HYDROGEOPHYSICAL CHARACTERIZATION OF GROUNDWATER FORMATIONS BASED ON WELL LOGS: CASE STUDY ON CENOZOIC CLASTIC AQUIFERS IN EAST HUNGARY

NORBERT PÉTER SZABÓ[1,2*], ANETT KISS[1], ANETT HALMÁGYI[1]

[1]Department of Geophysics, University of Miskolc
[2]MTA-ME Geoengineering Research Group, University of Miskolc
**e-mail: norbert.szabo.phd@gmail.com*

1. Abstract

Ground geophysical surveys can be effectively used for detecting and delineating shallow groundwater structures. For calculating the freshwater reserves, well-logging measurements need to be made in prospecting wells. In this paper, groundwater formations are evaluated using electric and nuclear logging data to extract the petrophysical and hydraulic parameters of aquifers and aquitards. To quantify the relative fractions of rock constituents, the effective porosity, shale content, water saturation and rock-matrix volumes should be estimated. The effective layer-thicknesses of permeable beds are of importance in locating the depth-intervals of water exploitation. The above parameters cannot be calculated reliably when the zone parameters such as cementation exponent, tortuosity factor, pore-water density and resistivity are not certain. With accurate petrophysical and zone parameters, an estimate can be given to hydraulic conductivity, which is one of the most important hydraulic rock properties in solving hydrogeophysical problems. First, a comprehensive interpretation method proposed by Professor János Csókás (1918-2000), former head of the Department of Geophysics, University of Miskolc, is used to give an estimate of hydraulic conductivity and critical velocity of flow without the need for grain-size data. Then, shale volume and hydraulic conductivity are determined separately by statistical factor analysis of well logs. Effective porosity, specific surface of grains and water saturation can be derived by well-known deterministic equations. Those of core measurements confirm the results of well log analysis. A set of detailed regression analyses is performed to specify the local regression relationships between the estimated parameters. It is also shown that there is a strong correlation between shale volume and hydraulic conductivity (and other quantities) and that the independent interpretation results are consistent. The advantage of the Csókás and factor analysis-based approaches is that instead of using a single well log, they utilize all types of well logs sensitive to the relevant petrophysical/hydraulic parameters for a more reliable hydrogeophysical characterization of aquifers.

2. Introduction

Well-logging methods are widely used in the reconnaissance of mineral and hydrocarbon resources, as they provide detailed in-situ information about the geometrical and petrophysical parameters of geological structures [1]. They are applicable also for the investigation of shallow formations, for instance, in water prospecting and solving environmental and engineering geophysical problems [2]. The main task encountered by well log analysts when solving hydrogeophysical problems is to estimate the layer thickness, effective porosity, and water (sometimes air) saturation, amount of shaliness, matrix volumes and hydraulic conductivity of unconsolidated beds as accurately as possible. Some petrophysical properties can be given with relatively high accuracy, for instance porosity is normally estimated with an error of 1-2 p.u. depending on the uncertainty of measured data. At the same time, the estimation error of hydraulic conductivity can reach an order of magnitude, thus, well logs can be primarily applied to detect only the variation of hydraulic conductivity along a borehole or between neighboring boreholes. Suitable borehole geophysical methods for estimating hydrogeophysical parameters with related applications are detailed in [3].

Almost all types of well-logging suites used in hydrocarbon exploration can be implemented in groundwater prospecting. Natural gamma-ray intensity and spontaneous potential logs are applied for lithology identification and layer-thickness determination. The former measures the natural radioactivity of formations caused by different amounts of potassium, thorium and uranium content. The latter records the values of electric potential between a surface and a downhole electrode excited by the ion movement between the drilling mud and original pore fluid as well as the presence of shale. Both of them can be used to predict shale content quantitatively. For porosity determination, nuclear well logs such as formation density and neutron porosity are used. Bulk density observed by gamma-gamma probes shows a strong inverse proportionality to porosity. Neutron-neutron measurements are mainly sensitive to the hydrogen index of formations and allow us to infer the total porosity in the absence of hydrogen atoms in the rock matrix. Several types of resistivity tools with different depth of investigation and vertical resolution can be applied to detect the invasion profile and to estimate water saturation in different zones around the borehole. Since freshwater has higher resistivity than brine, traditional (non-focused) resistivity probes are generally suitable for groundwater exploration. While shallow resistivity tools measure the apparent resistivity of the zone invaded by mud, the deep resistivity instrument observes the same quantity in the original (non-invaded) formation. The resistivity readings are corrected to predict the true resistivity of groundwater formations, which is of great importance in calculating water saturation in aquifers. Well-logging methods usually encompass also technical measurements that are not used directly in the petrophysical characterization of rocks, but that give important information on the technical conditions of the borehole wall and its environment, pressure, temperature, flow rate and composition of the original pore fluid along the borehole.

In addition to conventional well-logging techniques, there is a possibility for using advanced tools in the evaluation of aquifers. In oilfield applications, the direct determination of permeability as a related quantity is possible by means of the nuclear magnetic resonance log. The ground geophysical application of this technique, known as magnetic resonance sounding, is an emerging method in hydrogeology. Borehole nuclear resonance magnetism has been lately adapted from the oilfield for hydrogeological applications, using boreholes typical of environmental and hydrogeological investigations [4]. Although nuclear magnetic resonance is a very expensive method, it has the added advantage of not only providing free-fluid porosity, it can also be used to determine the distribution of pore sizes and fluid characteristics to provide a better estimate of the hydraulic properties of rocks.

The petrophysical parameters can be calculated by deterministic, inverse or statistical modeling procedures. The most common deterministic approaches are based on the individual analysis of well logs. Estimation that is more reliable can be given by using several well logs simultaneously. In this study, two advanced well-logging methodologies are discussed and compared to each other. In shallow clastic sediments, the evaluation of petrophysical and hydraulic parameters generally requires the preliminary knowledge of grain sizes and several fluid parameters. In the absence of direct geophysical measurements, one is confined to measuring some related physical parameters or taking rock samples from the borehole to extend hydrogeological information to a local area. In order to avoid core sampling, CSÓKÁS [5] worked out a comprehensive interpretation method to give an estimate of the hydraulic conductivity of unconsolidated freshwater-bearing formations based solely on well-logging data. By incorporating the field experiments of ALGER [6], Hazen's effective grain-size can be substituted by Archie's formation factor, which can be measured directly from well logs. The derived formula, including porosity and true resistivity of aquifers, gives a continuous estimate of hydraulic conductivity for the entire length of a borehole. In addition, the Csókás method comprises the determination of critical velocity of flow, which can be used to estimate the highest value of sand-free yield from the technical data of the filtering surface.

Factor analysis is traditionally used to reduce the dimensionality of multivariate statistical problems [7]. Technically, it is possible to decompose a data matrix of observed physical quantities of any dimension to a matrix of fewer statistical variables. This is called an exploratory statistical method, because it allows us to extract information on latent variables not directly measurable by well-logging probes. In this study, factor analysis is used to reveal correlation relations between well-logging data and petrophysical/hydraulic parameters of aquifers. An earlier study of hydrocarbon wells showed a non-linear relationship between one of the new variables derived by factor analysis and shale volume, which proved to be nearly independent of the measurement area [8].

A Hungarian feasibility study is presented in this paper. The multivariate statistical method applied to the investigated water well finds a similarly strong regression relation between the statistical factor and shale volume (and hydraulic conductivity) as a key parameter in the prediction of exploitable reserves of aquifers. Those of core measurements validate the interpretation results obtained independently by the Csókás method and factor analysis. In addition, regression relations and correlation coefficients between petrophysical/hydraulic parameters are given for the area of Baktalórántháza, East Hungary. It is shown that there is a strong correlation between shale volume and hydraulic conductivity (and other quantities). The independent interpretation results are consistent. The advantage of the presented approaches is that instead of using a single well log they process simultaneously all suitable well logs sensitive to the relevant petrophysical parameters in order to support groundwater exploration with reliable petrophysical/hydrogeological information.

3. Conventional formation evaluation

Darcy's equation is one of the basic relationships of hydrogeology to describe the flow of water through a porous formation

$$\frac{\partial \mathbf{u}}{\partial t} = -\frac{\kappa}{\Phi\mu}\nabla p, \qquad (1)$$

where κ (m^2) denotes rock permeability, Φ (volume/volume = v/v) is formation porosity, μ (Ns/m^2) is dynamic viscosity, $\mathbf{u}$ (m) is the relative displacement vector of the fluid and p (N/m^2) is the pore pressure. Hydraulic conductivity $k=\kappa\rho_w g/\mu$ (m/s) as a related quantity expresses the ease with which the water flows through the pore spaces. Quantity k is influenced by several properties of the rock matrix and pore fluid of primary porosity aquifers such as viscosity and density of pore-fluid (ρ_w in g/cm3), grain-size distribution, porosity, and water saturation (S_w in v/v).

The Kozeny-Carman equation has achieved widespread use as a standard model for estimating hydraulic conductivity of aquifers. In theory, the rock with primary porosity is treated as an assembly of capillary channels, which satisfies the Navier-Stokes equation. The following form of the Kozeny-Carman equation is one of the most widely used models for the estimation of hydraulic conductivity [9]

$$k = \frac{\rho_w g}{\mu}\frac{d^2}{180}\frac{\Phi^3}{(1-\Phi)^2}\ , \qquad (2)$$

where d (cm) is the dominant grain diameter and g (cm/s^2) is the normal acceleration of gravity (k is given in units of cm/s). Since rock samples can be taken from boreholes, the dominant grain diameter can be estimated from the grain-size distribution curve [10]

$$d = \frac{d_{10} + d_{60}}{2} \left(\frac{d_{10}}{d_{60}} \right)^{1/2} , \tag{3}$$

where d_{10} (cm) and d_{60} (cm) are the grain diameters at 10% and 60% cumulative frequencies, respectively. The dominant grain size in Eq. (3) can be defined as the diameter of a homo-disperse conglomerate of grains, the surface of which equals that of the real sample with actual grain-size distribution and the same density. The effective porosity of shaly sandy beds fully saturated with freshwater (S_w=1) can be derived from well logs, for instance from gamma-gamma measurements by neglecting the term of air saturation

$$\Phi = \frac{\rho_{sd} + V_{sh}(\rho_{sh} - \rho_{sd}) - \rho_b^{(m)}}{\rho_{sd} - 1} , \tag{4}$$

where $\rho_b^{(m)}$ (g/cm^3) is the bulk density measured and ρ_{sd} and ρ_{sh} denote the density of sand and shale components, respectively. The amount of shaliness appearing in Eq. (4) can be estimated by the following empirical formula generally used in young sedimentary formations [11]

$$V_{sh} = 0.083 \left[2^{3.7 \left(\frac{GR - GR_{\min}}{GR_{\max} - GR_{\min}} \right)} - 1 \right], \tag{5}$$

where GR (cps) is the measured natural gamma-ray intensity and GR_{min} and GR_{max} are the extreme values of the gamma-ray log in the groundwater zone. Equation (5) does not depend on the water type, but caution should be used in rocks including radioactive non-clay minerals or fractures filled with uranium- or thorium-rich water. The correlation between grain size and porosity should be revealed before well logs are applied to calculate the continuous curve of hydraulic conductivity, which is of great importance in the determination of water reserves as well as in the management and protection of groundwater supply.

In more advanced approaches, several well logs are processed together in one interpretation procedure. If the number of unknowns equal to that of the observed logs the modeling functions are treated as a set of linear equations, which can be solved graphically or numerically. If we measure more data types than unknowns some inversion method is generally used to extract petrophysical parameters. In the evaluation of domestic aquifers normally natural gamma-ray intensity (GR in cpm), spontaneous potential (SP in mV), neutron-neutron intensity (NN in kcpm), bulk density (ρ_b in g/cm^3), shallow and deep resistivity (R_s and R_d in ohmm) logs are recorded. The petrophysical parameters can be related to well-logging data by means of empirical modeling equations. These mathematical relations, called probe response equations, can be used to predict data in a forward modeling procedure. The values of theoretical data would be measured along the borehole if the geological structure were characterized by the assumed (exactly known) model parameters. In general, borehole geophysical data can be expressed as a set of nonlinear equations including the physical properties of rock matrix and fluid components weighted by the relative volumes of rock constituents. The following set of response functions can be used for the solution of the forward problem in fully saturated shaly sandy aquifers [12]

$$SP = SP_{sh} V_{sh} - C \cdot \lg \frac{R_{mf}}{R_w} \left(1 - V_{sh}\right), \tag{6}$$

$$GR = GR_{sd} + \frac{1}{\rho_b} \left(V_{sh} GR_{sh} \rho_{sh} + V_{sd} GR_{sd} \rho_{sd}\right), \tag{7}$$

$$\rho_b = \Phi \rho_{mf} + V_{sh} \rho_{sh} + V_{sd} \rho_{sd}, \tag{8}$$

$$NN = \Phi NN_f + V_{sh} NN_{sh} + V_{sd} NN_{sd}, \tag{9}$$

$$R_s = \left[\left(\frac{V_{sh}^{(1-0.5V_{sh})}}{R_{sh}^{1/2}} + \frac{\Phi^{m/2}}{\left(aR_{mf}\right)^{1/2}} \right) \right]^{-2}, \tag{10}$$

$$R_d = \left[\left(\frac{V_{sh}^{(1-0.5V_{sh})}}{R_{sh}^{1/2}} + \frac{\Phi^{m/2}}{\left(aR_w\right)^{1/2}} \right) \right]^{-2}, \tag{11}$$

$$\Phi + V_{sh} + V_{sd} = 1, \tag{12}$$

where V_{sd} (v/v) is the volume of sand and R_{mf}, R_w and R_{sh} (ohmm) denote the resistivity of mud filtrate, pore-water and shale, respectively. In Eqs. (6)–(11), there are some model parameters that do not vary (or vary just slowly) with depth, called zone parameters, which are practically kept constant during the forward modeling procedure. Zone parameters with subscripts *sh*, *sd*, *mf*, *w* refer to physical properties of shale, sand, mud-filtrate and water, respectively, whereas constants *m*, *n*, *a* represent the textural properties of rocks given from literature or empirical techniques. The definition, including the measurement unit of petrophysical and zone parameters, can be found in the section of list of symbols. Equation (12) is the material balance equation for the rock environment, which is used to specify the relative fractions of rock constituents per unit volume of rock and constrain the domain of volumetric parameters in the interpretation procedure. In inverse modeling, the model parameters (i.e. porosity, water saturation, shale content and matrix volume) are estimated simultaneously. Inversion procedures have the added advantage of also giving the estimation errors of petrophysical parameters, which characterize quantitatively the accuracy and reliability of inversion results. The hydraulic conductivity can be derived from the inversion results by using proper empirical equations [13]. The methodology of inverse modeling and some shallow examples can be found in DRAHOS [14] and SZABÓ and DOBRÓKA [15].

4. The Csókás method

CSÓKAS [5] worked out a comprehensive interpretation methodology to extract petrophysical and hydraulic parameters of aquifers solely from well logs. The Csókás' model is an empirically modified form of Eq. (2) applicable in shallow formations, based on the relation between the effective grain size and formation factor of freshwater-bearing unconsolidated sediments. The method gives a continuous estimate of hydraulic conductivity along a borehole by using electric and nuclear logging measurements without the need for grain-size data. In fully saturated aquifers the formation factor (*F*) is defined as the ratio of the resistivity of rock (R_0 in ohmm) to that of the pore-water

$$F = \frac{R_0}{R_w}. \tag{13}$$

ALGER [6] found a direct proportionality between the formation factor and grain size of freshwater saturated sediments in the laboratory that is the opposite of that experienced in hydrocarbon fields (i.e. in brine-saturated reservoirs). Based on these observations, Hazen's effective grain diameter (d_{10}) determined from sieve analysis was related to the formation factor of unconsolidated sediments

$$d_{10} = C_d \lg F , \tag{14}$$

where C_d=5.22·10^{-4} was proposed for not too poorly sorted sediments with a formation factor less than 10. This requirement is normally met in shallow clastic aquifers. An interpretation methodology to evaluate freshwater-bearing (shaly) sands based on Eq. (14) was proposed by ALGER [6] for different types of well-logging suites adapted from the oilfield. Equation (14), as a constraint relation, forms the bridge between well-logging measurements and hydraulic conductivity of aquifers. The dominant grain size can be given from the grain-size distribution curve. The uniformity coefficient $U=d_{60}/d_{10}$ acts as a shape parameter which characterizes the form of the grain-size distribution curve and quantifies the degree of uniformity in a granular material. KOVÁCS [16] connected the uniformity coefficient of sands to the dominant grain diameter as d/d_{10}=1.919·lg$U$+1. For poorly sorted sediments (U>5), quantity U is inversely proportional to the logarithm of hydraulic conductivity. For not so badly sorted sands (2.0≤U≤ 2.5) the previous equation takes the form $d = 1.671 \cdot d_{10}$, with which Eq. (14) modifies to

$$d = 1.671 \cdot C_d \lg \frac{R_0}{R_w} . \tag{15}$$

ARCHIE [17] suggested an empirical formula developed from laboratory measurements made on numerous samples

$$F = a\Phi^{-m} , \tag{16}$$

where m is the cementation exponent (in poorly compacted sediments m~1.5–1.7) and a is the tortuosity coefficient (a~1). The study of ALGER [6] showed that the formation factor not only depends on porosity, but also on the resistivity of pore water and grain size in freshwater saturated sediments with primary porosity. Dispersed clay particles can considerably modify the resistivity of pore water, which then affects the value of the formation factor appearing in Eq. (15). OGBE and BASSIOUNI [18] coupled the tortuosity factor with porosity and the formation factor

$$a^2 = \left(\frac{R_0}{R_w} \Phi \right)^{1.2} . \tag{17}$$

To find an estimate to hydraulic conductivity from well logs, Eq. (2) should be properly modified [16]

$$k = \frac{1}{5} \frac{g}{\upsilon} \frac{\Phi^3}{(1-\Phi)^2} \left(\frac{d}{\alpha_0} \right)^2 , \tag{18}$$

where α_0 is the average shape factor of sample particles in the range of 7 and 11 for sands (the average is 10). The kinematic viscosity of water $\upsilon=\mu/\rho_w$ (m^2/s) can be expressed in the function of formation temperature. The ratio of gravity acceleration and kinematic viscosity for water is $g/\upsilon=5.517\cdot10^4\cdot C_t$ ($m^{-1}s^{-1}$), where C_t is a temperature dependent coefficient calculated as $C_t=1+3.37\cdot10^{-2}T+2.21\cdot10^{-4}T^2$ (where T is given in units of °C). PIRSON [19] published another form of the Kozeny equation used to predict permeability

$$k=\frac{1}{5}\frac{g}{\upsilon}\frac{\Phi^3}{(1-\Phi)^2}\left(\frac{1}{a\cdot S}\right)^2, \tag{19}$$

where S (1/m) denotes the specific surface of the rock. By comparing Eqs. (18)–(19) the following identical equation is derived

$$\left(\frac{d}{\alpha_0}\right)^2=\left(\frac{1}{a\cdot S}\right)^2. \tag{20}$$

GÁLFI and LIEBE [20] summarized several empirical relations between the specific electric resistance and hydraulic conductivity for sands and gravels. In freshwater aquifers, the electric current is hardly conducted through the spaces between the grains, rather mainly on the surfaces of particles. Thus, the resistivity is inversely proportional to the specific surface of rock grains. By assuming that the sedimentary rock is composed of spherical particles the specific surface can be calculated as

$$S=6\frac{(1-\Phi)}{d}. \tag{21}$$

The combination of Eqs. (15) and (21) gives

$$S^2=\frac{36(1-\Phi)^2}{\left(1.671\cdot C_d\cdot \lg\frac{R_0}{R_w}\right)^2}. \tag{22}$$

The Csókás model based on Eqs. (17), (18), (20), (22) can give an estimate to hydraulic conductivity in units of m/s

$$k=C_k\frac{\Phi^3}{(1-\Phi)^4}\frac{\left(\lg\frac{R_0}{R_w}\right)^2}{\left(\frac{R_0}{R_w}\Phi\right)^{1.2}}, \tag{23}$$

where $C_k=855.7\cdot C_tC_d^{\,2}$ is a proportionality constant. Good aquifers are characterized by hydraulic conductivities k (m/s)$>10^{-6}$, while aquitards are indicated by k (m/s)$<3\cdot10^{-8}$. The uniqueness of the Csókás formula resides in the fact that all parameters in Eq. (23) can be derived from well logs, and thus therewith a continuous (in-situ) estimate can be given for hydraulic conductivity along a borehole.

The Csókás method gives further possibilities for the hydraulic characterization of aquifers. Tangential stresses are developed at the surfaces of rock grains because of the water flow. The critical velocity of flow at which the grains of size d_{10} are made to move to the direction of the borehole can be estimated from hydraulic conductivity [21]

$$v_c \approx 2(d_{10})^{1/2} \approx 0.067(k)^{1/2}, \tag{24}$$

where v_c and k are both given in m/s, while d_{10} is given in mm. For maintaining a sand-free exploitation process, the above information can be used for setting an optimal pumping rate. Above the critical velocity, there is a risk that the water becomes of worse quality or sand grains invade the well. The volume capacity of the well in m^3/s can be calculated as

$$Q_{max} = 2\pi r_0 h_0 v_c, \tag{25}$$

where r_0 (m) is the radius of filter pack and h_0 (m) is its length. Well-site experience shows that significantly higher velocities can be applied than suggested in Eq. (24). Consequently, a higher value of water discharge can be produced than that of Q_{max}. From grain-size analysis a more realistic estimate to critical velocity was suggested by KASSAI and JAMBRIK [22]

$$v_c = 10^{0.446\lg(d)+0.1654-3}, \tag{26}$$

where v_c is measured in m/s, d is given in mm. Equation (26) gives a good approximation in the range of 0.09–5 mm dominant grain sizes. By combining Eqs. (25)–(26), the optimal water discharge can be estimated by

$$Q_{max} = 2\pi \cdot 10^{-3} \cdot r_0 h_0 \left(10^{0.446\lg(d)+0.1654}\right), \tag{27}$$

which is multiplied by $6 \cdot 10^4$ to obtain the output in the unit of l/min. The Csókás procedure is tested and compared to statistical factor analysis using well-logging data collected in Baktalórántháza, East-Hungary (Section 6).

5. Exploratory factor analysis

The multivariate statistical procedure is applicable to transform numerous geophysical data types into smaller number of variables called factors. As a result, a few factors explain the determinant amount of total variance of measurement data, which can be connected to petrophysical/hydraulic properties of the investigated geological structure. In the first step, well logs are standardized

$$d_{ln}^{(obs)'} = \frac{\left(d_{ln}^{(obs)} - \bar{d}_l^{(obs)}\right)}{\sqrt{\frac{1}{N-1}\sum_{n=1}^{N}\left(d_{ln}^{(obs)} - \bar{d}_l^{(obs)}\right)^2}}, \tag{28}$$

where $d_{ln}^{(obs)'}$ denotes the n-th scaled data of the l-th observed well log, $\bar{d}_l^{(obs)}$ is the average value of raw data of the l-th well log (L is the number of borehole geophysical tools and N is the number of measuring points in the processed depth-interval). All standardized data are gathered into data matrix **D'**, which is decomposed by the model of factor analysis as

$$\mathbf{D}' = \mathbf{F}\mathbf{W}^{\mathrm{T}} + \mathbf{E}, \tag{29}$$

where **F** denotes the N-by-M matrix of factor scores, **W** is the L-by-M matrix of factor loadings, **E** is the N-by-L matrix of residuals, M is the number of extracted factors (T indicates the operator of matrix transpose). The observed variables are developed as the linear combination of factors. The scores of the first factor as the elements in the first column of matrix **F** give the well log of the first factor, which explains the largest part of variance of the well-logging data. Other subsequent factors represent a relatively lower portion of variances. The individual weights of each data type associated with the factors are given in the matrix of factor loadings **W**, which measures the degree of correlation between the factors and original data, respectively. Since the factors are assumed linearly independent $\left(\mathbf{F}^{\mathrm{T}}\mathbf{F}/N = \mathbf{I}\right)$, the correlation matrix of the standardized well-logging data is

$$\mathbf{R} = \frac{1}{N}\mathbf{D}'^{\mathrm{T}}\mathbf{D}' = \frac{1}{N}\left(\mathbf{F}\mathbf{W}^{\mathrm{T}}\right)^{\mathrm{T}}\left(\mathbf{F}\mathbf{W}^{\mathrm{T}}\right) + \mathbf{E}^2 = \mathbf{W}\mathbf{W}^{\mathrm{T}} + \mathbf{\Psi}, \tag{30}$$

where $\mathbf{\Psi}$ is the diagonal matrix of specific variances (**I** is the identity matrix). The diagonal elements of matrix **R** are equal to unity and are added up by the variances of the standardized observed variables. If one neglects the term $\mathbf{\Psi}$ in Eq. (30) the following reduced correlation matrix is introduced

$$\mathbf{R}^* = \mathbf{W}\mathbf{W}^{\mathrm{T}} = \mathbf{R} - \mathbf{\Psi} = \begin{pmatrix} h_1^2 & r_{12} & \cdots & r_{1L} \\ r_{12} & h_2^2 & \cdots & r_{2L} \\ \vdots & \vdots & \ddots & \vdots \\ r_{1L} & r_{2L} & \cdots & h_L^2 \end{pmatrix}, \tag{31}$$

where the i-th element in the main diagonal represents the communality defined as

$$h_i^2 = \sum_{l=1}^{L} W_{il}^2 \le 1. \tag{32}$$

From Eq. (31) it follows that matrix $\mathbf{\Psi}$ represents the part of variance of the observations that are not explained by the common factors, which is calculated in the knowledge of communalities

$$\mathbf{\Psi} = \mathbf{I} - \mathbf{H}^2. \tag{33}$$

If the elements of the L-by-L matrix of communalities ($\mathbf{H}^2$) are much smaller than unity, the observed variables can hardly be explained with the common factors. Otherwise, the information of original variables can be represented well by some factors.

The factor loadings should be estimated first, to which three types of solutions are normally applied. In case of $\mathbf{\Psi} = \mathbf{0}$ the problem reduces to the solution of an eigenvalue problem, which is equivalent to Principal Component Analysis (PCA). In this approximation the reduced correlation matrix $\mathbf{R}^*$ is decomposed as

$$\mathbf{W}\mathbf{W}^{\mathrm{T}} = \mathbf{Z}\mathbf{\Lambda}\mathbf{Z}^{\mathrm{T}}, \tag{34}$$

where $\mathbf{Z}$ is the L-by-M matrix of eigenvectors, $\mathbf{\Lambda}$ is the M-by-M diagonal matrix of eigenvalues. From Eq. (34) the matrix of factor loadings is

$$\mathbf{W} = \mathbf{Z}\mathbf{\Lambda}^{1/2}, \tag{35}$$

where the matrix element $\Lambda_{ll}^{1/2} = \lambda_l^{1/2}$ is computed by the l-th eigenvalue. Since the covariance matrix of observed variables (i.e. correlation matrix of standardized data) contains the data variances in its main diagonal, the factors can be classified by the eigenvalues ($\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_M \geq 0$). The direction of the largest data variance is indicated by the eigenvector corresponding to the largest eigenvalue, which is called the first principal direction (first factor). The second largest eigenvalue and its vector represent the second principal direction (second factor), which is perpendicular to the first one etc. The factor scores are calculated from factor loadings and measured data. PCA ignores matrix $\mathbf{E}$ in Eq. (30), which allows for the unique solution of a set of linear equations. For instance, consider five different types of well logs (GR, SP, NN, DEN, RD) represented by standardized data at six different depth levels (L=5, N=6). Factor analysis gives an estimate to two uncorrelated factors

$$\begin{pmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \\ F_{31} & F_{32} \\ F_{41} & F_{42} \\ F_{51} & F_{52} \\ F_{61} & F_{62} \end{pmatrix} = \begin{pmatrix} GR_{11} & SP_{12} & NN_{13} & DEN_{14} & RD_{15} \\ GR_{21} & SP_{22} & NN_{23} & DEN_{24} & RD_{25} \\ GR_{31} & SP_{32} & NN_{33} & DEN_{34} & RD_{35} \\ GR_{41} & SP_{42} & NN_{43} & DEN_{44} & RD_{45} \\ GR_{51} & SP_{52} & NN_{53} & DEN_{54} & RD_{55} \\ GR_{61} & SP_{62} & NN_{63} & DEN_{64} & RD_{65} \end{pmatrix} \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \\ W_{31} & W_{32} \\ W_{41} & W_{42} \\ W_{51} & W_{52} \end{pmatrix}, \tag{36}$$

where F_{nm} represents the score of the m-th factor at the n-th depth, L_{lm} is the load of the m-th factor put on the l-th observed variable. The drawback of the above method is that the portion of data variance unexplained by the factors is neglected, thus, the resultant principle components do not represent the total variance of measured data. With the second approach the matrices $\mathbf{W}$ and $\mathbf{\Psi}$ are estimated in a simultaneous optimization procedure. JÖRESKOG [23] suggested the minimization of the following objective function

$$\Omega(\mathbf{W}, \mathbf{\Psi}) = \mathrm{tr}\left(\mathbf{R} - \mathbf{W}\mathbf{W}^{\mathrm{T}} - \mathbf{\Psi}\right)^2 = \min . \tag{37}$$

For solving the optimization problem the use of the maximum likelihood method is generally applied, which can give a robust solution [24]. The third alternative for estimating the factor loadings is the use of a non-iterative approximate algorithm [23]. At first the following matrix is calculated

$$\mathbf{S}^* = \left(\mathrm{diag}\,\mathbf{S}^{-1}\right)^{1/2}\mathbf{S}\left(\mathrm{diag}\,\mathbf{S}^{-1}\right)^{1/2}, \tag{38}$$

where $\mathbf{S}$ denotes the sample covariance matrix of the standardized data. The eigenvalues λ and eigenvectors $\boldsymbol{\omega}$ of matrix $\mathbf{S}^*$ should be computed before the matrix of factor loadings is given

$$\mathbf{W} = \left(\mathrm{diag}\,\mathbf{S}^{-1}\right)^{-1/2}\mathbf{\Omega}\left(\mathbf{\Gamma} - \theta\,\mathbf{I}\right)^{-1/2}\mathbf{U}, \tag{39}$$

where $\mathbf{\Gamma}$ is the diagonal matrix of the first M number of sorted eigenvalues, $\mathbf{\Omega}$ is a matrix of the first M number of eigenvectors, $\mathbf{U}$ is an arbitrary M-by-M orthogonal matrix. Parameter θ in Eq. (39) specifies the smallest number of factors when

$$\theta = \frac{1}{L - M}\left(\lambda_{M+1} + \lambda_{M+2} + \ldots + \lambda_L\right) < 1 . \tag{40}$$

Assuming that $\mathbf{W}$ and $\mathbf{\Psi}$ are known quantities, the well logs of factor scores can be extracted by the maximization of the following log-likelihood function

$$\lg\left(P\right) = -\frac{1}{2}\left[\lg\left|2\pi\,\mathbf{\Psi}\right| + \left(\mathbf{D}' - \mathbf{F}\mathbf{W}^{\mathrm{T}}\right)\mathbf{\Psi}^{-1}\left(\mathbf{D}' - \mathbf{F}\mathbf{W}^{\mathrm{T}}\right)^{\mathrm{T}}\right] = \max . \tag{41}$$

After solving Eq. (40) an unbiased estimate to factor scores can be given by the hypothesis of linearity [25]

$$\mathbf{F}^{\mathrm{T}} = \left(\mathbf{W}^{\mathrm{T}}\mathbf{\Psi}^{-1}\mathbf{W}\right)^{-1}\mathbf{W}^{\mathrm{T}}\mathbf{\Psi}^{-1}\mathbf{D}'^{\mathrm{T}} . \tag{42}$$

The optimal number of factors can be set by statistical tests [26] or by Eq. (40) according to Jöreskog's approximate algorithm. The resultant factors are usually rotated for an easier interpretation. Since factor loadings are defined non-uniquely, an orthogonal transformation $\mathbf{W}\mathbf{W}^{\mathrm{T}} = \mathbf{W}^{*}\mathbf{W}^{*\mathrm{T}}$ can be applied to factor loadings, where $\mathbf{W}^{*} = \mathbf{W}\mathbf{V}$ holds for a suitably chosen M-by-M orthogonal matrix $\mathbf{V}$. In this study, the *Varimax* algorithm suggested by KAISER [27] is used to generate rotated factors, which are directly compared to petrophysical/hydraulic parameters of aquifers by regression analysis. Singular value decomposition of matrix $\mathbf{R}^{*}$ gives the proportions of total variance explained by each factor

$$\mathbf{R}^{*} = \mathbf{U}\mathbf{S}_e\mathbf{V}^{\mathrm{T}} , \tag{43}$$

where $\mathbf{U}$ and $\mathbf{V}$ are L-by-L orthogonal matrices and $\mathbf{S}_e$ is a diagonal matrix including the positive singular values sorted in descending order. The proportions of singular values estimate the information represented by the factors. The total variance is given by the trace of matrix $\mathbf{S}_e$. The variance explained by the q-th factor is

$$\sigma_q = \frac{S_{e,qq}}{\mathrm{tr}(\mathbf{S}_e)} \cdot 100\ (\%) . \tag{44}$$

Regression tests are performed to reveal the relations between the factors and petrophysical/hydraulic properties of rocks. In this study, the shale volume and hydraulic conductivity are estimated from the simultaneous processing of well logs. Earlier studies showed an exponential relation between the first factor and shale volume [8]. Hydraulic conductivity as a related quantity also correlates with the same statistical variable. In Section 6, the factor-analysis-based interpretation procedure is tested and compared to the Csókás method in the Baktalórántháza well site. The quality of interpretation results are checked by numerical analysis. The degree of linear dependence between the petrophysical variables is measured by the classical Pearson's correlation coefficient (R). In case of nonlinear relationships, the Spearman's (rank) correlation coefficient characterizes better the strength of correlation

$$\rho = 1 - \frac{6\sum_{n=1}^{N}\delta_n^2}{N(N^2-1)} \,, \qquad (45)$$

where δ_n is the difference between the ranks of the n-th data of the two investigated variables. When the value of the above correlation coefficient is nearly unity, a close (nonlinear) relationship is shown. The deviation between the well logs of petrophysical or hydraulic quantities is characterized by the Root-Mean-Square (*RMS*) error normalized to the relevant parameter.

6. Case study

The Csókás method and factor analysis are tested in Well Baktalórántháza-1 located in Szabolcs-Szatmár-Bereg County, in North-East Hungary. The original aim of the ground geophysical surveys was the investigation of the geological structure for the purpose of hydrocarbon exploration. Although neither oil nor gas was found, the borehole was still suitable for producing thermal water. According to VES soundings the upper 80–100 m part of the well was drilled through Pleistocene sediments dominated by sands, in which only the variation of grain sizes was detectable by geoelectric methods. A well-logging survey was conducted to provide information on deeper formations. Borehole logs indicated that the underlying rock of the sandy sequence was shale. Between 100–160 m sands had been deposited, followed by a shaly formation, and 5–15 m thick coarse-grained beds could be found. The boundary of Pleistocene and Pannonian periods was detected at the depth of 240 m. The Pannonian shaly complex mainly consisted of clayey sand, some gravel, clayey silt, clayey marl and bituminous clay.

The investigated interval for the present study is between 124.6–470.5 m, where the pore spaces of sediments are fully saturated with freshwater. The available well logs, such as natural gamma-ray intensity (*GR*), spontaneous potential (*SP*), gamma-gamma (*GG*) and neutron-neutron (*NN*) and shallow resistivity (*RS*) are plotted in Figure 1. Grain-size data are also provided from laboratory measurements made on 176 core specimens. The values of d_{10} and d_{60} are derived from the grain-size distribution curves, while the dominant grain size (d) was calculated from Eq. (3). The grain size data are plotted with dots in the last two tracks in Figure 1.

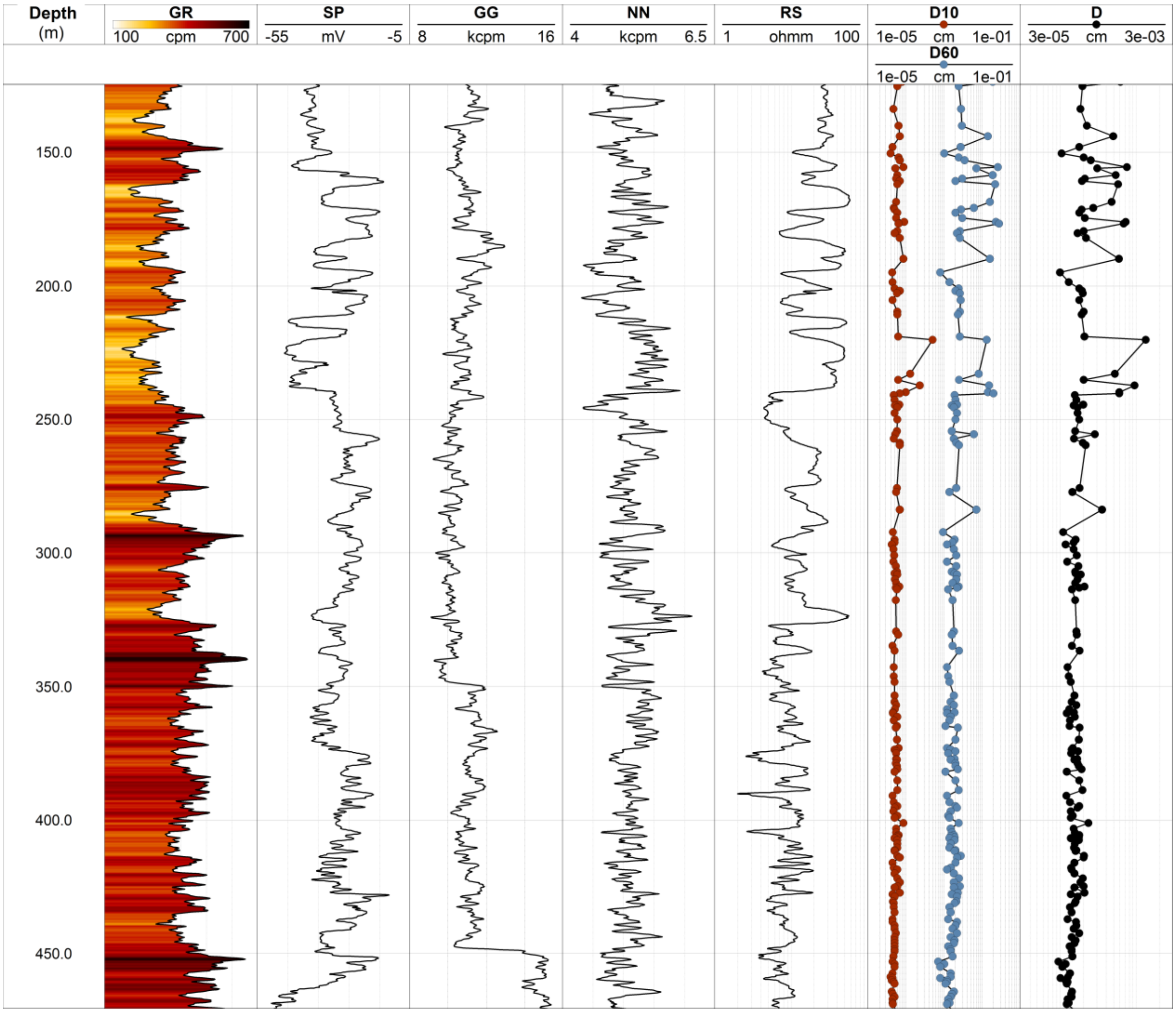

Figure 1

Observed well logs and grain-size data in Baktalórántháza-1

First, the result of factor analysis is presented. The factor loadings and scores are estimated by the solution of Eqs. (37) and (42). Singular value decomposition of the data covariance matrix shows that the total variance of input data can be explained by two lithological factors. The first factor is responsible for 82 % of data variance, while the second factor explains 18 % of observed information. The greatest loads go to the resistivity and natural gamma-ray logs: –0.77 (*RS*), 0.80 (*GR*). The average correlation between the measured variables is weak (*R*=0.16), where the highest coefficient is indicated between the resistivity and natural gamma-ray log (*R*=0.62). The information of these two well logs is culminated mainly in the first factor. For comparing the results of different interpretation methods, the scores of the first factor are scaled into the interval of 0 and 1. Regression tests show that the exponential relation between the first factor and shale volume suggested earlier by SZABÓ [8] is valid in the area of Baktalórántháza (Figure 2). The general formula between the first scaled factor (F_1') and shale volume calculated by Eq. (5) is

$$V_{sh} = \alpha e^{\beta F_1'} + \gamma \,, \tag{46}$$

where the local regression coefficients are given with 95 % confidence bounds: α=0.0412 [0.0373, 0.0452], β=3.204 [3.102, 3.306], γ=0.0285 [0.0196, 0.0374]. The rank correlation coefficient (ρ=0.95) shows a strong correlation between the statistical variable and the shale content (the linear correlation is also strong at R=0.88).

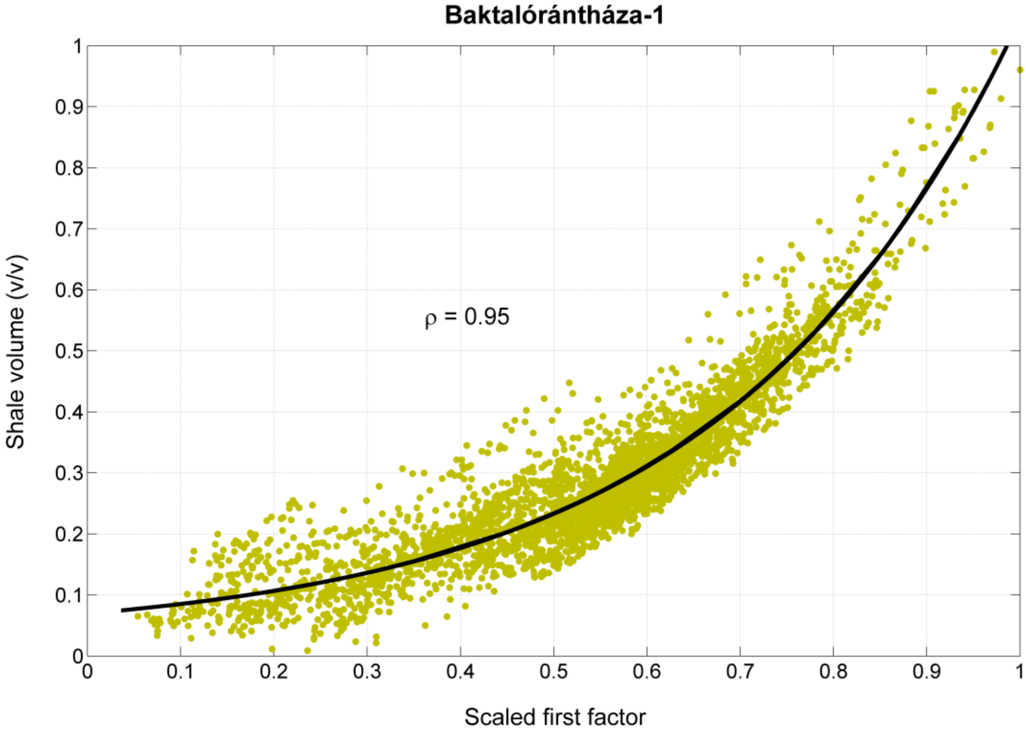

Figure 2

Regression relation between the first factor and shale volume in Baktalórántháza-1

Hydraulic conductivity as a related quantity to shale volume can be directly derived by factor analysis of well-logging data. The first statistical factor can be compared to the values of hydraulic conductivity calculated by the Kozeny-Carman model using dominant grain-sizes measured on the core samples and porosity by Eq. (2). The neutron-neutron readings are used to calculate the total porosity by using Eq. (9), where the constants of the response function of the neutron probe can be read from the *NN–GG* crossplot (NN_{sh}=4 kcpm, NN_{sd}=7.5 kcpm, NN_{f}=1 kcpm). The general regression model is

$$\lg(K) = c_1 F_1' + c_2 , \tag{47}$$

where K is the dimensionless hydraulic conductivity ($K=k/k_0$, where k_0=1 cm/s), c_1 and c_2 are site specific constants estimated with 95 % level of significance as c_1=–4.353 [–4.956, –3.75], c_2=–3.46 [–3.795, –3.125]. The Pearson's correlation coefficient (R=–0.74) shows a strong inverse proportionality between the first scaled factor and the decimal logarithm of hydraulic conductivity (Figure 3).

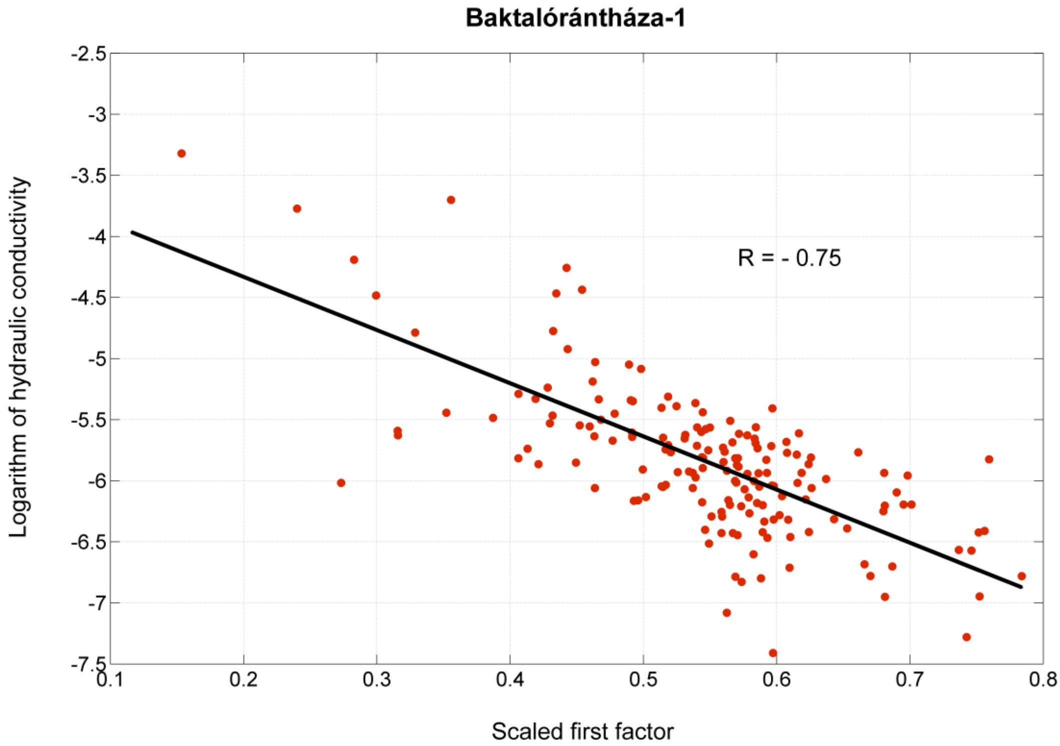

Figure 3
Regression relation between the first factor and hydraulic conductivity in Baktalórántháza-1

The use of the Csókás method requires the preliminary knowledge of effective porosity and the formation factor. The effective porosity and sand volume can be calculated from neutron (total) porosity and shale volume by Eq. (12). The amount of shaliness can be calculated by factor analysis or independently by using Eq. (5), where the natural gamma-ray intensities of sands and shales are derived from the deflections of the gamma-ray log (GR_{min}=188 cpm, GR_{max}=685 cpm). The formation factor is normally derived from *SP* and R_0 logs. In the lack of some related quantities, the Humble formula is used to calculate the formation factor ($F=0.62/\Phi^{2.15}$). With the formation factor and porosity, the Csókás method based on Eq. (23) gives an estimate of hydraulic conductivity without having to use grain-size data along the entire length of the borehole. Equation (2) is applied to validate the resultant hydraulic conductivity log, where the dynamic viscosity is set to 0.019 Pa·s and the acceleration of gravity is 981 cm/s^2. Since the Kozeny-Carman model cannot discard the measurement of grain sizes, the hydraulic conductivity is estimated only to that depth level where rock samples have been previously taken from the borehole. The numerical results show that the result of the Csókás method is in close agreement with that of the Kozeny-Carman procedure, because the *RMS* error is 3.4% averaged for the places of recovered rock samples. The hydraulic conductivity estimated by multivariate factor analysis fits acceptably to that of the Csókás procedure, where the *RMS* is 8.8% for the processed interval. The estimation results of the Csókás’ formula and factor analysis are represented in Figure 4. The critical velocity of water flow (VC_CS) can be derived from the hydraulic conductivity log (k_CS) by Eq. (24), moreover an estimate can be given for the specific surface of grains by Eq. (21). The latter is applicable to separate permeable and impervious intervals, i.e. aquifers appear at lower values of specific surfaces, while aquitards are represented by suddenly increased values of the interpolated well log of quantity S. The results of factor analysis are also included in the figure. Shale volume (VSH_FA) estimated by Eq. (46) is in track 4, which shows a close agreement with core data (VSH_CORE). The well log of VSH_LAR represents the shale volume based on the Larionov formula. Hydraulic conductivity estimated by factor analysis (k_FA) is validated by core derived hydraulic conductivity data (k_CORE). Both methods gives acceptable estimates, although perhaps factor analysis works better in the higher range of hydraulic conductivities (e.g. in the aquifer at the depth 225–240 m). The volumetric rock composition is illustrated in the last track in the figure for reference.

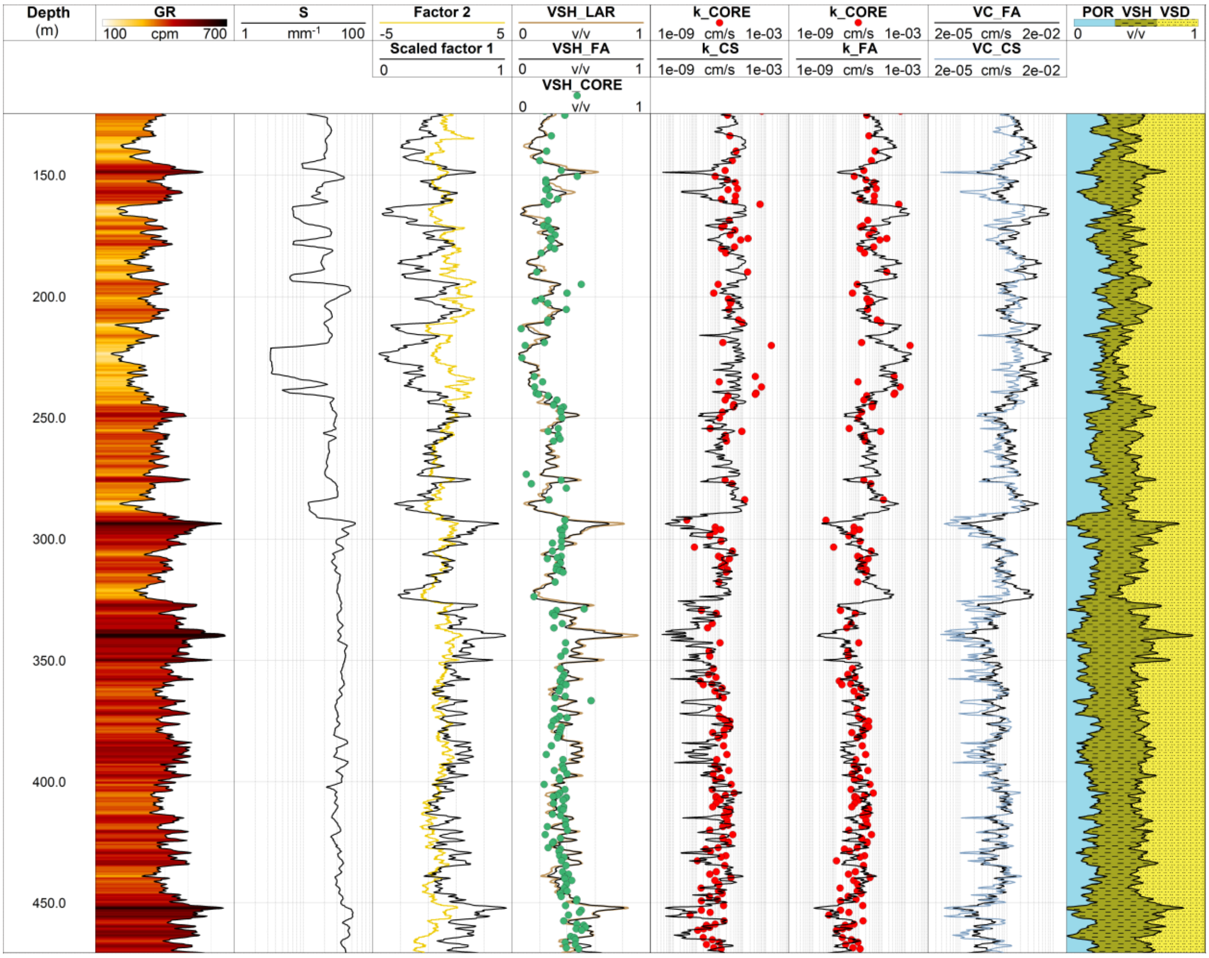

Figure 4
Well logs of estimated petrophysical and hydraulic parameters
in Baktalórántháza-1

A comparative study is made between the interpretation results of independent well log analyses. Figure 5 shows a strong linear correlation between the shale volumes estimated by the Larionov formula, factor and core analyses. The regression functions including the expected values of regression coefficients and their estimation errors are listed in Table 1. The correlation between the results of Larionov method and core analysis is somewhat weaker than that between results of factor and core analyses. The *RMS* error between the factor and core analyses also shows some 4 % relative decrease.

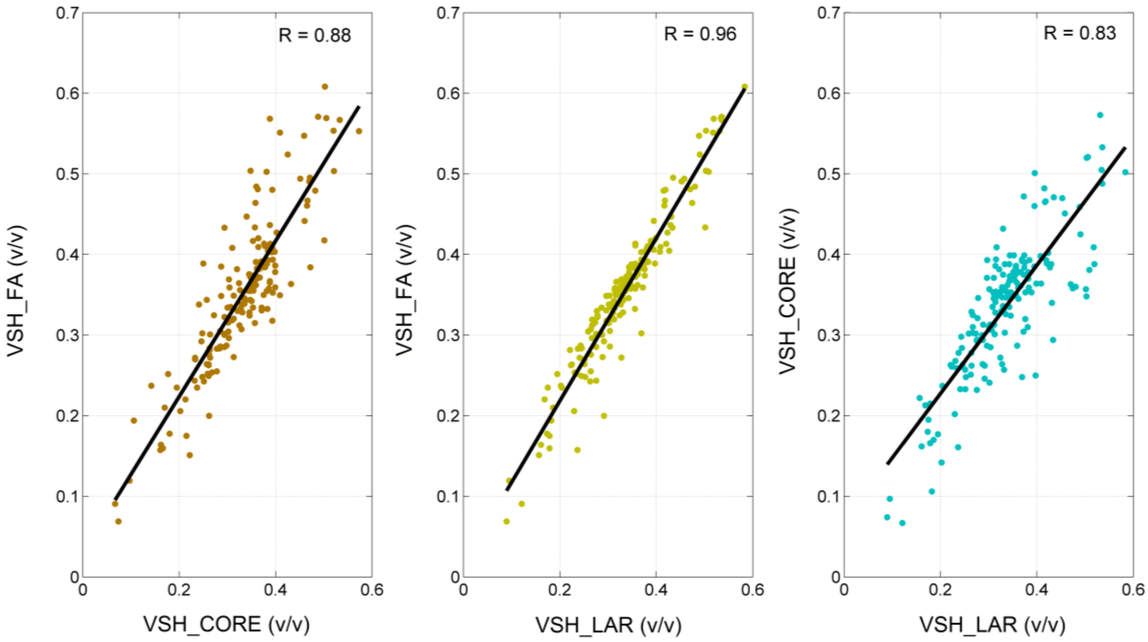

Figure 5
Regression relations between shale volumes estimated from different sources in Baktalórántháza-1

The Csókás formula-derived hydraulic conductivity log can also be compared to the result of factor analysis in the entire logging interval (Figure 6). The hydraulic conductivity estimated by multivariate factor analysis fits acceptably to that of the Csókás procedure, where the correlation coefficient is R=0.79 and the RMS error is 5.3 %. In the places of core sampling the correlation coefficient between the Csókás and Kozeny-Carman model-based hydraulic conductivity is R=0.75, while it is R=0.74 between the results of factor and core analyses. Even so, the RMS errors in Table 1 show that the factor analysis-derived hydraulic conductivities are closer to those of core measurements. This seems to be caused by the larger discrepancies between the Csókás and Kozeny-Carman methods compared to factor analysis and Kozeny-Carman procedure in the higher range of hydraulic conductivities.

The regression connections between different petrophysical/hydraulic quantities are examined below. It is found that there is a slight nonlinear relation between the shale volume and hydraulic conductivity, both estimated by factor analysis. The suggested regression function is

$$\lg\left(K^{(\mathrm{FA})}\right)=c_1^*\left(V_{sh}^{(\mathrm{FA})}\right)^{c_2^*}+c_3^*, \tag{48}$$

where the regression coefficients are obtained with 95 % level of confidence: c_1^*=–6.168 [–6.264, –6.072], c_2^*=–2.148 [–2.27, –2.026], c_3^*=0.4849 [0.467, 0.5028]. The rank correlation coefficient (ρ=–0.97) shows a strong correlation between the above variables. The strength of correlation between shale volume estimated by the Larionov formula and factor analysis-based hydraulic conductivity is also strong (ρ=–0.94); this can be approximated also by Eq. (48). The regression coefficients are modified a little as c_1^*=–6.339 [–6.556, –6.122], c_2^*=–1.696 [–1.947, –1.446], c_3^*=0.398 [0.3719, 0.424]. Both regression functions can be seen in Figure 7. Other linear regression relations between shale volume and hydraulic conductivity are listed in Table 1, where the independent estimation results agree with each other. In Figure 8, the specific surface calculated by Eq. (21) can be compared to porosity, and dimensionless hydraulic conductivity and critical velocity of flow, whose linear connections are quantified in Table 1.

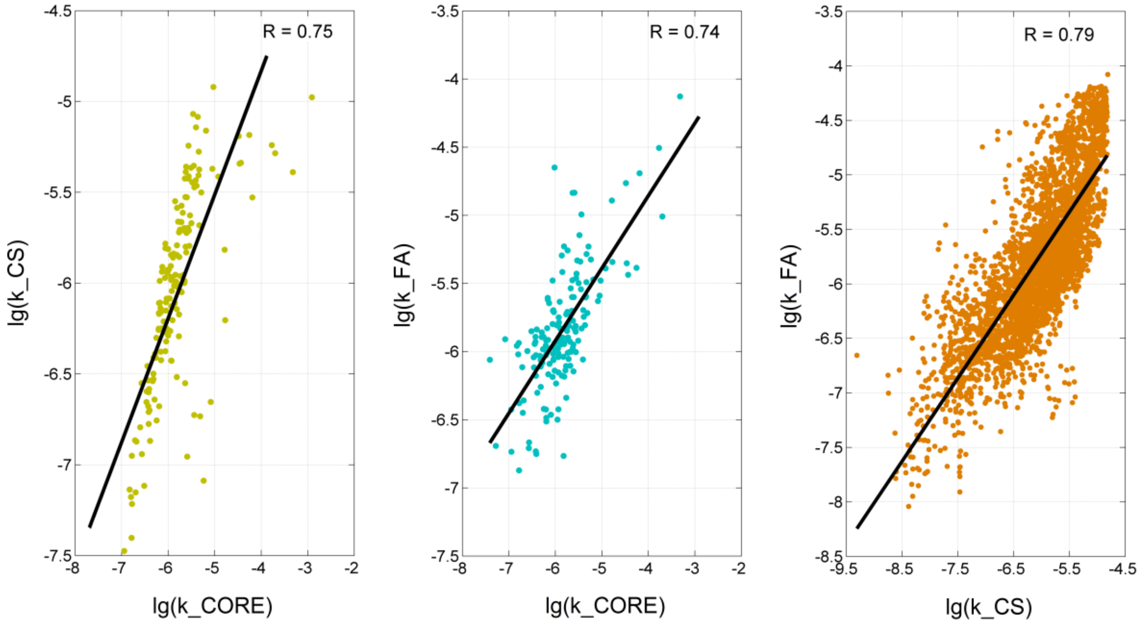

Figure 6

Regression relations between hydraulic conductivity estimation results in Baktalórántháza-1

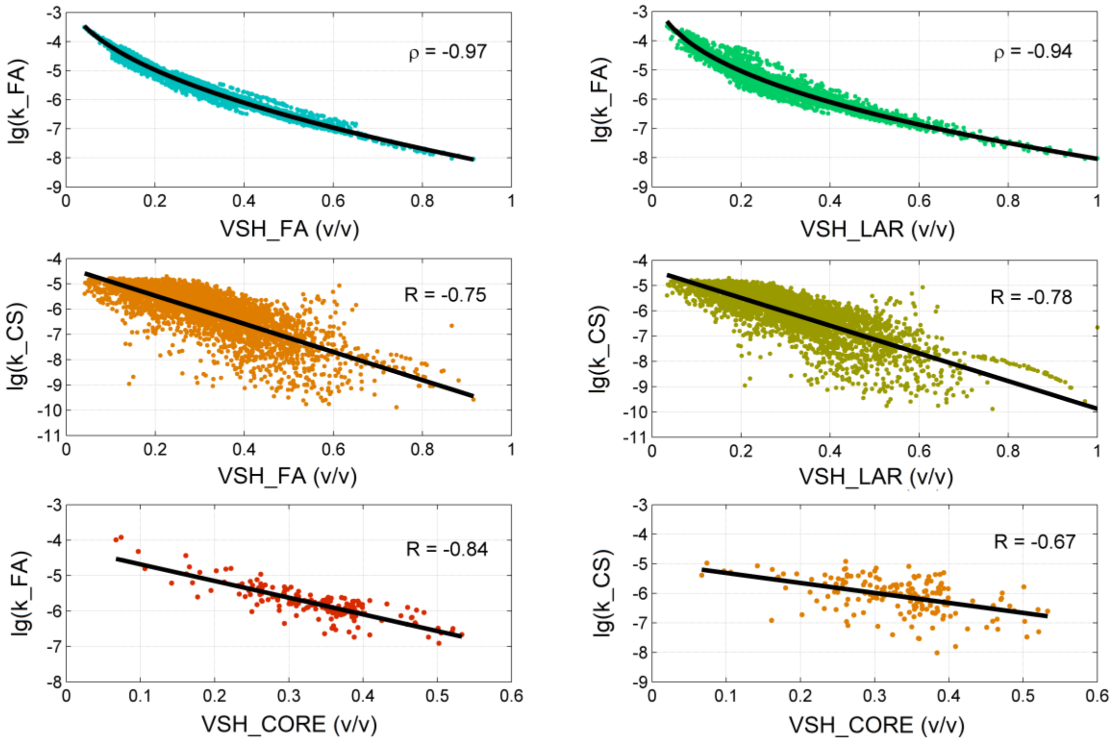

Figure 7

Regression relations between shale volume and hydraulic conductivity in Baktalórántháza-1

Table 1

Regression relations between petrophysical/hydraulic parameters of groundwater formations estimated from different methods in Baktalórántháza-1

Regression function	a_{min}	a	a_{max}	b_{min}	b	b_{max}	RMS	R
$V_{sh}^{(FA)} = aV_{sh}^{(CORE)} + b$	0.8875	0.9657	1.044	0.0038	0.0307	0.0577	4.89 %	0.88
$V_{sh}^{(FA)} = aV_{sh}^{(LAR)} + b$	0.9669	1.010	1.053	0.0018	0.0165	0.0313	3.25 %	0.96
$V_{sh}^{(CORE)} = aV_{sh}^{(LAR)} + b$	0.7177	0.7964	0.8752	0.0411	0.0683	0.0955	5.08 %	0.83
$\lg K^{(CS)} = a \lg K^{(CORE)} + b$	0.5886	0.6811	0.7735	-2.65	-2.107	-1.564	6.10 %	0.75
$\lg K^{(FA)} = a \lg K^{(CORE)} + b$	0.4587	0.5322	0.6057	-3.158	-2.726	-2.294	5.82 %	0.74
$\lg K^{(FA)} = a \lg K^{(CS)} + b$	0.7432	0.7632	0.7833	-1.26	-1.139	-1.018	5.31 %	0.79
$\lg K^{(CS)} = aV_{sh}^{(FA)} + b$	-5.733	-5.567	-5.401	-4.413	-4.355	-4.297	-	-0.75
$\lg K^{(CS)} = aV_{sh}^{(LAR)} + b$	-5.627	-5.482	-5.337	-4.443	-4.392	-4.341	-	-0.78
$\lg K^{(FA)} = aV_{sh}^{(CORE)} + b$	-5.122	-4.676	-4.229	-4.373	-4.219	-4.065	-	-0.84
$\lg K^{(CS)} = aV_{sh}^{(CORE)} + b$	-4.313	-3.371	-2.428	-5.297	-4.974	-4.651	-	-0.67
$V_{sh}^{(FA)} = a S^{(CS)} + b$	$8 \cdot 10^{-4}$	$9 \cdot 10^{-4}$	$9 \cdot 10^{-4}$	6.884	7.697	8.51	-	0.74
$\lg K^{(FA)} = a S^{(CS)} + b$	$-5 \cdot 10^{-5}$	$-5 \cdot 10^{-5}$	$-5 \cdot 10^{-5}$	-4.36	-4.313	-4.267	-	-0.73
$\Phi^{(NN)} = a S^{(CS)} + b$	$-4 \cdot 10^{-6}$	$-4 \cdot 10^{-6}$	$-4 \cdot 10^{-6}$	0.3063	0.3123	0.3183	-	-0.58
$\lg v_c^{(FA)} = a S^{(CS)} + b$	$-2 \cdot 10^{-5}$	$-2 \cdot 10^{-5}$	$-2 \cdot 10^{-5}$	-2.356	-2.333	-2.309	-	-0.73

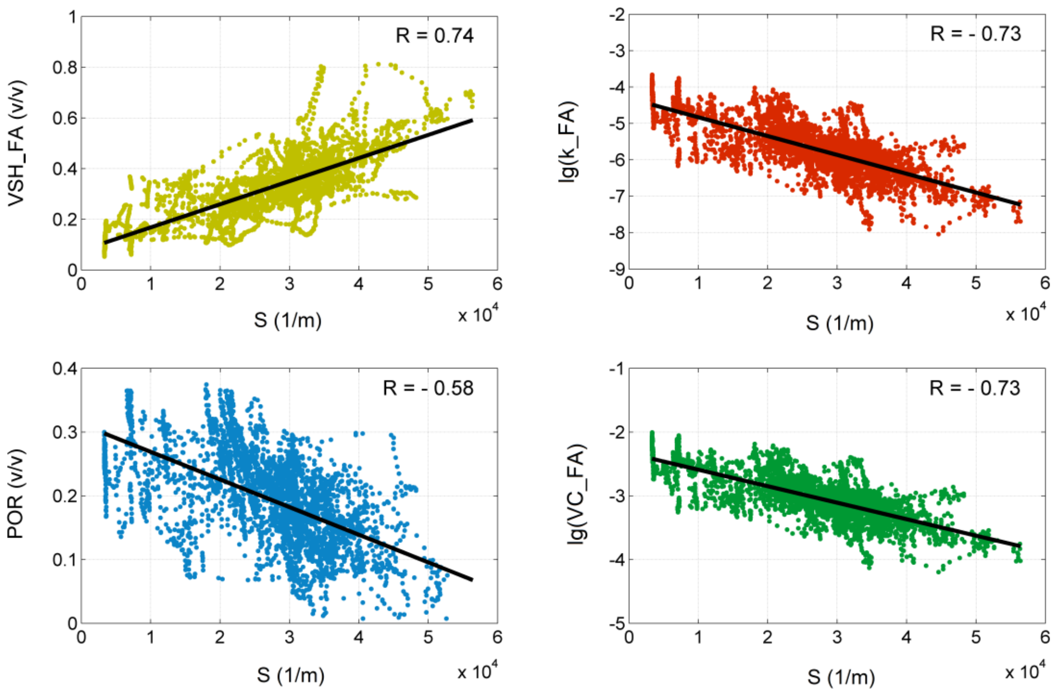

Figure 8

Regression relations between specific surface and petrophysical/hydraulic quantities in Baktalórántháza-1

7. Conclusions

The feasibility of the Csókás method and statistical factor analysis have been demonstrated in a shallow hydrogeological environment. Both methods using only well-logging information give a continuous (in-situ) estimate of petrophysical/hydraulic parameters in the form of well logs. The Csókás method is applicable to estimate hydraulic conductivity, critical velocity of flow, and specific surface of grains in typical unconsolidated aquifers, which are of high importance in extending the information of aquifer tests. The interpretation results are confirmed by the Kozeny-Carman procedure using core measurements. Factor analysis of the same well logs gives a reliable estimate of the amount of shaliness and hydraulic conductivity. The independent estimation for shaliness helps us to refine the lithology of formations and improve the aquifer storage model. Factor analysis uses all available well logs sensitive to lithology and water content (including caliper and temperature logs that cannot be used in inverse modeling without probe response functions), integrated into one statistical procedure to maximize the accuracy and reliability of the interpretation results. The petrophysical/hydrogeophysical information can be given also to a larger investigation area by using a special factor analysis algorithm extended to multidimensional model geometries.

The comparative regression study shows that the results of factor analysis are close to that of core analysis. Factor analysis uses five types of well logs and larger statistical sample in a joint statistical procedure for the estimation of hydraulic conductivity in Baktalórántháza-1. On the contrary, the Csókás method utilizes two types (i.e. resistivity and porosity logs) in estimating the relevant parameter. The Csókás method can give an estimate to a larger number of petrophysical/hydraulic parameters, which are connected to well-logging data deterministically. Factor analysis has been tested earlier in deep-seated (hydrocarbon-bearing structures), while the Csókás formula was developed specifically for shallow sediments. According to field experience, an optimal solution can be obtained with the Csókás procedure in medium or coarse-grained (well-sorted) unconsolidated sediments with a formation factor of less than 10. In the case of highly cemented aquifers, the estimation results show considerable deviations from the Kozeny-Carman model. Also in very fine-grained rocks, the hydraulic conductivities show sometimes a difference of more than one order of magnitude, which may require the revision of the Alger formula. Taking one thing with another, the independent estimation results agree with each other, which is proven by the regression connections revealed at the well site. The well-logging techniques presented in this paper is hereby recommended to the community of hydrogeophysicists; the techniques can be complemented with new types of measurements or advanced techniques used for the observation of input parameters to increase the efficiency of the evaluation of groundwater formations.

8. Acknowledgements

The author as the leading researcher of OTKA project No. PD-109408 thanks the support of the Hungarian Scientific Research Fund as well as to the support of the János Bolyai Research Fellowship of the Hungarian Academy of Sciences. The authors thank Professor Mihály Dobróka and László Hursán, honorary associate professor, for their valuable scientific advice, and László Bucsi Szabó of the Háromkő Deposit Company and Béla Latrán of Geokomplex Ltd. for professional consultation on well-site applications.

9. List of symbols

Symbol	**Description**	**Unit**
a	Tortuosity factor in Archie's equation	-
c_1, c_2	Regression coefficients for K vs. F'_1 relation	-

c_1^*, c_2^*, c_3^*	Regression coefficients for K vs. V_{sh} relation	-
C	Temperature coefficient for calculating SP response	-
C_d	Regression coefficient for d_{10} vs. F relation	-
C_t	Temperature coefficient of ratio g/υ	-
d, D	Dominant grain size	m
d_{10}, D10	Hazen's effective grain diameter	
d_{60}, D60	Grain diameter derived from the grain-size distribution curve at 60 % cumulative frequency	m
D'	Matrix of standardized well-logging data	-
E	Matrix of residuals in the decomposition of **D'**	-
F	Archie's formation factor	-
F	Matrix of factor scores	-
F_1'	First (scaled) factor	-
g	Normal acceleration of gravity	cm/s^2
GG	Gamma-gamma log reading	cps
GR	Natural gamma-ray intensity log reading	cps
GR_{sd}, GR_{min}	Natural gamma-ray intensity of sand	cps
GR_{sh}, GR_{max}	Natural gamma-ray intensity of shale	cps
h_0	Length of filter pack	m
h	Communality in the main diagonal of $\mathbf{R}^*$	-
$\mathbf{H}^2$	Matrix of communalities	-
I	Identity matrix	-
k	Hydraulic conductivity	m/s
K	Dimensionless hydraulic conductivity	-
L	Number of applied well log types	-
m	Cementation exponent in Archie's equation	-
M	Number of extracted statistical factors	-
N	Number of measuring points along a borehole	-
NN	Neutron-neutron intensity log reading	cpm
NN_f	Neutron-neutron intensity of pore-fluid	cpm
NN_{sh}	Neutron-neutron intensity of shale	cpm
NN_{sd}	Neutron-neutron intensity of sand	cpm
p	Pore pressure	Pa
P	Likelihood function used for the estimation of **F**	-
Q_{max}	Maximal theoretical water discharge	m^3/s
r_0	Radius of filter pack	m
R	Pearson's correlation coefficient	-
R	Correlation matrix of well-logging data	-
$\mathbf{R}^*$	Reduced correlation matrix of well-logging data	-
RMS	Root-mean-square error	%
S	Specific surface of rock grains	1/m
S	Sample covariance matrix of standardized data	-
S_w	Water saturation in the undisturbed formation	v/v
SP	Spontaneous potential log reading	mV
SP_{sh}	Spontaneous potential of shale	mV
t	Time	s
T	Temperature	$^{\circ}C$
R_d, RD	Deep resistivity log reading	ohmm
R_{mf}	Resistivity of mud filtrate	ohmm
R_s, RS	Shallow resistivity log reading	ohmm

R_w	Resistivity of pore water	ohmm
R_0	Resistivity of rock fully saturated with water	ohmm
$\mathbf{S}_e$	Matrix of singular values	-
$\mathbf{u}$	Relative displacement vector	m
U	Uniformity coefficient	-
$\mathbf{U}$, $\mathbf{V}$	Orthogonal matrices	-
v_c	Critical flow velocity	m/s
V_{sh}, VSH	Shale volume	v/v
V_{sd}, VSD	Sand volume	v/v
$\mathbf{W}$	Matrix of factor loadings	-
z	Measured depth along a borehole	m
$\mathbf{Z}$	Matrix of eigenvectors of $\mathbf{R}^*$	-
α_0	Shape factor of sample particles	-
α, β, γ	Regression coefficients for V_{sh} vs. F'_1 relation	-
$\mathbf{\Gamma}$	Matrix of sorted eigenvalues	-
θ	Constant for testing the number of statistical factors	-
$\mathbf{\Lambda}$	Matrix of eigenvalues (λ) of $\mathbf{R}^*$	-
κ	Intrinsic permeability	m^2
μ	Dynamic viscosity of water	Ns/m^2
υ	Kinematic viscosity of water	m^2/s
ρ	Spearman's rank correlation coefficient	-
ρ_b, DEN	Bulk density log reading	gcm^{-3}
ρ_{mf}	Density of mud filtrate	gcm^{-3}
ρ_{sd}	Density of sand	gcm^{-3}
ρ_{sh}	Density of shale	gcm^{-3}
ρ_w	Density of pore water	gcm^{-3}
σ	Data variance explained by a statistical factor	-
Φ, POR	Formation porosity	v/v
$\mathbf{\Psi}$	Matrix of specific variances	-
Ω	Objective function used for estimating $\mathbf{W}$	-
$\mathbf{\Omega}$	Matrix of eigenvectors	-

10. References

[1] SERRA, O.: *Fundamentals of Well-log Interpretation*. Elsevier, Amsterdam, 1984.

[2] TSELENTIS, G. A.: The processing of geophysical well logs by microcomputers as applied to the solution of hydrogeological problems. *Journal of Hydrology*, 1985, 80, 215-236.

[3] RUBIN, Y.-HUBBARD, S. S.: *Hydrogeophysics*. Water Science and Technology Library Series 50, Springer-Verlag, Dordrecht, Berlin, Heidelberg, New York, 2005.

[4] WALSH, D.-TURNER, P.-GRUNEWALD, E.-ZHANG, H.-BUTLER, J. J.-REBOULET, E.-KNOBBE, S.-CHRISTY, T.-LANE, J. W.-JOHNSON, C. D.-MUNDAY, T.-FITZPATRICK A.: A small-diameter NMR logging tool for groundwater investigations. *Groundwater*, 2013, 51, 914-926.

[5] CSÓKÁS, J.: Determination of yield and water quality of aquifers based on geophysical well logs (in Hungarian). *Magyar Geofizika*, 1995, 35(4), 176-203.

[6] ALGER, R. P.: Interpretation of electric logs in fresh water wells in unconsolidated formation. *SPE Reprint Series 1*, 1971, 255.

[7] LAWLEY, D. N.-MAXWELL, A. E.: Factor analysis as a statistical method. *The Statistician*, 1962, 12, 209-229.
[8] SZABÓ, N. P.: Shale volume estimation based on the factor analysis of well-logging data. *Acta Geophysica*, 2011, 59, 5, 935-953.
[9] BEAR, J.: *Dynamics of Fluids in Porous Media*. Dover Publications Inc., New York, 1972.
[10] JUHÁSZ, J.: *Hydrogeology* (in Hungarian). Akadémiai Kiadó, Budapest, 2002.
[11] LARIONOV, V. V.: *Radiometry of Boreholes* (in Russian). Nedra, Moscow, 1969.
[12] ALBERTY, M.-HASHMY, K.: Application of ULTRA to log analysis. *SPWLA Symposium Transactions*, 1984, Paper Z, 1-17.
[13] TIMUR, A.: An investigation of permeability, porosity, and residual water saturation relationships. *SPWLA 9th Annual Logging Symposium*, 1968, SPWLA-1968-J.
[14] DRAHOS, D.: Inversion of engineering geophysical penetration sounding logs measured along a profile. *Acta Geodetica et Geophysica*, 2005, 40, 193-202.
[15] SZABÓ, N P-DOBRÓKA, M.: Float-encoded genetic algorithm used for the inversion processing of well-logging data. In: Michalski, A. (ed.) *Global optimization: Theory, developments and applications*. Mathematics Research Developments, Computational Mathematics and Analysis Series, Nova Science Publishers Inc, New York, 2013.
[16] KOVÁCS, GY.: *Hydraulics of Filtration* (in Hungarian). Akadémiai Kiadó, Budapest, 1972.
[17] ARCHIE, G. E.: The electrical resistivity log as an aid in determining some reservoir characteristics. *SPE, Transactions of the AIME*, 1942, 146(1), 54-62.
[18] OGBE, D.-BASSIOUNI, Z.: Estimation of aquifer permeabilities from electric well logs. *The Log Analyst*, 1978, 19(5), 21-27.
[19] PIRSON, S. J.: *Handbook of Well Log Analysis*. Prentice-Hall Inc., New York, 1963.
[20] GÁLFI, J.-LIEBE, P.: The permeability coefficient in clastic water bearing rocks (in Hungarian). *Vízügyi Közlemények*, 1981, 63(3), 437-448.
[21] SCHMIEDER, A.: *Water-Risk and Water Economy in Mining* (in Hungarian). Műszaki Könyvkiadó, Budapest, 1975.
[22] KASSAI, F.-JAMBRIK, R.: *Water Mining II.* (in Hungarian). Tankönyvkiadó, Budapest, 1986.
[23] JÖRESKOG, K. G.: Factor analysis and its extensions. In: Cudeck, R.-MacCallum, R. C. (eds) *Factor analysis at 100, Historical developments and future directions.* Lawrence Erlbaum Associates, New Jersey, 2007.
[24] JÖRESKOG, K G: A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 1969, 34(2), 183-202.
[25] BARTLETT, M S: Factor analysis in psychology as a statistician sees it. *Nordisk Psykologi's Monograph Series 3*, 1953, Almqvist & Wiksell, Uppsala, 23-34.
[26] BARTLETT, M S: Tests of significance in factor analysis. *British Journal of Psychology*, 1950, 3(2), 77-85.
[27] KAISER, H F: The varimax criterion for analytical rotation in factor analysis. *Psychometrika*, 1958, 23, 187-200.