, veins with wall rock inclusions, breccia-like veins, and sigmoidal shaped en-échelon veins. These are primarily filled by calcite while small amounts of anhydrite, barite, celestine, etc. are also present as vein-filling phases (Hrabovszki et al. 2020, 2022).

DFN modeling and hydrogeological evaluation of the simulated fracture network of BAF-2 revealed three zones whose hydrogeological properties deviate from the average of the claystone body (Tóth et al. 2022a). BAF-2 is more than 900 m deep (Fig. 2), and the first anomalous zone is located in the upper 100 m, where the porosity decreases faster than the permeability. This may be because the BCF was exposed at the surface for a considerable amount of time (Konrád et al. 2015), and weathering changed the hydrogeological characteristics in the uppermost part. Several features of the upper 400-m and bottom 400-m sections of BAF-2 differ considerably. The upper half of BAF-2 has a significantly higher P10 than the lower half, and the geophysical logs also differ (Tóth et al. 2022a). The average P10 is 8.7 m−1 in the upper 400 m and 4.6 m−1 in the lower 400 m (Fig. 3). Considering the tectonic history of the area, the contact between the two regimes can be interpreted as a large-scale structural boundary that is probably a reverse fault zone. The border of the two sections also behaves as an independent hydraulic unit. The third zone, which differs from the average hydrogeological characteristics, is at a depth of about 700 m, and it displays a unique behavior attributed to fine sandstone layers that are common at depths below 758 m or another large-scale tectonic structure could also influence the hydrogeological behavior of this section (Tóth et al. 2022a).

3.1 Fracture system of the BAF-4 well

BAF-4 is a 900-m-deep borehole that was also considered in the present study. Below the BCF, BAF-4 unexpectedly ran into the Gyűrűfű Formation and drilled into it for about 50 m (Fig. 2). The same method presented by Tóth el al. (2022a) was used to characterize the fracture network of BAF-4. The P10 values of the two boreholes, which are approximately 3.5 km apart (Fig. 1), differ significantly. The two key 400-m-thick sections defined in BAF-2 (Tóth et al. 2022a; Fig. 3a) cannot be identified in BAF-4 (Fig. 3b). In BAF-4, P10 increased continuously with increasing depth from 0 to 900 m without any sudden changes. According to BHTV data, the average fracture density is 3.6 fractures per meter.

Fig. 2 Lithologies of BAF–2 and BAF–4. Legend: (1) quaternary sediments; (2) claystone with siltstone beds; (3) claystone with siltstone and sandstone beds; (4) reductive claystone; (5) rhyolite
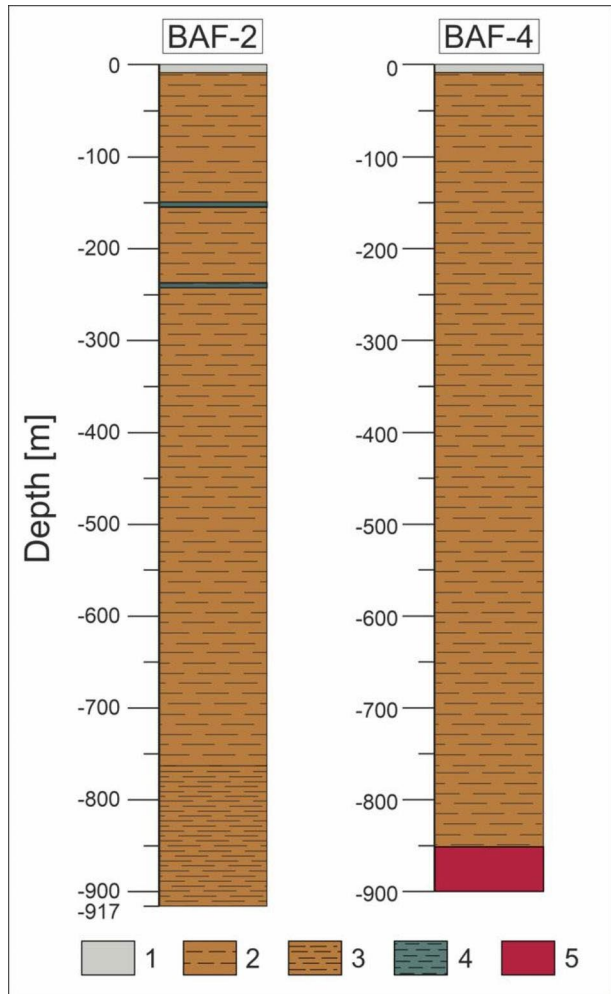
4 Results and discussion

4.1 Multiple regression analysis between the fracture density and the geophysical log data

This study examined the relationship between geophysical log data and fracture density of the rock body by using the software IBM SPSS Statistics, version 28.0. Geophysical well log data from the top section of BAF-4 (220–530 m) were used to train the multiple linear regression model. This section has a lithology typical of the BCF and all of the required geophysical log data. The lower part of BAF-4 (530–840 m) was used to validate the model. Then, the regression function computed in BAF-4 was used to estimate P10 of BAF-2, which was then compared to the measured data. The P10 values for both boreholes were calculated by using BHTV data. The independent variables comprised a standard well log set of the neutron porosity, short-spaced density, natural gamma, and resistivity (e10, e40,

Fig. 3 Fracture density in (**a**) BAF–2 and (**b**) BAF–4
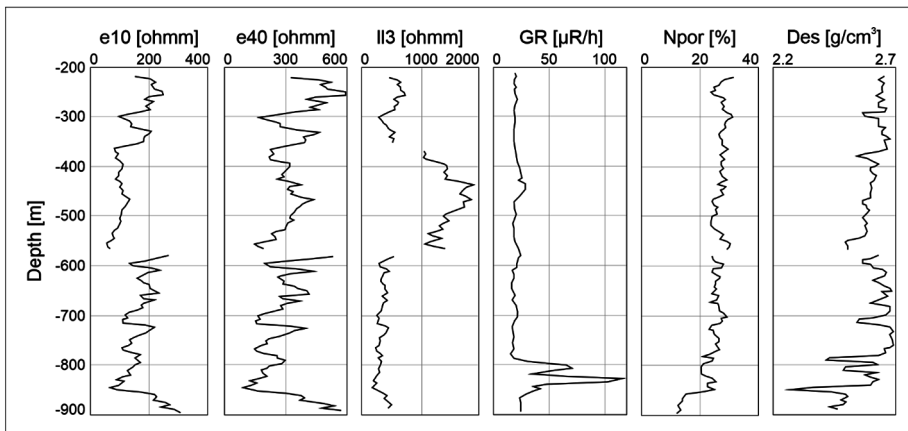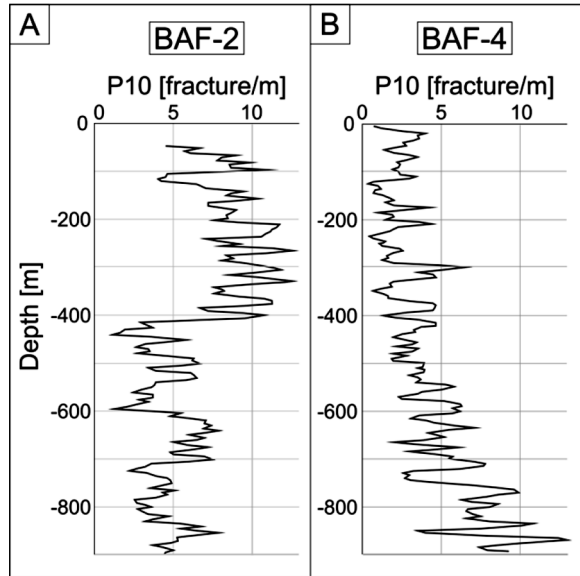

Fig. 4 Geophysical logs of BAF–4. Each log is averaged into 10 m wide sections with a 5 m overlap. e10, e40 and ll3: resistivity logs, GR: natural gamma ray, Npor: neutron porosity, Des: density log measured with short-spaced detector. Discontinuities of the logs are caused by a lack of data

and ll3) (Figs. 4 and 5). Conventional log data and the BHTV measurements were carried out by Geo–Log Ltd. There were short sections for which geophysical log data were not available due to technical reasons. These sections were not included in the study. The vertical resolution of the logs was 10 cm. Each log was averaged into 10-m sections with a 5-m overlap (i.e., moving window). These were then compared with the observed P10 values, which were also calculated for 10-m-wide sections from the BHTV data.
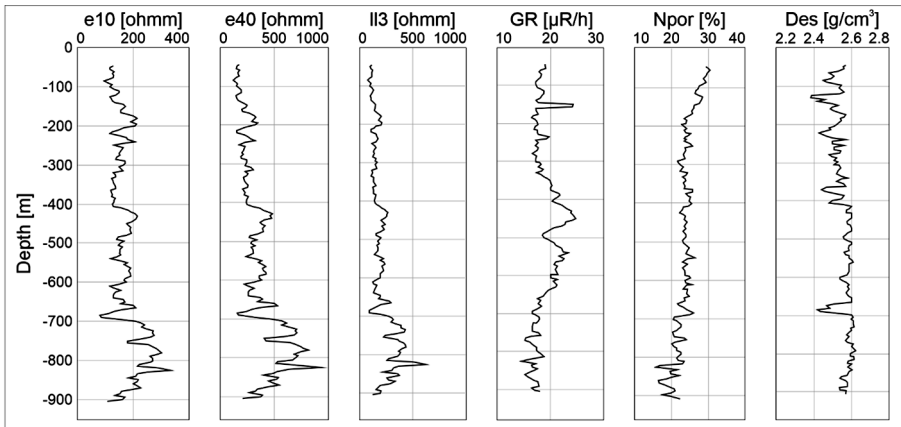
Fig. 5 Geophysical logs of the BAF–2. Each log is averaged into 10 m wide sections with a 5 m overlap. e10, e40 and ll3: resistivity logs, GR: natural gamma ray, Npor: neutron porosity, Des: density log measured with short-spaced detector. Discontinuities of the logs are caused by a lack of data

Table 1 Workflow of the multiple linear regression model with backward elimination method, where a variables was removed from the model if partial F p-value was greater or equal to 0.100. During the backward regression analysis ll3, lg(GR) and Npor were removed from the model

Variables Entered/Removed			
Model	Variables Entered	Variables Removed	Method
1	lg(GR), lg(Des), Npor, e40, ll3, e10	-	Enter
2	-	ll3	Backward (criterion: Probability of F-to-remove >=0.100).
3	-	lg(GR)	Backward (criterion: Probability of F-to-remove >=0.100).
4	-	Npor	Backward (criterion: Probability of F-to-remove >=0.100).

4.2 Results of the multiple linear regression analysis

A multiple linear regression analysis was run to predict P10 from well log data. The normal distribution of the variables was checked, a log transformation of the natural gamma ray and the density values were required. There was linearity as shown by partial regression plots. With the backward stepwise regression method all predictors were entered (e10, e40, ll3, lg(GR), lg(Des), Npor, Table 1.). The elimination criterion was that variables having partial F p-values greater or equal to 0.100 were eliminated from the model (Table 1).

Independence of residuals was identified, as assessed by a Durbin-Watson statistic of 1.426. All three variables of the model are statistically significant to the prediction, $p < .001$ based on t-statistic (Table 3.) Multicollinearity was assessed by tolerance values greater than 0.1 (VIF smaller than 10; Franke, 2010). The backward elimination process found two types of resistance logs to be significant, between which there may be a correlation (Table 3.). The residuals have normal distribution, with a standard deviation close to 1, MEA=0.092 (Fig. 6).

Table 2 Summary of the models of the backward multiple regression analysis with the R^2, standard error of each model and the Durbin-Watson statistics

Model Summary[e]

Model	R^2	Std. Error of the Estimate	Durbin-Watson
1	0.784	0.66463	
2	0.784	0.65724	
3	0.779	0.65690	
4	0.767	0.66829	1.426

e. Dependent Variable: P10

a. Predictors: (Constant). lg(GR), lg(Des), Npor, e40, LL3, e10

b. Predictors: (Constant). lg(GR), lg(Des), Npor, e40, e10

c. Predictors: (Constant). lg(Des), Npor, e40, e10

d. Predictors: (Constant). lg(Des), e40, e10

Table 3 Coefficients of the multiple regression analysis of model 4 with the collinearity and t-statistics

Coefficients[a]

Model		Unstandardized Coefficients		t	Sig.	Collinearity Statistics	
		β	Std. Error			Tolerance	VIF
4	(Constant)	76.174	13.285	5.734	<.001		
	lg(Des)	-173.499	33.095	-5.242	<.001	.321	3.114
	e10	.024	.006	3.997	<.001	.098	10.178
	e40	−.018	.002	-7.287	<.001	.170	5.888

a. Dependent Variable: P10

Fig. 6 Distribution of the residuals with a mean of -0.01 and a standard deviation of 0.966
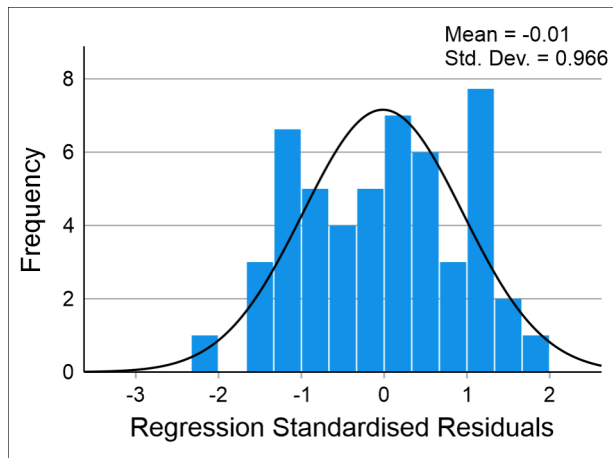
The multiple regression model statistically significantly predicted P10, $F_{(3, 47)} = 51.511$, $p < .001$, with $R^2 = 0.767$ and $S_e = 0.668$ based on e10, e40 and density data in the upper part of the BAF-4 well (Table 4.). The best fitting regression function was as follows:

$$P10 = 76.174 + 0.024*e10 − 0.018*e40 − 173.499*\lg{(Des)} \tag{4}$$

**Table 4** ANOVA (analysis of variance) table of the multiple linear regression

ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig.
4	Regression	69.016	3	23.005	51.511	<.001[b]
	Residual	20.990	47	0.447		
	Total	90.006	50			

a. Dependent Variable: P10

b. Predictors: (Constant), lg(DES), e40, e10

Fig. 7 Linear regression analysis: observed P10 values plotted against (**a**) predicted P10 in BAF–4 at 220–530 m and (**b**) residuals (i.e., difference between the data points and regression line)
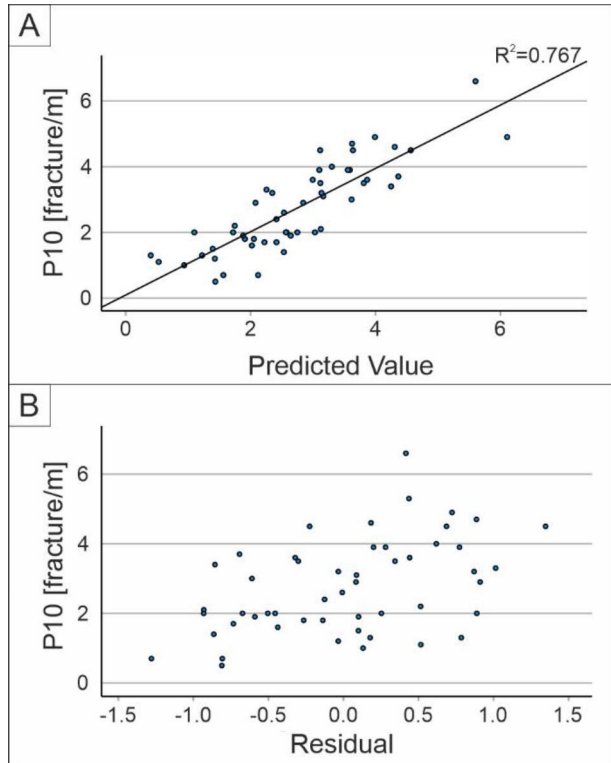
where P10 denotes the fracture density (m^{-1}), e10 and e40 indicate the resistivity [Ωm], Des is the density [g/cm^3].

In the upper section of BAF-4 (220–530 m), a good fit was obtained by the regression analysis with $R^2=0.767$, (Figs. 7a and 8a). The residuals are the difference between the observed value and value predicted by the model. A slight correlation was obtained between the residuals and P10, and the residuals increased slightly with P10 (Fig. 7b).

In the lower section of BAF-4 (530–850 m), fracture density can be predicted using the same function with good accuracy, exept for the section between 750 and 780 m, where the predicted and measured P10 values differ noticeably (Fig. 8b). In this section, the predicted P10 values underestimated the measured P10. Correlation coefficient between 530 and 750 m is $R^2=0.630$, $S_e = 1.623$, MEA=1.024.

Fig. 8 Predicted and measured
P10 values with the depth of
BAF-4: (**a**) in the training sec-
tion at 230–530 m and (**b**) in the
predicted section at 530–850 m.
Black line: measured P10,
orange line: predicted P10, gray
are: section with lower predic-
tion accuracy
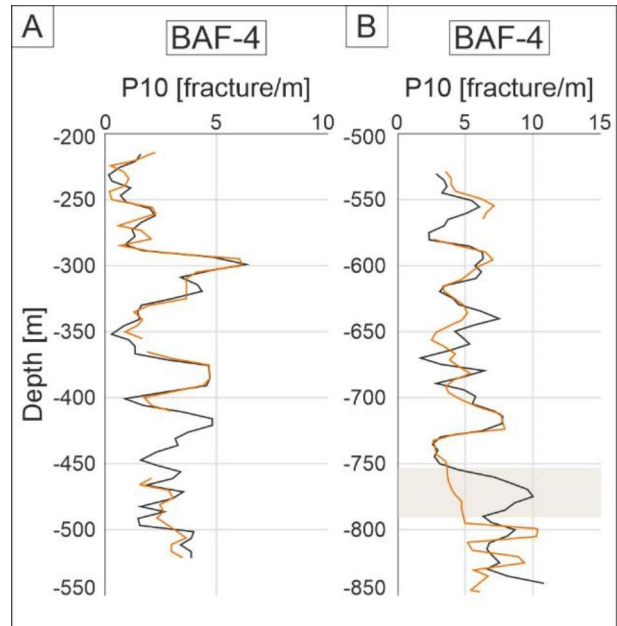
The regression model function computed on the upper segment of BAF-4 was applied
to predict P10 along BAF-2 (Fig. 9). The correlation between the measured and predicted
P10 values was strong with $R^2 = 0.701$, $S_e = 2.239$, MEA$= 2.533$. The predicted P10 values
closely matched the dichotomy of the borehole. However, P10 was slightly underestimated
in the upper 400 m, which was highly fractured. Below 400 m, in the lower half of the
well, the prediction was accurate including the section at 400–600 m, which was the least
fragmented section of the well (Fig. 9). At 700–860 m, P10 was underestimated. In some
extreme cases (e.g., at 720, 780, and 820 m), the model function predicted a negative P10
value because it underestimated the very low measured P10 value. Obviously, a negative
fracture density is not possible.

4.3 Limitations and implications

Results suggest that resistivity and density were the primary influencing factors of P10. The
resistivity is related to formation fluid saturation and depends on the rock type, porosity,
type, composition, and volume of fluid (Archie 1942). Resistivity measurements are usu-
ally used to identify permeable sections and estimate the porosity. Because resistivity logs
are available to locate permeable intervals in an essentially impermeable host rock, they
are suitable for determining fractured zones. The neutron porosity and density also showed
correlations in the partial regression plot with P10. Because open fractures may increase
porosity and decrease bulk density, these logs may also be used to identify fractured zones.
Natural gamma-ray logs are used to detect gamma radiation, which is linked to the amount
of clay in the host rock. This lithological factor may affect the rheology of the claystone
through the grain size and clay content, which can influence the formation and propaga-
tion of fractures (Eisenstadt and Sims 2005). Based on the multiple regression analysis of

Fig. 9 Predicted (orange line) and measured (black line) P10 values with the depth in BAF-2
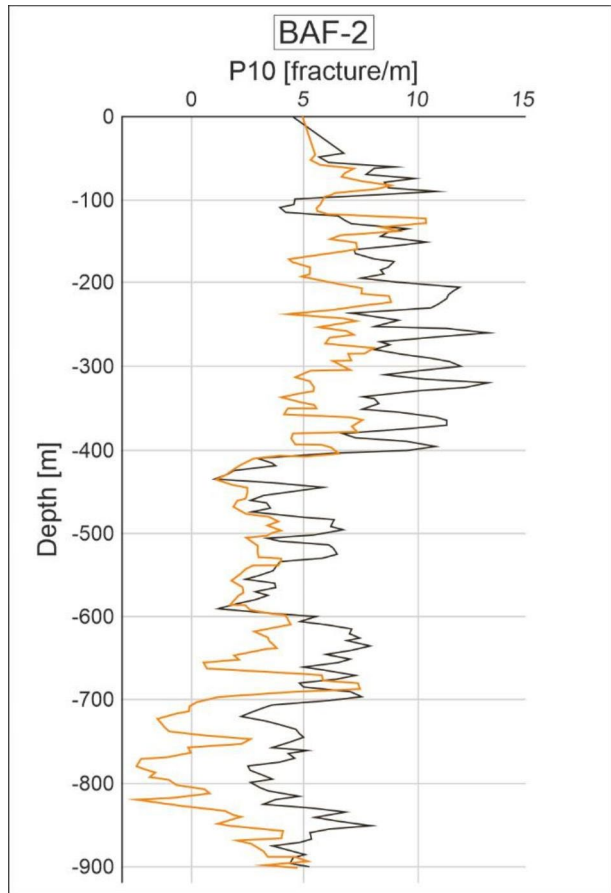
the BCF, neutron porosity and gamma radiation did not influence significantly the fracture density in the formation.

The determined correlation coefficients were significantly influenced by the lithology of the section under study. Because the rock body of the BCF showed little variation in its lithology (Halász 2009; Halász and Halmai 2015), P10 can be predicted using the above multiple linear regression function for much of the formation. However, the estimated and measured P10 values diverged more than usual at several sections in the analyzed boreholes: at 750–780 m in BAF-4 and at 700–860 m in BAF-2.

In BAF-4, the model function underestimated P10 at 750–780 m, at this section, the BHTV detected more discontinuities than predicted. Core images (Fig. 10) show that this interval contains substantially thinner layers than typical for the BCF. The BHTV registers all planar objects in the borehole; therefore, the raw data used to calculate P10 included both the bedding planes and fractures. The P10 prediction based on the geophysical log data was less accurate for this section because these beds do not affect the geophysical characteristics of the rock body. The difference between the observed and estimated values shows that in the laminated zones of the claystone, the actual P10 value may be significantly lower than the calculated value from considering all discontinuities.
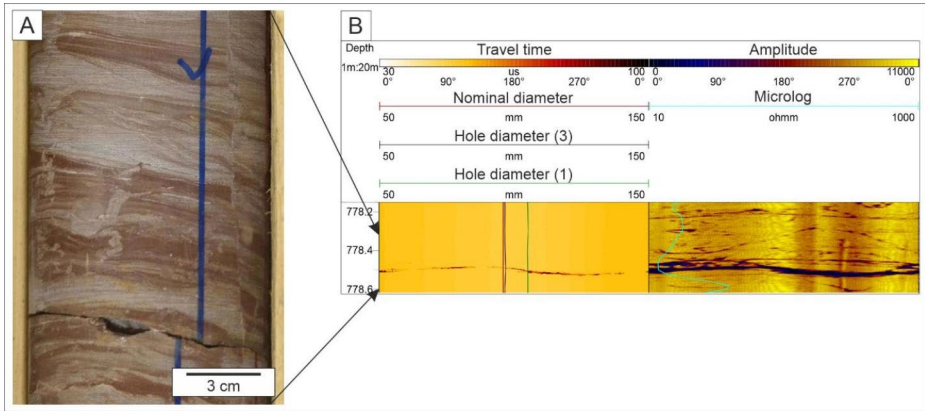
Fig. 10 (**a**) Core sample of BAF–4 between 778.37 and 778.52 m. Thin-layered claystone was detected by the BHTV. Blue arrow: downward direction. (**b**) BHTV image from the same depth, thin layers of the claystone can be detected on the BHTV image on the right

Since the function of the linear regression model was trained by using the upper portion (220–530 m) of BAF-4, the prediction was less accurate in lower parts of the BCF, which had a larger average grain size. When the regression model function was applied to BAF-2, it showed a similar behaviour. Here, the slight decrease in accuracy toward the bedding may be attributed to a gradual change in the lithology (i.e., coarsening grain size). P10 was most significantly underestimated below 700 m, where the fine sandstone layers were common. For accurate prediction of the sandy parts, another regression analysis should be performed on this section. However, it was not possible to validate the calculation on another section with fine sandstone layers among the available boreholes of the BCF. In addition, the sandy part of the BCF is a less important concern regarding the choice of the repository site because the repository should be built in a part with the best sealing properties.

The above results suggest that the prediction accuracy of the regression analysis was primarily affected by the lithology of the rock body. For intervals of typical lithology and consequently representative well log data, P10 can be predicted with a high degree of accuracy. However, special attention must be paid to intervals with atypical well log data suggesting unique changes in the lithology, such as a larger than average grain size or laminar layers.

5 Conclusion

In this study, the relationship between P10 and different geophysical log data was examined for the BCF, which is a potential host for a high-level nuclear waste repository. Two, ~900 m deep boreholes provided a unique opportunity to explore the fracture network characteristics of the BCF. Regression analysis showed a strong linear relationship between P10 and selected geophysical log data. The main influencing factors for P10 were the resistivity (e10, e40) and density. The coefficients of the regression model function were trained on the upper half of BAF-4. The function was then used to predict P10 with high accuracy for other sections typical of the BCF.

In some sections where the lithological properties were atypical, the prediction was less accurate. The lower part of the BCF contains fine sandstone layers, and this change in grain size may have influenced the geophysical properties. In these sandy sections, the function underestimated P10.

The established linear regression model function for the BCF may be used to determine the fracture density in older boreholes that intersect the formation where BHTV log data are not available. By involving more boreholes, the presented approach may provide more information about the spatial extension of communicating fracture clusters and large-scale structural elements such as fault zones.

Authors Contribution Emese Tóth: Writing- Original draft preparation, Visualization, Investigation, Conceptualization, Ervin Hrabovszki: Writing - Review & Editing, Tivadar M. Tóth: Conceptualization, Methodology, Supervision, Writing - Review & Editing.

Declarations

Conflict of Interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Aghli G, Soleimani B, Moussavi-Harami R, Mohammadian R (2016) Fractured zones detection using conventional petrophysical logs by differentiation method and its correlation with image logs. J Petrol Sci Eng 142:152–162. https://doi.org/10.1016/j.petrol.2016.02.002

Ahn J, Apted MJ (eds) (2010) Geological repository systems for safe disposal of spent nuclear fuels and radioactive waste. Woodhead Publ, Oxford Cambridge

Alin A (2010) Multicollinearity. WIREs Computational Statistics, 2: 370–374. https://doi.org/10.1002/wics.84

Anders MH, Laubach SE, Scholz CH (2014) Microfractures: a review. J Struct Geol 69:377–394. https://doi.org/10.1016/j.jsg.2014.05.011

Archie GM (1942) The Electrical Resistivity Log as an aid in determining some Reservoir characteristics. Trans AIME 146:54–62

Árkai P, Demény A, Fórizs I et al (2000) Composition, diagenetic and post-diagenetic alterations of a possible radioactive waste repository site: the Boda Albitic claystone formation, southern Hungary. Acta Geol Hung 43:351–378

Benkovics L, Mansy J-L, Csontos L, Bergerat F (1997) Folding in the Abaliget road cut (Mecsek Mountains). Acta Geol Hung 40:425–440

Bergerat F, Csontos L (1988) Brittle tectonics and paleo-stress field in the Mecsek and Villány Mts (Hungary): correlation with the opening mechanism of the Pannonian Basin. Acta Geol Hung 31:81–100

Bowerman BL, O'Connell RT, Koehler AB (2005) Forecasting, time series and regression, 4 edn. Brooks/Cole Thomson Learning Inc, th ed. United States of America

Csontos L, Benkovics L, Bergerat F et al (2002) Tertiary deformation history from seismic section study and fault analysis in a former european tethyan margin (the Mecsek– Villány area, SW Hungary). Tectonophysics 357:81–102. https://doi.org/10.1016/S0040-1951(02)00363-3

Csontos L, Bergerat F (1992) Reevaluation of the Neogene brittle tectonics of the Mecsek– Villány are (SW Hungary). Ann Universitatis Scientiarum Budapestinensis de Rolando Eötvös Nominatae Sectio geologica 29:3–12

Dong S, Zeng L, Lyu W et al (2020) Fracture identification and evaluation using conventional logs in tight sandstones: a case study in the Ordos Basin, China. Energy Geoscience 1:115–123. https://doi.org/10.1016/j.engeos.2020.06.003

Eberly LE (2007) Multiple linear regression. Methods Mol Biol 404:165–187. https://doi.org/10.1007/978-1-59745-530-5_9

Eisenstadt G, Sims D (2005) Evaluating sand and clay models: do rheological differences matter? J Struct Geol 27:1399–1412. https://doi.org/10.1016/j.jsg.2005.04.010

Fodor L, Csontos L, Bada G et al (1999) Tertiary tectonic evolution of the Pannonian Basin system and neighbouring orogens: a new synthesis of palaeostress data. Geological Society, London, Special Publications 156:295–334. https://doi.org/10.1144/GSL.SP.1999.156.01.15

Franke GR (2010) Multicollinearity. In Wiley International Encyclopedia of Marketing (eds J. Sheth and N. Malhotra).https://doi.org/10.1002/9781444316568.wiem02066

Gamal M, El-Araby AA, El-Barkooky AN, Hassan A (2022) Detection and characterization of fractures in the Eocene Thebes formation using conventional well logs in October field, Gulf of Suez, Egypt. Egyptian J Petroleum 31:1–9. https://doi.org/10.1016/j.ejpe.2022.06.001

Ge Z, Pan S, Li J et al (2020) Estimation of weak anisotropy parameters and fracture density of asymmetric fractures in monoclinic medium. J Geophys Eng 17:1049–1064. https://doi.org/10.1093/jge/gxaa063

Halász A (2009) Cycles and rhythms within the Boda Claystone formation in WELL Ib-4. Cent Eur Geol 52:325–342. https://doi.org/10.1556/CEuGeol.52.2009.3-4.7

Halász A, Halmai Á (2015) The colour analysis method applied to homogeneous rocks. Geologos 21:233–239. https://doi.org/10.1515/logos-2015-0019

Hrabovszki E, Tóth E MTóth T, et al (2020) Potential formation mechanisms of early diagenetic displacive veins in the Permian Boda Claystone formation. J Struct Geol 138:104098. https://doi.org/10.1016/j.jsg.2020.104098

Hrabovszki E, Tóth E MTóth T, et al (2022) Geochemical and microtextural properties of veins in a potential high-level radioactive waste disposal site. J Struct Geol 154:104490. https://doi.org/10.1016/j.jsg.2021.104490

Hrabovszki E, Tóth E, Raucsik B et al (2017) A BAF–2 fúrás töréses szerkezeti elemeinek mikroszerkezeti és cementáció vizsgálata (Bodai Agyagkő Formáció). Földt Közl 147:245. https://doi.org/10.23928/foldt.kozl.2017.147.3.245

Ja'fari A, Kadkhodaie-Ilkhchi A, Sharghi Y, Ghanavati K (2012) Fracture density estimation from petrophysical log data using the adaptive neuro-fuzzy inference system. J Geophys Eng 9:105–114. https://doi.org/10.1088/1742-2132/9/1/013

Konrád G, Hámos G (2006) Geological aspects of determining high activity radioactive waste depository sites in Hungary and the results of the recent research. Acta Geogr ac Geologica et Meteorologica Debrecina 01:33–38

Konrád G, Sebe K (2010) New details of young tectonic phenomena in the western Mecsek Mts and their surroundings. Bull Hung Geol Soc 140:135–162

Konrád G, Sebe K, Halász A (2015) Földtani tektonikai értékelés. In: Sámson M (ed) BAF-2 fúrás dokumentáló és értekelő jelentése. Public Limited Comany for Radioactive Waste Management, Pécs

Konrád G, Sebe K, Halász A, Babinszki E (2010) Sedimentology of a Permian playa lake: the Boda Claystone formation, Hungary. Geologos 16:27–41

Laubach SE, Baumgardner RW, Monson ER et al (1988) Fracture detection in low-permeability Reservoir Sandstone: a comparison of BHTV and FMS Logs to Core. All days. SPE, Houston, Texas, pp SPE–18119

Maros G, Koroknai B, Palotás K et al (2004) Tectonic analysis and structural evolution of the north-eastern Mórágy Block. Annual Report of the Geological Institute of Hungary 371–386

Martinez-Torres LP (2002) Characterization of naturally fractured reservoirs from conventional well logs. Msc Thesis, University of Oklahoma

Massiot C, Townend J, Nicol A, McNamara DD (2017) Statistical methods of fracture characterization using acoustic borehole televiewer log interpretation: BHTV FRACTURE DATA ANALYSIS. J Geophys Res Solid Earth 122:6836–6852. https://doi.org/10.1002/2017JB014115

Nelson RA (2001) Geologic analysis of naturally fractured reservoirs, 2nd edn. Gulf Professional Pub, Boston

Neuzil CE, Tracy JV (1981) Flow through fractures. Water Resour Res 17:191–199. https://doi.org/10.1029/WR017i001p00191

Olive DJ (2017) Linear regression. Springer Cham, Switzerland. https://doi.org/10.1007/978-3-319-55252-1

Pöppelreiter M, García-Carballido C, Kraaijveld M (2010) Borehole Image Log Technology: application across the exploration and production life cycle. Dipmeter and Borehole Image Log Technology. American Association of Petroleum Geologists, pp 1–13

Shalaby MR, Islam MA (2017) Fracture detection using conventional well logging in carbonate Matulla formation, Geisum oil field, southern Gulf of Suez, Egypt. J Petrol Explor Prod Technol 7:977–989. https://doi.org/10.1007/s13202-017-0343-1

Taherdangkoo R, Abdideh M (2016) Application of wavelet transform to detect fractured zones using conventional well logs data (case study: southwest of Iran). IJPE 2:125. https://doi.org/10.1504/IJPE.2016.078818

Tokhmchi B, Memarian H, Rezaee MR (2010) Estimation of the fracture density in fractured zones using petrophysical logs. J Petrol Sci Eng 72:206–213. https://doi.org/10.1016/j.petrol.2010.03.018

Tóth E, Hrabovszki E, Schubert MTóthT F (2020) Shear strain and volume change associated with sigmoidal vein arrays in the Boda Claystone. J Struct Geol 138:104105. https://doi.org/10.1016/j.jsg.2020.104105

Tóth E, Hrabovszki E, Schubert F, Tóth M T (2022a) Discrete fracture network (DFN) modelling of a high-level radioactive waste repository host rock and the effects on its hydrogeological behaviour. J Struct Geol 156:104556. https://doi.org/10.1016/j.jsg.2022.104556

Tóth E, Hrabovszki E, Schubert F, Tóth M T (2022b) Lithology-controlled hydrodynamic behaviour of a fractured sandstone–claystone body in a Radioactive Waste Repository Site, SW Hungary. Appl Sci 12:2528. https://doi.org/10.3390/app12052528

Tranmer M, Murphy J, Elliot M, Pampaka M (2020) Multiple Linear Regression (2nd Edition); Cathie Marsh Institute Working Paper 2020-01. https://hummedia.manchester.ac.uk/institutes/cmist/archive-publications/working-papers/2020/2020-1-multiple-linear-regression.pdf

Varga A (2009) Petrology and Geochemistry of the Paleozoic–Lower Triassic Siliciclastic Rocks from Southern Transdanubia, Hungary. Eötvös Loránd University, Budapest, Hungary

Varga A, Raucsik B, Szakmány G, Máthé Z (2006) Mineralogical, petrological and geochemical characteristics of the siliciclastic rock types of Boda Siltstone formation. Bull Hung Geol Soc 136:201–231

Varga A, Szakmány G, Raucsik B, Máthé Z (2005) Chemical composition, provenance and early diagenetic processes of playa lake deposits from the Boda Siltstone formation (Upper Permian), SW Hungary. Acta Geol Hung 48:49–68. https://doi.org/10.1556/AGeol.48.2005.1.2

Vinod HD (1973) Generalization of the durbin-watson statistic for higher order autoregressive processes. Commun Stat 2(2):115–144. https://doi.org/10.1080/03610927308827060

Witherspoon PA, Wang JSY, Iwai K, Gale JE (1980) Validity of Cubic Law for fluid flow in a deformable rock fracture. Water Resour Res 16:1016–1024. https://doi.org/10.1029/WR016i006p01016

Zazoun RS (2013) Fracture density estimation from core and conventional well logs data using artificial neural networks: the Cambro-Ordovician reservoir of Mesdar oil field, Algeria. J Afr Earth Sc 83:55–73. https://doi.org/10.1016/j.jafrearsci.2013.03.003

Zemanek J, Glenn EE, Norton LJ, Caldwell RL (1970) Formation evaluation by inspection with the borehole televiewer. Geophysics 35:254–269. https://doi.org/10.1190/1.1440089