



Permeability extraction from multiple well logs using particle swarm optimization based factor analysis

Norbert Péter Szabó^{1,2} · Armand Abordán^{1,2} · Mihály Dobróka¹

Received: 1 January 2022 / Accepted: 26 May 2022 / Published online: 25 June 2022
© The Author(s) 2022

Abstract

In this paper, we present an innovative factor analysis algorithm for hydrocarbon exploration to estimate the intrinsic permeability of reservoir rocks from well logs. Unlike conventional evaluation methods that employ a single or a limited number of data types, we process simultaneously all available data to derive the first statistical factor and relate it to permeability by regression analysis. For solving the problem of factor analysis, we introduce an improved particle swarm optimization method, which searches for the global minimum of the distance between the observed and calculated data and gives a quick estimation for the factor scores. The learning factors of the intelligent computational technique such as the cognitive and social constants are specified as hyperparameters and calculated by using simulated annealing algorithm as heuristic hyperparameter estimator. Instead of the arbitrary fixation of these hyperparameters, we refine them in an iterative process to give reliable estimation both for the statistical factors and formation permeability. The estimated learning parameters are consistent with literature recommendations. We demonstrate the feasibility of the proposed well-log analysis method by a Hungarian oilfield study involving open-hole wireline logs and core data. We determine the spatial distribution of permeability both along a borehole and between more wells using the factor analysis approach, which serves as efficient and reliable multivariate statistical tool for advanced formation evaluation and reservoir modeling.

Keywords Factor analysis · Particle swarm optimization · Hyperparameter · Intrinsic permeability · Hydrocarbon reservoir

Mathematics Subject Classification 62H12 · 62H25 · 86A99 · 68T20 · 62J02

✉ Norbert Péter Szabó
gfnmail@uni-miskolc.hu

¹ Department of Geophysics, University of Miskolc, Miskolc-Egyetemváros 3535, Hungary

² MTA-ME Geoen지니어ing Research Group, University of Miskolc, Miskolc-Egyetemváros 3535, Hungary

1 Introduction

Wireline logs including abundant in situ measurement information taken from boreholes can be effectively used to derive basic petrophysical parameters of hydrocarbon reservoirs (Serra 1984). Most of these formation properties (e.g., porosity, shale volume, water saturation etc.) are usually estimated in reliable quality-checked inversion procedures (Ball et al. 1987; Alberty and Hashmy 1984; Dobróka et al. 2016). However, permeability of rock formations cannot be directly estimated by the inversion of conventional well logs, since the available tool response equations do not include it explicitly. Therefore, it is generally estimated empirically by using some types of well logs (Wyllie and Rose 1950; Timur 1968; Coates and Dumanoir 1974) and these estimations often need to be calibrated by core data (Singh 2018). Due to different methods and conditions, the estimation error of permeability is rather large compared to the other petrophysical parameters (e.g., porosity and shale volume) that are required for evaluating the productivity of hydrocarbon reservoirs. Exploratory data analysis methods including regression analysis, dimension reduction and pattern recognition techniques are widely used in geosciences for different purposes. Tang and White (2008) made a comparative study on several statistical tools for facies classification. Jafarpour and McLaughlin (2009) applied a discrete cosine transform for reservoir characterization. Soares et al. (2020) used dictionary learning for the sparse representation of 4D seismic data to highlight only the main features of the dataset. Cosultchi et al. (2012) utilized principal component analysis and cluster analysis for aquifer identification in petroleum reservoirs. Generally, some linear regression analysis approach is used for relating the permeability to existing petrophysical parameters (Dahraj and Bhutto 2014; Handhal 2016). Multivariate statistical tools have been applied for permeability estimation in AI-Mudhafar (2020).

We present a new alternative to predict permeability of hydrocarbon reservoirs, which is based on a metaheuristic approach for factor analysis. The simultaneous processing of well logs allows not just to reduce the dimensionality of the data space, but to find correlation between the extracted factor variables and reservoir parameters. As one of the most important parameters, we estimate the intrinsic permeability utilizing all available well logs leading to more reliable estimates. Szabó and Dobróka (2018) suggested a statistical approach for the processing of well logs, in which a differential evolution algorithm based factor analysis was introduced for shale volume estimation. For solving the problem of factor analysis, we currently apply the particle swarm optimization (PSO) method proposed by Kennedy and Eberhart (1995), which is less computationally expensive and requires less parametrization than genetic algorithm (Rahmat-Samii 2003; Donelli et al. 2006). This metaheuristic technique is based on swarm intelligence (Del Ser 2019), which is widely used in engineering applications. For instance, Nguyen et al. (2020) utilized it for feature selection in data mining. Self et al. (2016) suggested PSO for selecting optimal drilling parameters to minimize the overall cost. Shaw and Srivastava (2007) evaluated the applicability of PSO to inversion of direct current, induced polarization and magnetotelluric data and concluded that the results are consistent with ridge regression and genetic algorithm based inversions. Essa and Elhoussein (2018) presented a geophysical application, in which magnetic anomalies were interpreted successfully with the help of this metaheuristic.

Factor analysis is a multivariate statistical tool that is generally used to decrease the dimensionality of multi-parameter problems, while keeping most of the information contained in the original variables to simplify the data interpretation. The newly derived factors are necessarily uncorrelated and may reveal hidden information from the measurements (Lawley and Maxwell 1962). Thus, for processing wireline logging data it can be effectively applied because the measured datasets are usually quite large due to several applied logging tools and numerous measured depth points. The extracted factors from wireline logs are referred to as factor logs that can be linked to petrophysical parameters through regression analysis (Szabó 2011). In this way, factor analysis processes all available well logs to give an estimation to these parameters, while for instance, simple deterministic methods normally rely only on one log type to derive one parameter at a time. Since it is an effective approximate method, it can also process data for which an exact mathematical relation with petrophysical properties is not available (e.g., the caliper log), however they may contain useful lithological or physical information.

We show a two-level embedded computing algorithm for improving the performance of factor analysis. In the inner loop, the PSO algorithm runs to estimate the factor scores while the factor loadings are initially derived from the non-iterative approach of Jöreskog (2007). In the outer loop, the hyperparameters of the same PSO procedure called the cognitive and social parameters are refined by a Metropolis criterion based simulated annealing method as global optimization tool avoiding the possible localities of the objective function (Metropolis et al. 1953). These control parameters are important learning factors, which are normally chosen as arbitrary constants at the beginning of the PSO procedure. In this study, we suggest a much more intelligent manner as estimating them automatically. This approach was first proposed for the interpretation of direct push logs to estimate the volumetric soil moisture content of the shallow unsaturated zone (Abordán and Szabó 2021) as a further improved factor analysis technique used in Szabó et al. (2012). In this study, the factor scores are cross-plotted with reservoir properties. We assume the extracted factors hold quantitative information on the permeability of deep reservoir rocks. We extend the factor analysis algorithm to process multiwell logging data. Our permeability prediction results are shown involving single and multi-borehole data collected from a Hungarian gas field.

2 Methods

2.1 Traditional method of factor analysis

To initialize factor analysis, we build a common data matrix (\mathbf{D}) from the standardized data of S number of measured well logs, in which every column represents a dataset recorded by different logging tools, while N number of rows contain all measured depths along the processed interval

$$\mathbf{D} = \begin{pmatrix} D_{11} & D_{12} & \cdots & D_{1S} \\ D_{21} & D_{22} & \cdots & D_{2S} \\ \vdots & \vdots & \vdots & \vdots \\ D_{i1} & D_{i2} & \cdots & D_{iS} \\ \vdots & \vdots & \vdots & \vdots \\ D_{N1} & D_{N2} & \cdots & D_{NS} \end{pmatrix}, \quad (1)$$

Factor analysis is based on the following model

$$\mathbf{D} = \mathbf{F}\mathbf{L}^T + \mathbf{E}, \quad (2)$$

where \mathbf{F} is the factor scores matrix of size N -by- R (R is the number of derived factors). Quantity \mathbf{L} is the factor loadings matrix of size S -by- R and \mathbf{E} is the residual error matrix of size N -by- S (symbol T is the matrix transpose). According to Eq. (2), the observed wireline logs are expressed as a linear combination of the factors. The factor loadings in matrix \mathbf{L} show the partial correlations between the observed and derived statistical variables. The largest part of total variance represented by matrix \mathbf{D} is explained by the first factor, the scores of which are found in the first column of matrix \mathbf{F} . Given that the factors are uncorrelated, we write the correlation matrix of standardized data as

$$\mathbf{R} = N^{-1}\mathbf{D}^T\mathbf{D} = \mathbf{L}\mathbf{L}^T + \mathbf{\Psi}, \quad (3)$$

where $\mathbf{\Psi}$ denotes the diagonal matrix of error variances independent of those of the common factors in matrix \mathbf{F} . The calculation of correlation matrix \mathbf{R} is easy to handle if matrix \mathbf{E} is neglected in Eq. (2). This approximation helps to avoid the estimation of specific variances by cancelling matrix $\mathbf{\Psi}$ in Eq. (3). Since the part of the total variance that may not be described by the common factors is neglected, this approach leads to an approximate solution for the factors using the popular method of Principal Component Analysis (PCA). Although PCA does not account for residual matrix \mathbf{E} , but it gives a quick and unique (algebraic) solution for extracting the principal components (Maćkiewicz and Ratajczak 1993).

We approximate the factor loadings non-iteratively by a scaling-free estimation method (Jöreskog 2007)

$$\mathbf{L} = (\text{diag}\mathbf{\Sigma}^{-1})^{-1/2}\mathbf{\Omega}(\mathbf{\Gamma} - \theta\mathbf{I})^{1/2}\mathbf{U}, \quad (4)$$

where $\mathbf{\Gamma}$ denotes the matrix of the first R number of eigenvalues of the empirical covariance matrix $\mathbf{\Sigma}$, and the corresponding R number of eigenvectors are given in matrix $\mathbf{\Omega}$, and \mathbf{U} is a freely selected R -by- R orthogonal matrix. Quantity \mathbf{I} denotes the identity matrix and θ is a properly chosen constant that must be accepted smaller than 1. In traditional dimension reduction approaches, the factor scores are calculated by the maximum likelihood method. By assuming the hypothesis of linearity

$$\mathbf{F}^T = (\mathbf{L}^T \mathbf{\Psi}^{-1} \mathbf{L})^{-1} \mathbf{L}^T \mathbf{\Psi}^{-1} \mathbf{D}^T \tag{5}$$

gives an unbiased estimate to the factor scores (Bartlett 1937). Since Eq. (5) is analogous to a least squares solution weighted by the specific variances (Menke 2012), the localities of the objective function to be minimized cannot be effectively avoided by the gradient-based searching for the absolute minimum.

2.2 Particle swarm optimization based factor analysis

To improve the performance of data reduction, we suggest PSO as a global optimization procedure to determine the vector of factor scores. Abordán and Szabó (2018) established the basic principles of the PSO assisted factor analysis (FA-PSO). To give a metaheuristic estimation of the factors, we reformulate the model of factor analysis given in Eq. (2) as

$$\mathbf{d}^{(m)} = \tilde{\mathbf{L}}\mathbf{f} + \mathbf{e} \tag{6}$$

where $\mathbf{d}^{(m)}$ is the vector of standardized wireline logging data of size NS , $\tilde{\mathbf{L}}$ denotes the matrix of factor loadings of size NS -by- NR , \mathbf{f} is the expanded vector of factor scores of size RN and \mathbf{e} is the vector of approximation errors of size SN (Szabó and Dobróka 2017). As a first step, we determine matrix $\tilde{\mathbf{L}}$ by Eq. (4) and rotate it using the varimax technique for easier interpretation (Kaiser 1958). We give the optimal values of the factor scores \mathbf{f} by using PSO, which can be considered as an iterative inversion procedure. We use the following L_2 -norm based function as objective function

$$E = \left[\frac{1}{NS} \sum_{i=1}^{NS} (d_i^{(m)} - d_i^{(c)})^2 \right]^{1/2} \tag{7}$$

where $\mathbf{d}^{(c)}$ contains the standardized calculated wireline logging data. Here, calculated data refers to the term $\tilde{\mathbf{L}}\mathbf{f}$ from Eq. (6), which also gives the possibility for calculating theoretical well logs that can be regarded as the solution of the forward problem of the inversion procedure. Previous researches suggest that it can be even used to generate theoretical well logs for intervals where measurements are missing, due to the high correlation of certain log types to the first factor (Szabó et al. 2012). By the minimization of the objective function in Eq. (7), we update the factor scores by PSO, while we keep the factor loadings constant to reduce runtime. It must be mentioned that it is possible to simultaneously refine the factor scores and loadings using the well-known maximum likelihood method, or an iteratively reweighted factor analysis for seeking a more robust solution (Szabó and Dobróka 2017).

Particle swarm optimization originally introduced by Kennedy and Eberhart (1995) was developed to mimic the social behavior of animals. Generally, it was applied for its easy implementation and high efficiency in solving complex problems (Kumar et al. 2017). To find the optimal solution, PSO applies a swarm of particles generated randomly by following uniform distribution in a pre-defined search space. Given an

N -dimensional search space, vector $\mathbf{f}_i = (f_{i1}, f_{i2}, \dots, f_{iN})^T$ denotes the position of the i th particle. In our problem, quantity \mathbf{f}_i denotes the i th possible solution for the vector of factor scores. We store the particle's velocity in vector $\mathbf{v}_i = (v_{i1}, v_{i2}, \dots, v_{iN})^T$, which defines its movement in each iteration within the search domain. The swarm of particles seek the minimum of the objective function defined in Eq. (7) by utilizing the following equations

$$\mathbf{f}_i(t+1) = \mathbf{f}_i(t) + \mathbf{v}_i(t+1) \quad (8)$$

$$\mathbf{v}_i(t+1) = w\mathbf{v}_i(t) + r_1c_1(\mathbf{p}_i(t) - \mathbf{f}_i(t)) + r_2c_2(\mathbf{g}(t) - \mathbf{f}_i(t)) \quad (9)$$

where L denotes the number of particles in the swarm ($i = 1, 2, \dots, L$), t stands for iterations ($t = 1, \dots, t_{max}$) where t_{max} is the final iteration step. During the search for the global optimum, PSO utilizes a memory which is updated after each iteration. Every particle's best position is held in $\mathbf{p}_i = (p_{i1}, p_{i2}, \dots, p_{iN})^T$, while vector \mathbf{g} stores the best position found by the whole population. Kennedy and Eberhart (1995) suggested to set the cognitive and social constants c_1 and c_2 equal to 2 as default, separately, during the initialization. These control parameters define to what extent the personal best position of particles \mathbf{p}_i and the very best position of the swarm \mathbf{g} affect the movement of the i th particle during the search. In Eq. (9), r_1 and r_2 are uniformly distributed random parameters from 0 to 1, while w represents an inertia weight that was suggested by Shi and Eberhart (1998) to more effectively control the optimization procedure.

To start the FA-PSO procedure for finding the optimal vector of factor scores \mathbf{f} , we generate a population of particles with uniform distribution within the search space, which can be pre-defined by solving Eq. (5) for the factor scores. As for w , we use a chaotic descending inertia weight scheme to better control the velocity of particles (Feng et al. 2007). We perform it in three steps: first we generate a random number (z) from the range 0 to 1. Then we set z by logistic mapping according to $z' = 4z(1-z)$ and finally

$$w = (w_1 - w_2)(t_{max} - t)/t_{max} + w_2z' \quad (10)$$

where w_1 and w_2 represent the starting and the final values of inertia weight during the iterative procedure, respectively. Then in each iteration step, we re-determine the particles' positions according to Eqs. (8–9), and re-calculate the objective function defined in Eq. (7) with the new factor scores \mathbf{f} . We continuously update the particles until the maximum number of iterations is reached, thereafter we accept the particle with the lowest data misfit as the solution for the factor scores.

2.3 Permeability estimation by factor analysis

Previous studies show that the first statistical factor correlates well with the shale content of various geological formations (Szabó 2011; Szabó and Dobróka 2013; Puskarczyk et al. 2015; Puskarczyk 2020; Abordán and Szabó 2020). Advanced mathematical approaches have been developed to extract the shale volume from well logs

as efficient as possible. The iteratively reweighted factor analysis gives a reliable estimation to the factors independent of the statistical distribution of input data (Szabó and Dobróka 2017). In addition to give a robust solution, the calculation of factor scores was established on global optimization by using evolutionary computation (Szabó and Dobróka 2018). By analyzing multiple well logs, we take all the measured information into account during the interpretation, which makes a step forward in increasing the reliability of estimation compared to single depth-point approaches using one or few well logs. Under taking all measured information we mean that we select all suitable well logs as input being highly sensitive to shale volume, in other words only those data types are processed that are significantly influenced by the variation of the relevant petrophysical parameter along the borehole. Based on the above considerations, the first statistical factor extracted from multiple well logs can be interpreted as a lithology indicator, which holds valuable information on the shale content of sedimentary formations. When applying factor analysis to well logs, a strong exponential connection between the first factor (F_1) and shale volume (V_{sh}) can be set in geological formations of different ages. In the Hungarian and American examples given in Fig. 1, the consistence of the results is also confirmed by the validity of the same regression function (with approximately the same regression coefficients and by using proper scaling to the factor scores).

We assume a strong correlation between the absolute permeability and the first principal factor in sedimentary formations. This is partly due to the fact that permeability is to some extent inversely proportional to shale content of clastic formations (Schön and Georgi 2003; Revil and Cathles 1999). In our test site, an empirical relation is found between the decimal logarithm of permeability and shale volume both derived from wireline logs as well as core information (Fig. 2). The correlation coefficient for the first case is -0.91 , while for the latter it is -0.72 , both showing a clear relationship between the two quantities. In well log analysis, for estimating the absolute permeability (K) from wireline logging datasets, usually deterministic approaches are used which rely on porosity (Φ) and irreducible water saturation ($S_{w,irr}$) data. The most frequently applied manner is based on the use of similar empirical approaches like the formula of Timur (1968)

$$K = 0.136 \frac{\Phi^{4.4}}{S_{w,irr}^2} \quad (11)$$

Analogous studies show that the statistical factors include not only the information on lithological characteristics but also the hydraulic properties of reservoir rocks. For instance, Szabó (2015) explored the hydraulic conductivity of fractured groundwater formations successfully by factor analysis of hydrogeophysical logs. In this study, we test the proposed factor analysis approach to estimate permeability in six wells (Well 1–6) drilled in the Pannonian basin in East Hungary. We investigate unconsolidated upper Miocene gas-bearing formations with high average porosity showing a variability of different grain sizes and lithology as marl, silt and sands. Accordingly, we show that the well log of the first factor derived by the improved FA-PSO approach scaled into the range 0 to 1 can be directly used to estimate permeability along arbitrary

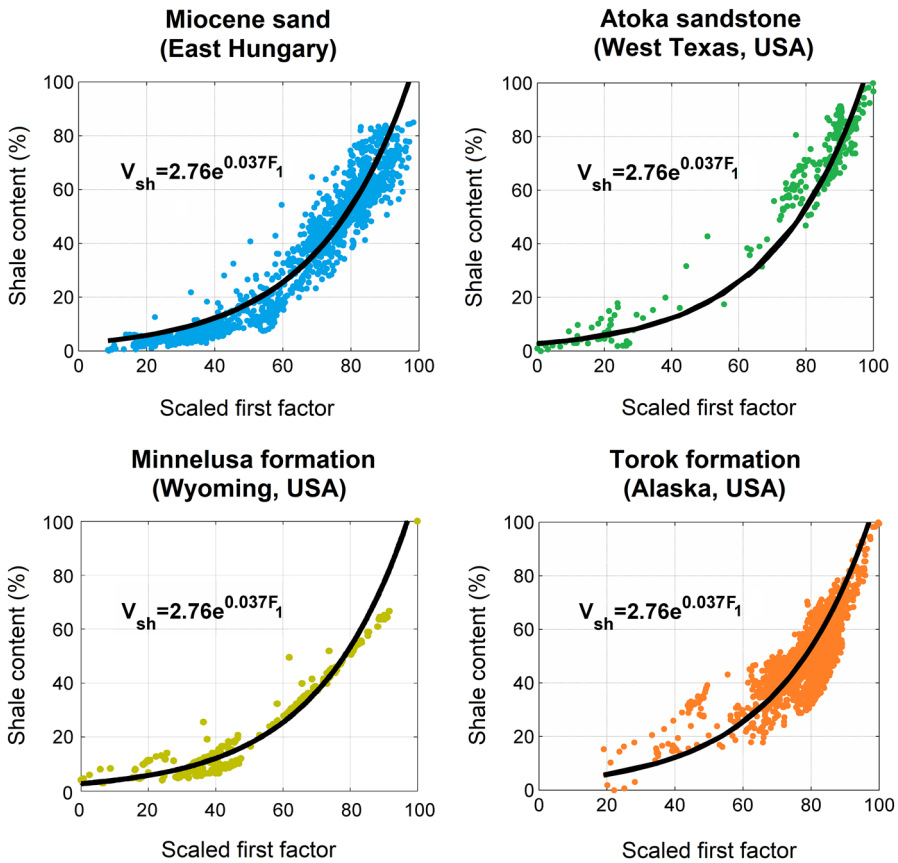


Fig. 1 Regression relationships between the first factor and shale volume derived by well log analysis for different hydrocarbon fields

intervals. For this purpose, we use a non-linear relationship, which connects the first factor to the intrinsic permeability of hydrocarbon reservoirs. In the given area, the first (properly scaled) factor extracted from the wireline logging dataset by FA-PSO is connected to the decimal logarithm of permeability (K) in the form as

$$\lg(K) = a(1 - F_1^b)^c + d \tag{12}$$

where a, b, c, d are area specific coefficients estimated by regression analysis using for instance core data and other preliminary information. The above non-linear relationship is applied to different wireline logging datasets in the investigated hydrocarbon field to demonstrate the feasibility of the proposed method.

Our FA-PSO method is a fully metaheuristic approach, therefore its output is highly affected by the control parameters (i.e., learning factors) that are worth to be set as optimal as possible at initialization. The optimal parameter selection of PSO has been studied by several researches (Jiang et al. 2007; Fernández Martínez et al. 2010;

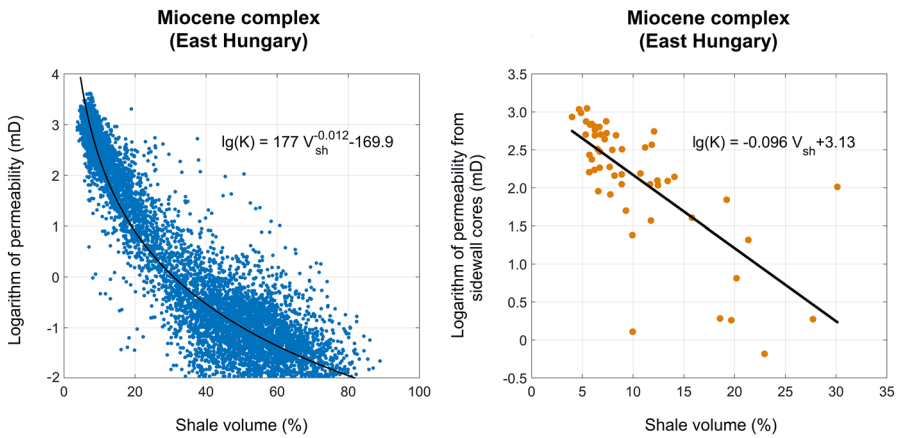


Fig. 2 Empirical connections between the intrinsic permeability and shale volume of Miocene hydrocarbon formations in the investigated site, Pannonian basin, East Hungary

Harrison et al. 2019). However, since the optimal control parameters highly depend on the optimization task, here an attempt is made to automatically select the optimal values of parameters c_1 and c_2 of PSO while solving factor analysis. These parameters are usually chosen according to empirical suggestions, which can make a problem ambiguous. To offer a more advanced solution of the presented PSO-based factor analysis, we propose a further developed algorithm to improve the FA-PSO method in regard of the c_1 and c_2 hyperparameters. For this purpose, we search for these parameters by Simulated Annealing method (SA) suggested by Metropolis et al. (1953) in an outer loop of our algorithm in an automated way. Then, in the inner loop, FA-PSO is run with the newly estimated c_1 and c_2 parameters. According to the previously outlined workflow, the procedure starts with the determination of the factor loadings according to Eq. (4). Then, we optimize the factor scores with the help of PSO by minimizing the objective function defined in Eq. (7). However, the control parameters c_1 and c_2 of PSO are also optimized during the procedure by SA in an outer loop as shown in Fig. 3. In the regression part of the workflow, we can correlate the first factor to available permeability values (and other petrophysical parameters). In the prediction phase, we estimate the permeability to missing logging intervals directly from the factor scores using the regression model in Eq. (12).

3 Numerical results

3.1 Single borehole application

We test the feasibility of factor analysis using in situ well logging data in a hydrocarbon well (Well-6). For permeability estimation, we simultaneously process the caliper (CAL), natural gamma-ray intensity (GR), potassium-thorium product ($K \times TH$), spontaneous potential (SP), deep resistivity (R_d), bulk density (ρ_b), neutron-porosity

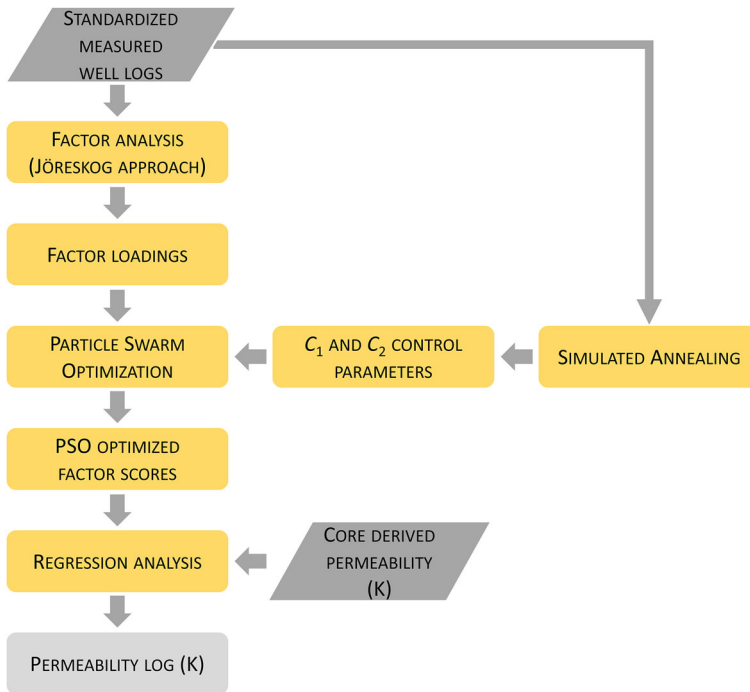


Fig. 3 Hyperparameter estimation using SA algorithm for improving the PSO based factor analysis method. Cognitive and social constants (c_1 , c_2) are automatically refined to give an estimate to the factor scores, while the permeability (K) is directly derived from the first factor using regression analysis

(Φ_N) and acoustic interval time (Δt) logs. Our investigation covers a 290 m long interval with a 0.2 m sampling interval. At the start of the FA-PSO procedure, the input logs are standardized and gathered into column vector $\mathbf{d}^{(m)}$ according to Eq. (6). Then the factor loadings are estimated for three factors by the non-iterative approximate method of Jöreskog (2007) based on Eq. (4). The resulting factor loadings are rotated by the varimax algorithm and are collected in Table 1. The first factor is strongly related to the natural gamma-ray intensity, potassium-thorium index and bulk density logs that make it a good lithological indicator, which is consistent with previous applications (Szabó and Dobróka 2017). The second factor log correlates with the neutron-porosity and acoustic logs, while the third factor is greatly sensitive to the variation of the deep resistivity log.

Once the factor loadings are determined, we optimize the factor scores by the algorithm of PSO. First, we give an initial solution for the factor scores by solving Eq. (5), thus we define the search space of the PSO process as $[-12, 12]$. We estimate the amount of variance explained by the factors based on the singular value decomposition of the reduced covariance matrix of standardized data $\mathbf{R}-\Psi = \mathbf{L}\mathbf{L}^T$. Since three factors explain more than 90% of total variance of the input data, we extract three factors from the dataset. Thus, the number of unknowns is 4,353 (3 factor \times 1451 depth points). For estimating their optimal values by PSO, we initialize 90 particles

Table 1 Orthogonally transformed factor loadings estimated in Well-6, Pannonian Basin, Hungary

Well log	Symbol	Factor 1	Factor 2	Factor 3
Borehole caliper	CAL	0.27	- 0.12	0.20
Natural gamma-ray intensity	GR	0.87	- 0.42	- 0.07
Potassium-thorium product	$K \times TH$	0.73	- 0.25	0.03
Spontaneous potential	SP	0.45	- 0.27	0.19
Deep resistivity	R_d	- 0.12	0.34	0.56
Bulk density	ρ_b	0.94	0.04	- 0.06
Neutron-porosity	Φ_N	0.38	- 0.64	- 0.40
Acoustic interval time	Δt	0.12	- 0.87	- 0.04

within the pre-defined search space. They search for the solution by moving around in the search domain according to Eqs. (8)–(9) to minimize the objective function in Eq. (7) in 5,000 iteration steps. In our experiment, we treat the cognitive and social parameters of PSO as unknown. Our algorithm allows changing them freely, instead of selecting them as arbitrary constants. We set the initial value of c_1 and c_2 equal to 1, respectively, while parameter w is set according to Eq. (10). At the end of the process, the estimated values of hyperparameters are $c_1 = 1.44$ and $c_2 = 2.06$. The variation of the control parameters of PSO and data distance as functions of iteration steps is plotted in Fig. 4. At the end of the optimization phase, the data distance calculated by Eq. (7) reaches 0.53. Both plots show a convergent and stable optimization procedure. The independence of results on the random initial particle positions in PSO algorithm has been earlier investigated in the context of inverse problem solutions (Abordán and Szabó 2020).

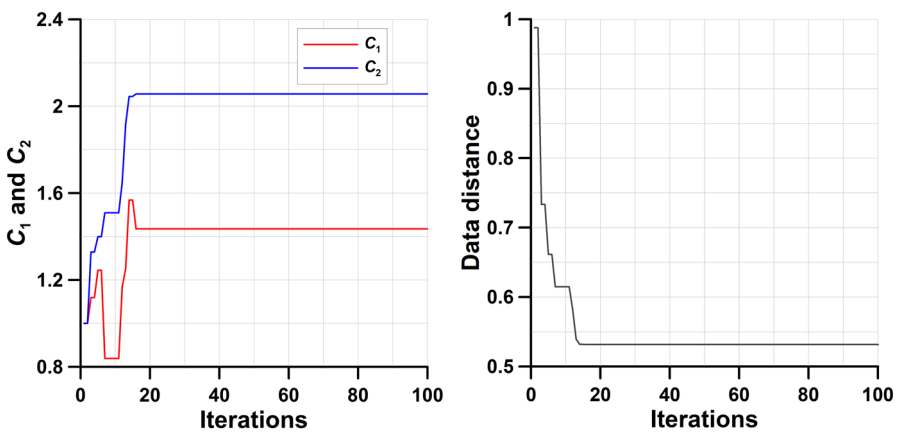
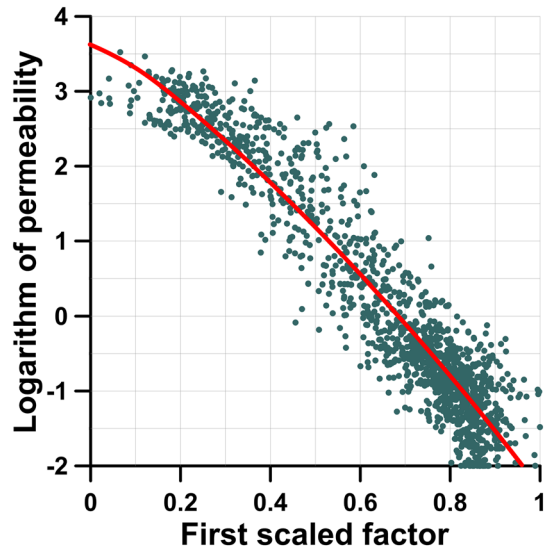


Fig. 4 Convergence of the 1D PSO based factor analysis procedure in Well-6, Pannonian Basin, Hungary. The variation of learning parameters (c_1 plotted with dashed line, c_2 plotted with continuous line) during the iterative procedure (*left panel*). The decrease of data distance (energy) vs. iteration steps (*right panel*)

Fig. 5 Regression relationship found between the first factor and the decimal logarithm of permeability measured in mD in Well-6, Pannonian Basin, Hungary



To find the regression relationship between the first factor found by the FA-PSO method and decimal logarithm of absolute permeability, we properly scale the former into the range of 0 to 1. For regression analysis, we use permeability data deterministically derived from Eq. (11). The regression coefficients in Eq. (12) are found to be $a = 5.90$, $b = 1.27$, $d = 2.28$ while c is fixed at 1, because it shows no sensitivity in this case. The correlation relationship between the scaled first factor and absolute permeability is plotted in Fig. 5. The Spearman's rank correlation coefficient of -0.77 between the first factor and the decimal logarithm of permeability indicates a significant (non-linear) inverse relationship. The input well logs of the FA-PSO procedure are plotted in Fig. 6 in the first 8 tracks (note that the depth scale represents relative depth coordinates). The last track indicates that the fit between the factor analysis derived permeability ($\lg(K)_{\text{FA-PSO}}$) and that of derived deterministically by Eq. (11) ($\lg(K)_{\text{DET}}$) is consistent along the investigated interval. The CPU time of the FA-PSO procedure for automatically determined hyperparameters is ~ 7 h (~ 3000 depth points), while for fixed hyperparameters it is 86 s using an octa-core processor workstation.

3.2 Test of hyperparameters

The learning parameters of classical PSO are usually set to default values based on recommendations (Kennedy and Eberhart 1995). We suggest an improved PSO algorithm to improve these parameters, which forms the basis of the workflow presented in Fig. 3. We focus on the estimation of parameters c_1 and c_2 to optimize the PSO process. As input, we select the GR, $K \times \text{TH}$, SP, R_d , ρ_b , Φ_N , Δt logs in a 20 m long interval in Well-3. At the beginning of the SA procedure, both parameters are set as 1 and in every iteration step their value is slightly altered by adding a random perturbation (δ) to both values. We generate parameter δ as random number in each

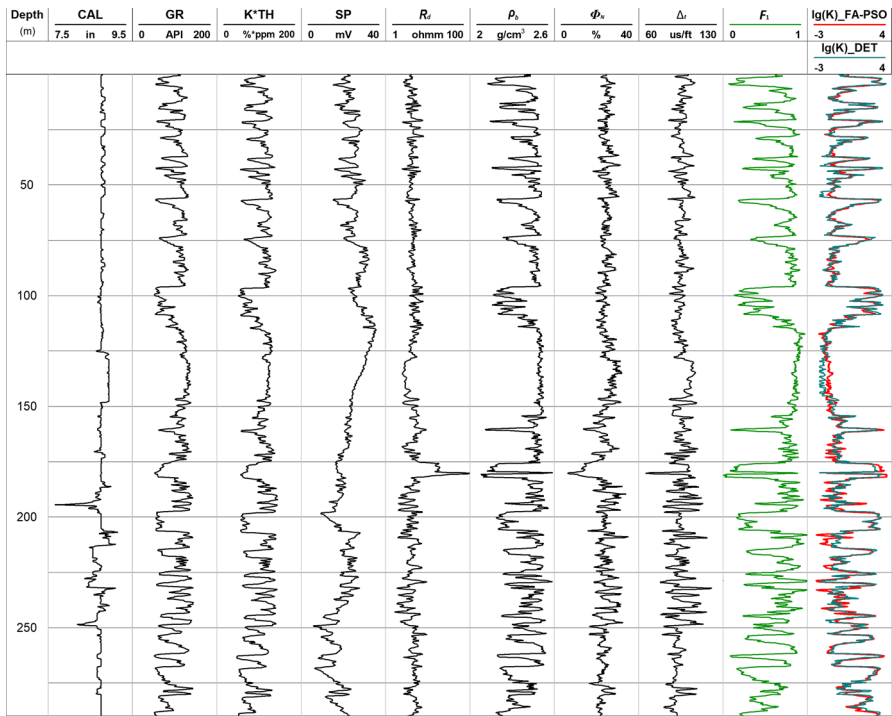


Fig. 6 The result of factor analysis in Well-6, Pannonian Basin, Hungary. The input logs of FA-PSO procedure are plotted in tracks 1-8: borehole caliper (CAL), natural gamma-ray intensity (GR), potassium-thorium product ($K \times TH$), spontaneous potential (SP), deep resistivity (R_d), bulk density (ρ_b), neutron-porosity (Φ_N) and acoustic (P-wave) transit time (Δ_t). The scaled first factor log (F_1) is plotted in track 9. In track 10, $lg(K)_{FA-PSO}$ represents the factor analysis derived logarithmic permeability (originally given in mD), while $lg(K)_{DET}$ shows the deterministic estimation for the same quantity

iteration step from the range $-\delta_{max}$ to δ_{max} . We also decrease the maximum value of perturbation (δ_{max}) iteratively using $\delta_{max} = \delta_{max} \times \varepsilon$, where constant ε is smaller than 1. Then with these new hyperparameters (c_1 and c_2), PSO optimizes the factor scores in 200 iterations. We define the energy function of SA as Eq. (7). If the energy difference (ΔE) in two subsequent SA iterations of the algorithm is negative (i.e., PSO reached a lower data distance), then the new values of c_1 and c_2 are accepted and the procedure is continued. If the energy difference is greater than 0 (i.e., the efficiency of PSO got worse with the new parameters), then the new c_1 and c_2 hyperparameters can be still accepted if a random number generated from 0 to 1 is smaller than acceptance probability $P_a = \exp(-\Delta E/T)$, where T denotes the actual temperature of the search. This mechanism is the basis of the SA algorithm because it enables the search to get out of local minima of the energy function. To ensure the convergence to the global optimum of the objective function, the temperature must be reduced logarithmically as suggested by Geman and Geman (1984). We change the temperature in the q th iteration as $T^{(new)} = T_0 / \lg(1 + q)$, where the initial temperature of the artificial system T_0 is experimentally set to 5×10^{-6} .

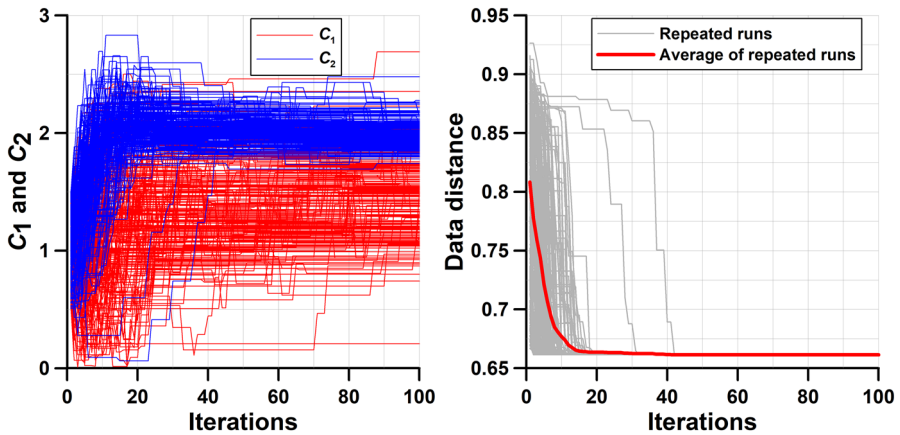


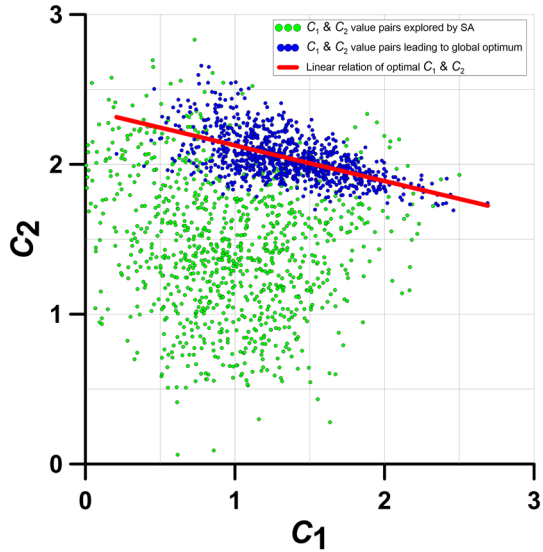
Fig. 7 Convergence of data distance (*right panel*) by altering control parameters c_1 and c_2 of the FA-PSO procedure (*left panel*) repeated 200 times using the dataset of Well-3, Pannonian Basin, Hungary

We run the SA procedure for 100 iteration steps to optimize the values of c_1 and c_2 as shown in Fig. 7. The newly generated control parameters are tested in each SA iteration step by running the FA-PSO method for 200 iterations for three consecutive times. This is particularly advantageous especially at the early stages, where data distances remain rather large due to the suboptimal values of c_1 and c_2 . The mean of the three consecutive FA-PSO runs for each SA iteration can be seen in Fig. 7 (on the right panel). As the figure indicates, the whole procedure was run 200 times to check its stability. After approximately 40 iteration loops, SA successfully optimizes the values of the learning parameters for the FA-PSO algorithm in all 200 separate program runs and data distance defined by Eq. (7) did not decrease any further. This shows the stability of the SA-based hyperparameter estimation method. In Fig. 8, we illustrate all those value pairs of c_1 and c_2 that were tried by SA during the iterations in 200 repeated program runs. Green dots indicate the ones that did not lead to the global optimum, while the blue ones enabled PSO to converge to the optimal solution of factor scores. According to the results, c_1 has less effect on the convergence and can be selected from a wider range (0.25–2.7), while the optimal range for c_2 is more restricted (1.7–2.6) when solving the problem of factor analysis. The fitted regression line indicates a moderate correlation between c_1 and c_2 . Our result agrees well with those of the literature, for instance Cleghorn and Engelbrecht (2014) derived similar convergent regions for the sum of the two parameters. This test justifies the advantage of using SA to find the optimal combination of hyperparameters. The CPU time of the FA-PSO procedure for automatically estimated hyperparameters is ~ 70 s, while for fixed hyperparameters is 0.27 s using an octa-core processor workstation.

3.3 Multi-borehole application

We extend the improved factor analysis to multi-dimensional cases, which allows the simultaneous processing of well logs measured from several neighboring boreholes.

Fig. 8 Cross-plot of learning parameters c_1 and c_2 estimated by the 1D Particle Swarm Optimization based factor analysis (green and blue dots) performed in Well-3, Pannonian Basin, Hungary. Red line indicates linear relationship between those points that led the PSO based factor analysis procedure to the global optimum. For the latter, the Pearson’s correlation coefficient -0.58 indicates moderate correlation between the two quantities



Let the measured data vector be $\mathbf{d}^{(h)}$ defined in Eq. (6) in the h th borehole ($h = 1, 2, \dots, H$). Then the model of factor analysis becomes

$$\begin{pmatrix} \mathbf{d}^{(1)} \\ \vdots \\ \mathbf{d}^{(h)} \\ \vdots \\ \mathbf{d}^{(H)} \end{pmatrix} = \text{diag}(\tilde{\mathbf{L}}^{(1)}, \dots, \tilde{\mathbf{L}}^{(h)}, \dots, \tilde{\mathbf{L}}^{(H)}) \times \begin{pmatrix} \mathbf{f}^{(1)} \\ \vdots \\ \mathbf{f}^{(h)} \\ \vdots \\ \mathbf{f}^{(H)} \end{pmatrix} + \begin{pmatrix} \mathbf{e}^{(1)} \\ \vdots \\ \mathbf{e}^{(h)} \\ \vdots \\ \mathbf{e}^{(H)} \end{pmatrix} \quad (13)$$

where $\tilde{\mathbf{L}}^{(h)}$ represents the matrix of factor loadings and vector $\mathbf{f}^{(h)}$ contains the estimated factor scores along the h th borehole. There is N_h number of measured depth points in the h th wellbore, therefore the total number of data processed by factor analysis is $N^* = N_1 + N_2 + \dots + N_H$. In Eq. (13), we determine the matrix of factor

Table 2 Rotated factor loadings estimated in Wells 1–5, Pannonian Basin, Hungary

Well log	Symbol	Factor 1	Factor 2
Natural gamma-ray intensity	GR	0.87	0.02
Shallow resistivity	R_s	-0.73	-0.01
Bulk density	ρ_b	0.83	-0.29
Neutron-porosity	Φ_N	0.82	0.18
Acoustic transit time	Δt	-0.01	0.64

loadings in a similar manner as in the previously presented one-dimensional case by Eq. (4), while we approximate the optimal values of factor scores by PSO.

We show a 2D application of FA-PSO to derive the permeability variations in a hydrocarbon zone in the same area of the Pannonian Basin. We gathered the wireline logging data collected from five wells (Well-1, -2, -3, -4, -5) into one dataset. The borehole investigated in Sect. 3.1 (Well-6) falls farther from the line of the other wells, thus we excluded it from our calculation. The GR, ρ_b , Φ_N , Δt and shallow resistivity (R_s) logs being completely available all depths serve as the input of 2D factor analysis. By combining the dataset of the five wells, the total number of data is 37,525. The wells are located along a 2,200 m long profile. By estimating two factors, the orthogonally transformed factor loadings are given in Table 2. When calculating the factor scores the learning parameters are simultaneously optimized. The resultant values of learning parameters are $c_1 = 1.59$ and $c_2 = 2.23$ (Fig. 9). Due to the increased number of unknowns, in this case the PSO routine requires 25,000 iteration steps with 90 particles. Based on the calculation of the singular values of the reduced covariance matrix of standardized wireline logs, the first factor explains 77% of the total data variance, while the remaining 23% is explained by the second factor. The first extracted factor correlates well with all processed well logs except the P-wave acoustic interval time log, while the second factor is quite the opposite.

We find a strong correlation between the first scaled factor and the decimal logarithm of permeability in the investigated Miocene formations (Fig. 10). The Spearman's rank correlation coefficient is -0.93. Using the proposed model in Eq. (12), we establish the local regression model, where the available permeability values are obtained from deterministic modeling using Eq. (11). The regression coefficients in this case are $a = 7.18$, $b = 4.5$, $c = 1.2$ and $d = -3.8$. By interpolating the output data by ordinary kriging, we derive the maps of the estimated factors and that of the permeability estimated by 2D FA-PSO method along the investigated cross-section (Fig. 11). The

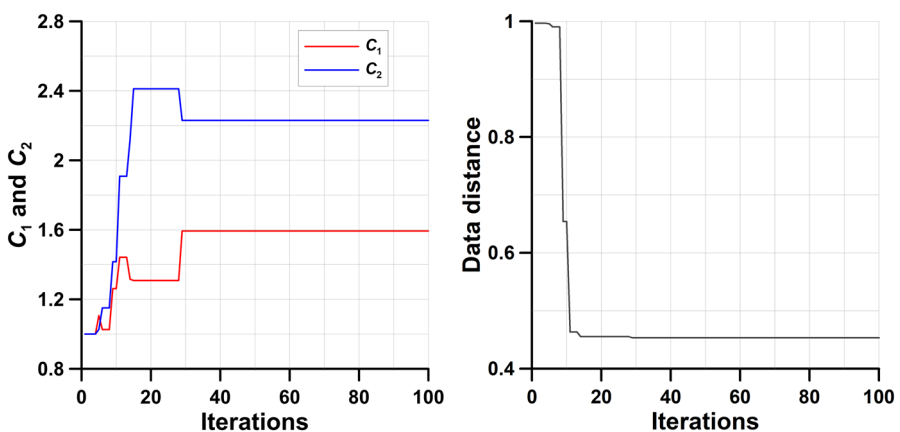


Fig. 9 Convergence of the 2D PSO based factor analysis procedure in Wells 1–5, Pannonian Basin, Hungary. The variation of learning parameters (c_1 plotted with dashed line, c_2 plotted with continuous line) during the iterative procedure (left panel), the decrease of data distance (energy) vs. iteration steps (right panel)

Fig. 10 Regression relationship found between the first factor and the decimal logarithm of permeability measured in mD in Wells 1–5, Pannonian Basin, Hungary

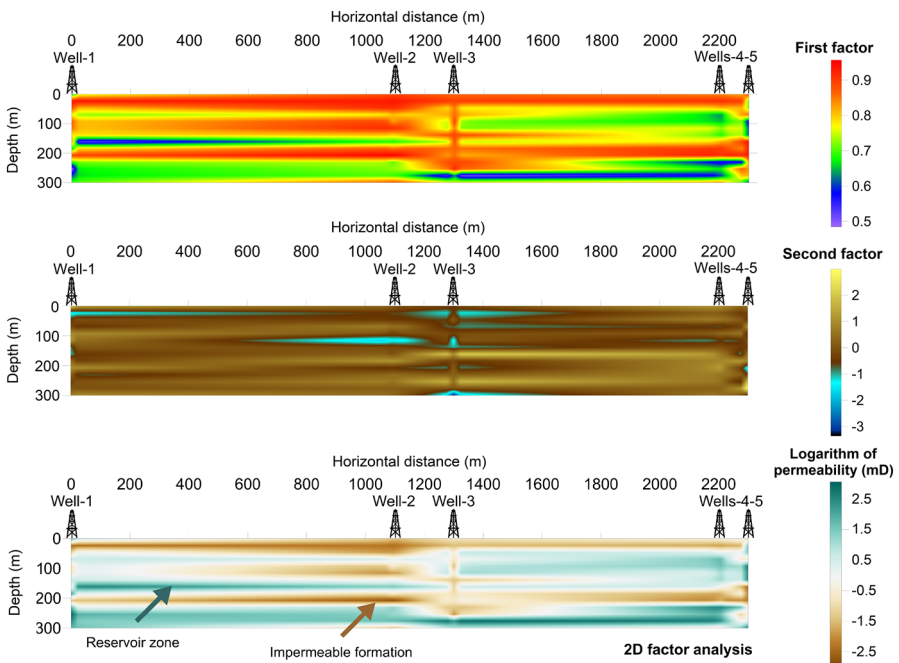
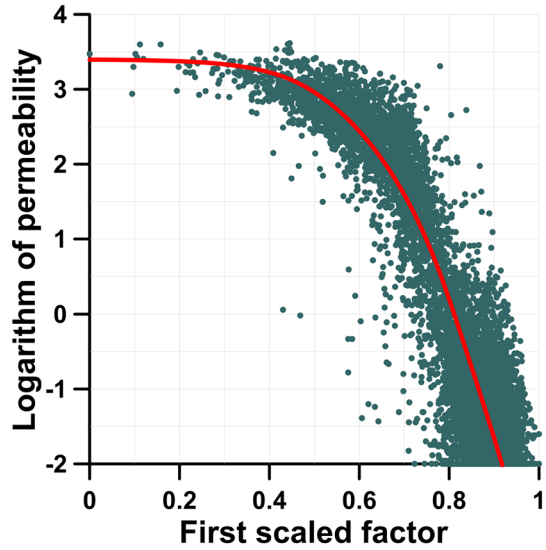


Fig. 11 Result of permeability prediction using 2D PSO assisted factor analysis for multiple wells. *Top panel* shows the section of the extracted first factor, the *middle panel* includes that of the second factor, *bottom panel* illustrates the estimated logarithmic permeability (originally given in mD) in Wells 1–5, Pannonian Basin, Hungary

depth of potential reservoirs and sealing units can be recognized well on the image (note that the depth scale represents relative depth coordinates). Of course, care must be taken, the farther the wells are from each other, the more uncertainty there is in permeability prediction because of lithological and structural variations. The reliability of the results can be ensured with the integration of core measurements and surface geophysical measurements. The CPU time of the FA-PSO procedure for automatically determined hyperparameters is ~ 100 h, while for fixed hyperparameters is ~ 20 min using an octa-core processor workstation.

3.4 Calibration by core data

The results of factor analysis highly depend not just on the way of optimization but also what type and quality of the input data we have in the phase of regression analysis (Fig. 3). We test the feasibility of the permeability prediction method using the combination of wireline logs and core plug data. Let us choose a well log suite from Well-3 in the investigated area, which include the natural gamma-ray intensity (GR), potassium-thorium product ($K \times TH$), spontaneous potential (SP), deep resistivity (R_d), shallow resistivity (R_s), bulk density (ρ_b), neutron-porosity (Φ_N) and acoustic interval time (Δt) logs. From core plugs, the equivalent liquid permeability (K_L) is available, which is derived from the measured gas permeability (K_G) as

$$K_L = \frac{K_G}{1 + (\kappa/P_m)} \quad (14)$$

where κ is a constant for a particular gas in a given rock type and P_m is the mean pressure (Klinkenberg 1941). We emphasize that that laboratory measurements on core plugs and wireline logging are carried out under different conditions (e.g., temperature and pressure etc.) yielding sometimes considerable differences in predicted permeability values. Moreover, core data provide only local information about permeability, while well logs are influenced by a larger extent of the rock formation, therefore the misfit between the well log derived permeability and that of core data sometimes can be significant. To improve the strength of correlation for the variables of Eq. (12), we let the most of the observed information concentrate into only one factor. We conduct the regression analysis to find the relationship between the first factor and lab permeability on those locations where core derived permeability data is available. In the investigated depth interval, core data is available at 54 non-equidistantly sampled locations, therefore wireline logging data is gathered only from these locations and standardized to serve as the input of factor analysis. The processed interval is 18 m long. First, factor loadings are estimated by Eq. (4) only for one factor. We show the calculated and rotated factor loadings in Table 3. Based on the table, we can conclude that the extracted factor from the wireline logging dataset shows a strong correlation with the GR, $K \times TH$, ρ_b and Φ_N logs, separately.

Once the factor loadings are calculated, the factor scores are optimized by PSO in the same manner as detailed in Sect. 3.1 (in Well-6). First, we solve Eq. (5) for estimating the initial values of factor scores to define the search space of the FA-PSO process [-3, 3]. We extract only one factor from the well logging dataset, therefore the number of

Table 3 Rotated loadings of the first factor estimated in Well-3, Pannonian Basin, Hungary

Wireline log	Symbol	Factor 1
Natural gamma-ray intensity	GR	0.97
Potassium-thorium product	$K \times TH$	0.94
Spontaneous potential	SP	0.32
Deep resistivity	R_d	-0.41
Shallow resistivity	R_s	-0.35
Bulk density	ρ_b	0.89
Neutron-porosity	Φ_N	0.87
Acoustic interval time	Δt	-0.65

unknowns adds up to only 54 (1 factor \times 54 depth points). For estimating their optimal values, we randomly initialize 45 particles in the pre-defined search space. They look for the solution of factor scores by moving around in the search space according to Eqs. (8–9) to minimize the objective function in 200 iteration steps earlier defined in Eq. (7). We let the control parameters (c_1, c_2) of the PSO algorithm change automatically. At the last iteration step, the resultant values of learning parameters are $c_1 = 1.98$ and $c_2 = 1.91$, while the data distance reaches 0.66 (Fig. 12).

To interpret the regression relationship between the first factor and the decimal logarithm of absolute permeability, first we scale the former into the range of 0 to 1. As a reference value, the equivalent liquid permeability data is used. The non-linear relationship according to the model suggested in Eq. (12) between the first scaled factor and permeability is plotted in Fig. 13. The regression coefficients are found to be $a = 2.95, b = 2.10, d = 0.17$ while c is fixed at 1. The Spearman’s rank correlation

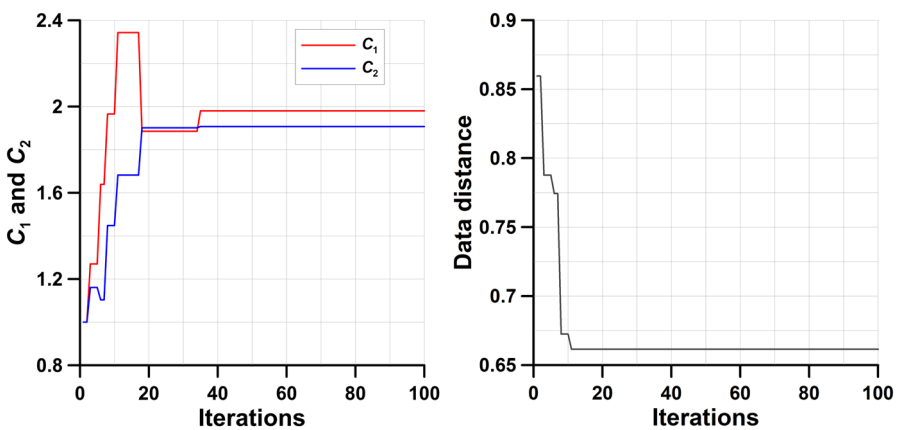
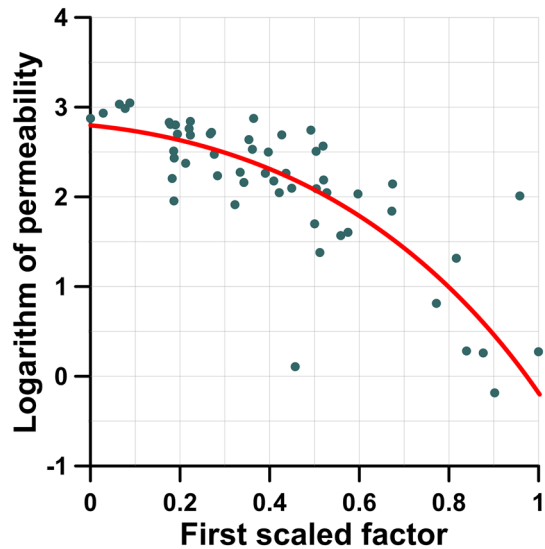


Fig. 12 Convergence of the 1D PSO based factor analysis procedure in Well-3, Pannonian Basin, Hungary. The variation of learning parameters (c_1 plotted with *dashed line*, c_2 plotted with *continuous line*) during the iterative procedure (*left panel*), the decrease of data distance (energy) vs. iteration steps (*right panel*)

Fig. 13 Regression relationship found between the first factor and the decimal logarithm of core permeability measured in mD in Well-3, Pannonian Basin, Hungary



coefficient is -0.78 , which indicates a strong non-linear relationship. Because of having some outlying values, we can see only a moderate deviation between the well log analysis derived permeability and the data available from core plugs, nevertheless, it is well described by Eq. (12). The input (observed) well logs of factor analysis are plotted in the first eight tracks in Fig. 14, the scaled first factor log is available in track 9 of the same figure. The last track contains the decimal logarithm of permeability derived by the FA-PSO method ($\lg(K)_{\text{FA-PSO}}$) and the core data ($\lg(K)_{\text{CORE}}$) (note that the depth scale represents relative depth coordinates). We conclude that the two independent methods determine similar permeability distribution with depth along the processed interval.

4 Discussion

Petrophysical properties including volumetric parameters such as porosity, shale volume and water saturation as well as permeability cannot be directly measured by downhole geophysical tools. Normally, we estimate them by inverting the observed wireline logs, which allows also the quantification of their estimation accuracy by using the law of error propagation. In the knowledge of the data covariance matrix including the data variances in its main diagonal, one can derive the covariance matrix of the estimated model parameters using linear inverse theory (Menke 1984). The PSO method as intelligent optimization tool has been implemented in a stochastic calibration context, specifically for the estimation of subsurface properties (Russian et al. 2019; Patani et al. 2021). Its efficiency has been shown in some applications in well logging inversion, too. For instance, by applying a series expansion based

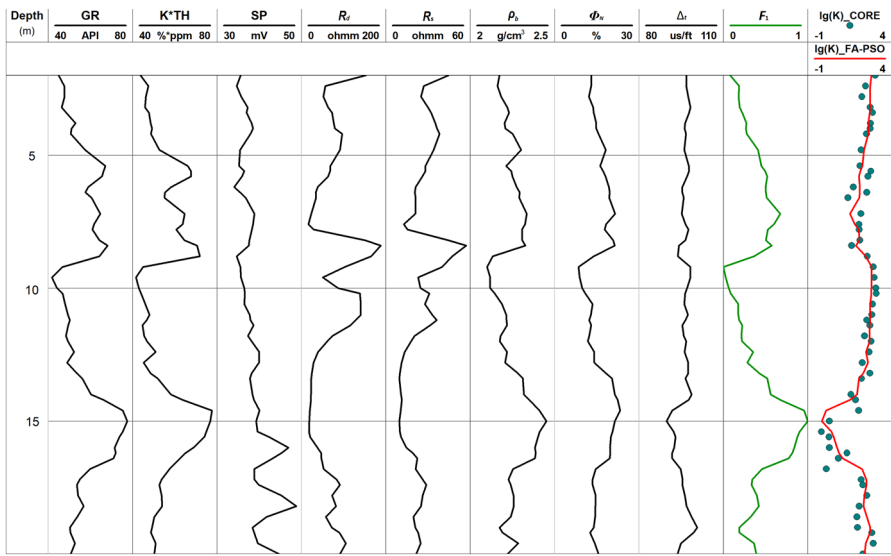


Fig. 14 The result of factor analysis in Well-3, Pannonian Basin, Hungary. The input logs of FA-PSO procedure are plotted in tracks 1-8: natural gamma-ray intensity (GR), potassiumthorium product ($K \times TH$), spontaneous potential (SP), deep resistivity (R_d), shallow resistivity (R_s), bulk density (ρ_b), neutron porosity (Φ_N) and acoustic (P-wave) transit time (Δ_t). The scaled first factor log (F_1) is plotted in track 9. In track 10, $lg(K)_{FA-PSO}$ represents the factor analysis derived logarithmic permeability (originally given in mD), while $lg(K)_{CORE}$ shows the core derived value for the same quantity

discretization scheme the so-called interval inversion method with a great over-determination (data-to-unknowns ratio) may significantly reduce the uncertainty of the inverted parameters (Dobróka et al. 2016). Global optimization tools like PSO, SA and genetic algorithms can be effectively combined with linearized inversion techniques to solve such an inverse problem. By the successive application of the PSO technique and the interval inversion method, not only the computer processing time of PSO inversion can be reduced, but there is a possibility to calculate the estimation errors of petrophysical properties from the Jacobi matrix generated in a given linear inversion step (Abordán and Szabó 2020; Szabó and Dobróka 2020).

The problem with the estimation of permeability is that the permeability is not included at all in the theoretical tool response functions connecting the model parameters with the measured data (Alberty and Hashmy 1984). Thus, the quality check of this parameter is not possible within the inversion procedure. In the practice, the permeability is derived from the inversion results (or from different other sources), by using some empirical relation. In this study, the suggested method for permeability prediction is based on factor analysis and its results are all set in a deterministic framework. Our approach provides a quite straightforward estimation for permeability using the factor scores. However, it is possible to test the reliability of the regression model by performing noise sensitivity tests using real data. Consider the well logs and core data presented in Sect. 3.4. First we add 5% Gaussian distributed noise to the input logs, then we run the FA-PSO procedure, and we repeat this process 30 times

(in different runs, the noise was always regenerated and added to the observed data). The histograms of the resultant regression coefficients (a, b, d) are plotted in Fig. 15. The average and ranges of their estimated values are: $a = 2.8 \pm 0.7, b = 2.25 \pm 0.75, d = 0.25 \pm 0.75$ (coefficient c is fixed as earlier). The result of regression analyses can be seen in the bottom right panel of the figure, which shows the consistency of regression model (12). The uncertainty of permeability prediction is acceptably low and is comparable with the level of data noise. This is also proven in Fig. 16, where the error intervals of the input logs, and those of the first factor and permeability are plotted, respectively. It can be seen that the uncertainty of permeability prediction is higher in formations with lower permeability, but it is favorable that the same parameter in the hydrocarbon bearing zones can be more reliably produced. The results are also confirmed by the core data available in the processed interval.

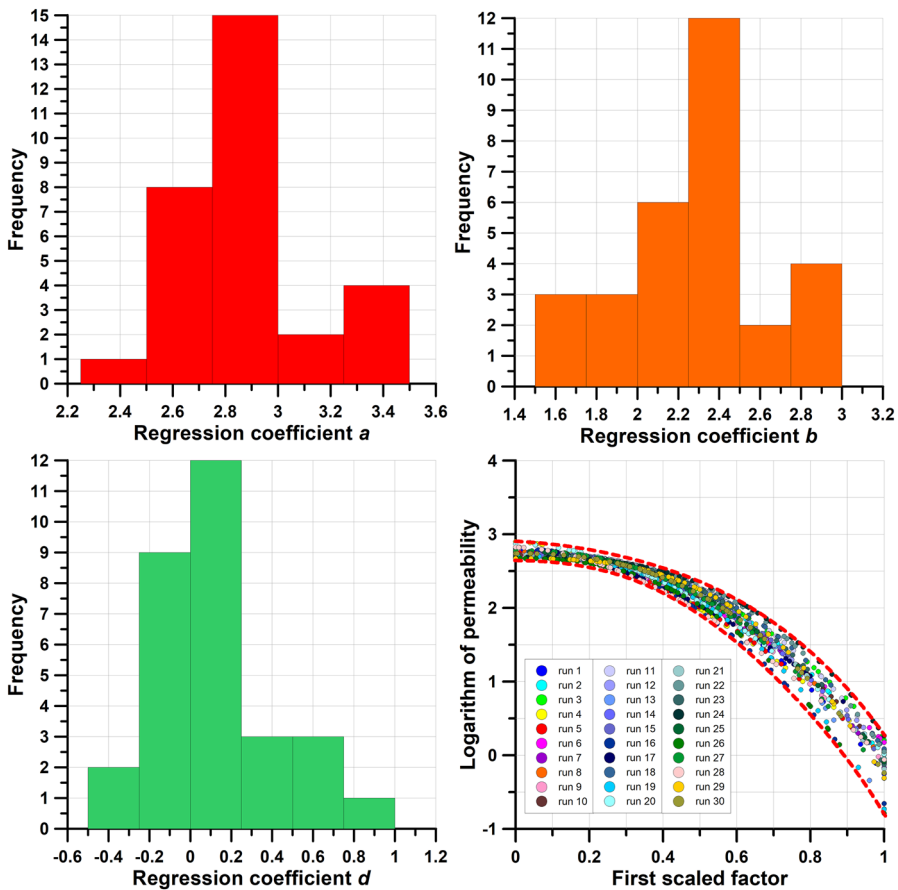


Fig. 15 Result of repeating the run of FA-PSO procedure 30 times. The input data were contaminated by 5% Gaussian distributed noise. Top panels and bottom left panel show the domain of estimated regression coefficients. Bottom right panel show the possible range of permeability values estimated by factor analysis

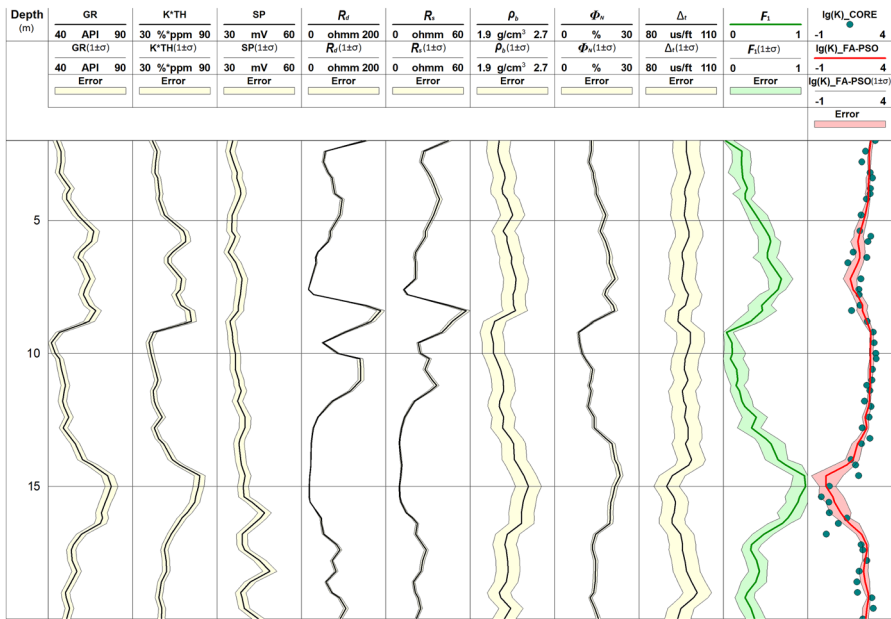


Fig. 16 The reliability of factor analysis based permeability estimation method tested by 30 repeated runs in Well-3, Pannonian Basin, Hungary. The input logs and their assumed error ranges (at 5% Gaussian noise) are plotted in tracks 1-8: natural gamma-ray intensity (GR), potassium-thorium product ($K \times TH$), spontaneous potential (SP), deep resistivity (R_d), shallow resistivity (R_s), bulk density (ρ_b), neutron-porosity (Φ_N) and acoustic (P-wave) transit time (Δ_t). The scaled first factor log (F_1) including its possible upper and lower boundaries is plotted in track 9. In track 10, $lg(K)_{FA-PSO}$ represents the factor analysis derived logarithmic permeability and its uncertainty interval, while $lg(K)_{CORE}$ shows the core derived value for the same parameter

5 Conclusions

In the paper, we introduce a particle swarm optimization assisted factor analysis with automatically optimized control parameters to permeability prediction in hydrocarbon reservoirs. The hyperparameter estimation based PSO process requires no prior knowledge on its learning parameters. A metaheuristic search is guided for estimating both the optimal set of cognitive and social constants and the factor scores. Based on the above, we perform exploratory factor analysis by a new alternative optimization tool. The advantages of the proposed method against traditional approaches are: (I) derivative-free solution is given, (II) optimal misfit is achieved between the measurements and predictions (as global optimization is made), (III) solution is practically independent from the initial selection of factor scores, (IV) highly adaptive procedure applied to different well log suites (see Table 1–3), (V) more reliable solution is given by automatically estimating the learning parameters of PSO than those treating arbitrarily chosen control parameters, (VI) robust solution can be assured using an appropriate choose of the misfit function. More robust norms like a most frequent value based or L_1 -norm can be substituted into the Euclidean norm based objective functions

(Szabó et al. 2018). At the moment, the only serious drawback of the hyperparameter estimation method is its long computational running time. For automatically calculated hyperparameters, the CPU times increase approximately 300 times compared to those of a single run using fixed values of c_1 and c_2 . However, based on preliminary program runs, parameters c_1 and c_2 can be determined for the investigated area and then fixed for factor analysis, which may reduce the running time by orders of magnitude. In the knowledge of proper initial values, instead of the SA algorithm, a linearized optimization procedure may be proposed to reduce the running time greatly.

We have established a well usable regression model between the first factor estimated by the particle swarm optimization based factor analysis and the permeability of hydrocarbon-bearing formations. It is noted that pilot tests confirmed the validity of the same regression relation in some hydrocarbon wells in the Carpathian Foredeep, South Poland (and hydraulic conductivity in Hungary and the USA). This allows the estimation of permeability from an independent well-log-analysis method, which is based on the comprehensive interpretation of multiple wireline logs. We have tested the developed method in single wells in Hungarian hydrocarbon-bearing formations, which was extended successfully to two-dimensional application to simultaneously evaluate the data measured in multiple boreholes. The factor analysis derived permeability is compared and verified both by deterministic modeling of well logs and laboratory based measurements. As a next research step, the inertia weight w might also be automatically selected for PSO, and the factor loadings will be used as unknowns as we have done it in our earlier applications. By using fast computers and the tools of parallel computing, a generalized FA-PSO method can be developed for a reliable and fast 3D permeability prediction in different oilfields.

Acknowledgements The research was carried out in the Project No. K-135323 supported by the National Research, Development and Innovation Office (NKFIH), Hungary. The authors thank for the permission of using the data to Anita É. Csoma, E&P Innovation Lead & Head of Subsurface, Hungarian Oil and Gas Company (MOL Plc).

Authors' contributions NPSZ conceived the idea of the relation between statistical factor and permeability as well as the new algorithm by combining PSO and FA. NPSZ and AA designed the data analysis strategy and carried it out. AA wrote the main part of the manuscript. MD supervised the research and reviewed and edited the manuscript. All authors contributed to the interpretation of results and the final version of the paper. The authors applied the EC approach for the sequence of authors. See <https://doi.org/10.1371/journal.pbio.0050018> for further details.

Funding Open access funding provided by University of Miskolc (ME).

Declarations

Conflict of interest The authors have declared that no competing interests exist.

Availability of data and materials Not applicable.

Code availability Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abordán, A., Szabó, N.P.: Particle swarm optimization assisted factor analysis for shale volume estimation in groundwater formations. *Geosci Eng* **6**, 87–97 (2018)
- Abordán, A., Szabó, N.P.: Uncertainty reduction of interval inversion estimation results using a factor analysis approach. *Int J Geomath* (2020). <https://doi.org/10.1007/s13137-020-0144-4>
- Abordán, A., Szabó, N.P.: Machine learning based approach for the interpretation of engineering geophysical sounding logs. *Acta Geod Geophys* (2021). <https://doi.org/10.1007/s40328-021-00354-4>
- Alberty, M., Hashmy, K.: Application of ULTRA to log analysis. In: SPWLA Symposium Transactions. Paper Z, pp. 1–17 (1984)
- Al-Mudhafar, W.J.: Integrating machine learning and data analytics for geostatistical characterization of clastic reservoirs. *J. Petrol. Sci. Eng.* (2020). <https://doi.org/10.1016/j.petrol.2020.107837>
- Ball, S.M., Chace, D.M., Fertl, W.H.: The well data system (WDS): advanced formation evaluation concept in a microcomputer environment. In: Proceedings of the SPE Eastern Regional Meeting, paper 17034, pp. 61–85 (1987)
- Bartlett, M.S.: The statistical conception of mental factors. *Br. J. Psychol.* **28**, 97–104 (1937)
- Cleghorn, C.W., Engelbrecht, A.P.: A generalized theoretical deterministic particle swarm model. *Swarm Intell* (2014). <https://doi.org/10.1007/s11721-013-0090-y>
- Coates, G. R., Dumanoir, J. L.: A new approach to improved log derived permeability. In: Proceedings of SPWLA 14th Annual Logging Symposium, Paper R (1973)
- Cosultchi, A., Sheremetov, L., Velasco-Hernandez, J., Batyrshin, I.: A method for aquifer identification in petroleum reservoirs: A case study of Puerto Ceiba oilfield. *J. Petrol. Sci. Eng.* (2012). <https://doi.org/10.1016/j.petrol.2012.06.026>
- Dahraj, N.U.H., Bhutto, A.A.: Linear mathematical model developed using statistical methods to predict permeability from porosity. In: PAPG/SPE Pakistan Section Annual Technical Conference, SPE-174716-MS (2014). <https://doi.org/10.2118/174716-MS>
- Del Ser, J., Osaba, E., Molina, D., Yang, X.S., Salcedo-Sanz, S., Camacho, D., Das, S., Suganthan, P.N., Coello, C.A., Herrera, F.: Bio-inspired computation: where we stand and what's next. *Swarm Evol. Comput.* (2019). <https://doi.org/10.1016/j.swevo.2019.04.008>
- Dobróka, M., Szabó, N.P., Tóth, J., Vass, P.: Interval inversion approach for an improved interpretation of well logs. *Geophysics* (2016). <https://doi.org/10.1190/geo2015-0422.1>
- Donelli, M., Franceschini, G., Martini, A., Massa, A.: An integrated multiscale strategy based on a particle swarm algorithm for inverse scattering problems. *IEEE Trans. Geosci. Remote Sens.* (2006). <https://doi.org/10.1109/TGRS.2005.861412>
- Essa, K.S., Elhoussein, M.: PSO (Particle Swarm Optimization) for interpretation of magnetic anomalies caused by simple geometrical structures. *Pure Appl. Geophys.* (2018). <https://doi.org/10.1007/s00024-018-1867-0>
- Feng, Y., Teng, G., Wang, A., Yao, Y.: Chaotic inertia weight in Particle Swarm Optimization. In: Second International Conference on Innovative Computing, Information and Control (2007). doi: <https://doi.org/10.1109/ICIC.2007.209>
- Fernández Martínez, J.L., García Gonzalo, E., Fernández Álvarez, J.P., Kuzma, H.A., Menéndez Pérez, C.O.: PSO: a powerful algorithm to solve geophysical inverse problems: Application to a 1D-DC resistivity case. *J. Appl. Geophys.* (2010). <https://doi.org/10.1016/j.jappgeo.2010.02.001>
- Geman, S., Geman, D.: Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.* (1984). <https://doi.org/10.1109/TPAMI.1984.4767596>

- Handhal, A.M.: Prediction of reservoir permeability from porosity measurements for the upper sandstone member of Zubair Formation in Super-Giant South Rumila oil field, southern Iraq, using M5P decision tress and adaptive neuro-fuzzy inference system (ANFIS): a comparative study. *Model. Earth Syst. Environ.* (2016). <https://doi.org/10.1007/s40808-016-0179-6>
- Harrison, K.R., Ombuki-Berman, B.M., Engelbrecht, A.P.: An Analysis of control parameter importance in the Particle Swarm Optimization algorithm. In: Tan, Y., Shi, Y., Niu, B. (eds.) *Advances in swarm intelligence: ICSI 2019 Lecture Notes in Computer Science*. Springer, Heidelberg (2019)
- Jafarpour, B., McLaughlin, D.B.: Reservoir characterization with the Discrete Cosine Transform. *SPE J.* (2009). <https://doi.org/10.2118/106453-PA>
- Jiang, M., Luo, Y.P., Yang, S.Y.: Stochastic convergence analysis and parameter selection of the standard particle swarm optimization algorithm. *Inf. Process. Lett.* (2007). <https://doi.org/10.1016/j.ipl.2006.10.005>
- Jöreskog, K.G.: Factor analysis and its extensions. In: Cudeck, R., MacCallum, R.C. (eds.) *Factor analysis at 100, historical developments and future directions*, pp. 47–77. Lawrence Erlbaum Associates (2007)
- Kaiser, H.F.: The varimax criterion for analytical rotation in factor analysis. *Psychometrika* (1958). <https://doi.org/10.1007/BF02289233>
- Kennedy, J., Eberhart, R.C.: Particle swarm optimization. In: *Proceedings of IEEE international conference on neural networks 4*, pp. 1942–1948 (1995). doi: <https://doi.org/10.1109/ICNN.1995.488968>
- Klinkenberg, L.J.: The permeability of porous media to liquids and gases. *Drilling and Production Practice, American Petroleum Inst.*, pp. 200–213 (1941). DOI:<https://doi.org/10.5510/OGP20120200114>
- Kumar, A., Pant, S., Ram, M., Singh, S.B.: On solving complex reliability optimization problem using multi-objective particle swarm optimization. In: Ram, M., Paulo Davim, J. (eds.) *Mathematics Applied to Engineering*, pp. 115–131, Academic Press. (2017)
- Lawley, D.N., Maxwell, A.E.: Factor analysis as a statistical method. *The Statistician* (1962). <https://doi.org/10.2307/2986915>
- Maćkiewicz, A., Ratajczak, W.: Principal components analysis (PCA). *Comput. Geosci.* (1993). [https://doi.org/10.1016/0098-3004\(93\)90090-R](https://doi.org/10.1016/0098-3004(93)90090-R)
- Menke, W.: *Geophysical data analysis: discrete inverse theory*. Academic Press Inc., London (1984)
- Metropolis, N., Rosenbluth, A.E., Rosenbluth, M.N., Teller, A.H., Teller, E.: Equation of state calculations by fast computing machines. *J. Chem. Phys.* (1953). <https://doi.org/10.1063/1.1699114>
- Nguyen, B.H., Xue, B., Zhang, M.: A survey on swarm intelligence approaches to feature selection in data mining. *Swarm Evol. Comput.* (2020). <https://doi.org/10.1016/j.swevo.2020.100663>
- Patani, S.E., Porta, G.M., Caronni, V., et al.: Stochastic inverse modeling and parametric uncertainty of sediment deposition processes across geologic time scales. *Math Geosci* (2021). <https://doi.org/10.1007/s11004-020-09911-z>
- Puskarczyk, E.: Application of multivariate statistical methods and Artificial Neural Network for facies analysis from well logs data: an example of Miocene deposits. *Energies* (2020). <https://doi.org/10.3390/en13071548>
- Puskarczyk, E., Jarzyna, J., Porebski, S.: Application of multivariate statistical methods for characterizing heterolithic reservoirs based on wireline logs - Example from the Carpathian Foredeep Basin (Middle Miocene, SE Poland). *Geological Quarterly* (2015). <https://doi.org/10.7306/gq.1202>
- Rahmat-Samii, Y.: Genetic algorithm (GA) and particle swarm optimization (PSO) in engineering electromagnetics. In: *17th International Conference on Applied Electromagnetics and Communications*, pp. 1–5 (2003). doi: <https://doi.org/10.1109/ICECOM.2003.1290941>
- Revil, A., Cathles, L.M.: Permeability of Shaly Sands. *Water Resources Research* (1999). <https://doi.org/10.1029/98WR02700>
- Russian, A., Riva, M., Russo, E.R., et al.: Stochastic inverse modeling and global sensitivity analysis to assist interpretation of drilling mud losses in fractured formations. *Stoch Environ Res Risk Assess* (2019). <https://doi.org/10.1007/s00477-019-01729-4>
- Schön, J.H., Georgi, D.T.: Dispersed shale, shaly-sand permeability - a hydraulic analog to the Waxman-Smits equation. In: *SPWLA 44th Annual Loggin Symposium, SPWLA-2003-S*, pp. 22–25 (2003)
- Self, R., Atashnezhad, A., Hareland, G.: Reducing drilling cost by finding optimal operational parameters using Particle Swarm Algorithm. In: *SPE Deepwater Drilling and Completions Conference, SPE-180280-MS* (2016). <https://doi.org/10.2118/180280-MS>
- Serra, O.: *Fundamentals of well-log interpretation*. Elsevier, Amsterdam (1984)
- Shaw, R., Srivastava, S.: Particle swarm optimization: A new tool to invert geophysical data. *Geophysics* (2007). <https://doi.org/10.1190/1.2432481>

- Shi, Y., Eberhart, R.A.: A modified particle swarm optimizer. In: IEEE International Conference on Evolutionary Computation Proceedings, pp. 69–73 (1998). pp 69–73 doi: <https://doi.org/10.1109/ICEC.1998.699146>
- Singh, N.P.: Permeability prediction from wireline logging and core data: a case study from Assam-Arakan basin. *J Petrol Explor Prod Technol* (2019). <https://doi.org/10.1007/s13202-018-0459-y>
- Soares, R.V., Luo, X., Evensen, G., Bhakta, T.: 4D seismic history matching: Assessing the use of a dictionary learning based sparse representation method. *J. Petrol. Sci. Eng.* (2020). <https://doi.org/10.1016/j.petrol.2020.107763>
- Szabó, N.P.: Shale volume estimation based on the factor analysis of well-logging data. *Acta Geophys.* (2011). <https://doi.org/10.2478/s11600-011-0034-0>
- Szabó, N.P.: Hydraulic conductivity explored by factor analysis of borehole geophysical data. *Hydrogeol. J.* (2015). <https://doi.org/10.1007/s10040-015-1235-4>
- Szabó, N.P., Dobróka, M.: Extending the application of a shale volume estimation formula derived from factor analysis of wireline logging data. *Math Geosci* (2013). <https://doi.org/10.1007/s11004-013-9449-2>
- Szabó, N.P., Dobróka, M.: Robust estimation of reservoir shaliness by iteratively reweighted factor analysis. *Geophysics* (2017). <https://doi.org/10.1190/geo2016-0393.1>
- Szabó, N.P., Dobróka, M.: Exploratory factor analysis of wireline logs using a Float-Encoded Genetic Algorithm. *Math Geosci* (2018). <https://doi.org/10.1007/s11004-017-9714-x>
- Szabó, N.P., Dobróka, M.: Interval inversion as innovative well log interpretation tool for evaluating organic-rich shale formations. *J. Petrol. Sci. Eng.* (2020). <https://doi.org/10.1016/j.petrol.2019.106696>
- Szabó, N.P., Dobróka, M., Drahos, D.: Factor analysis of engineering geophysical sounding data for water saturation estimation in shallow formations. *Geophysics* (2012). <https://doi.org/10.1190/geo2011-0265.1>
- Szabó, N.P., Balogh, G.P., Stickel, J.: Most frequent value-based factor analysis of direct-push logging data. *Geophys. Prospect.* (2018). <https://doi.org/10.1111/1365-2478.12573>
- Tang, H., White, C.D.: Multivariate statistical log-log-facies classification on a shallow marine reservoir. *J. Petrol. Sci. Eng.* (2008). <https://doi.org/10.1016/j.petrol.2008.05.004>
- Timur, A.: An investigation of permeability, porosity, and residual water saturation relationships. In: SPWLA 9th Annual Logging Symposium, SPWLA-1968-J (1968)
- Wyllie, M.R.J., Rose, W.D.: Some theoretical considerations related to the quantitative evaluation of the physical characteristics of reservoir rock from electrical log data. *J Pet Technol* (1950). <https://doi.org/10.2118/950105-G>

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