ORIGINAL PAPER

Uncertainty reduction of interval inversion estimation results using a factor analysis approach

Armand Abordán[1,2] · Norbert Péter Szabó[1,2]

Abstract

This paper aims to investigate the impact of the overdetermination (data-to-unknowns) ratio on the global inversion of wireline logging data. In the course of the so-called interval inversion method, geophysical data measured in a borehole over a longer depth range is jointly inverted and the depth variation of the investigated petrophysical parameters are expanded into series using Legendre polynomials as basis functions resulting in a highly overdetermined inverse problem. A metaheuristic Particle Swarm Optimization (PSO) approach is applied as a first phase of inversion for decreasing the starting model dependence of the interval inversion procedure. In the subsequent linear inversion steps, by using the measurement error of logging tools and the covariance matrix of the estimated petrophysical parameters, we can quantify the accuracy of the model parameters. The dataset used in this study consists of nuclear, resistivity and sonic logs which are inverted to compute porosity, shale volume and water saturation along the investigated interval. For increasing the data-to-unknowns ratio of the inverse problem, shale volume is estimated separately by a PSO-based factor analysis and fixed as known parameter for the interval inversion process. Since the shale volume has been described as high degree Legendre polynomial, a significant increase of the overdetermination ratio considerably decreases the uncertainty of the remaining model parameters allowing for a more reliable calculation of hydrocarbon content.

Keywords Factor analysis · Series expansion · Interval inversion · Particle swarm optimization

Mathematics Subject Classification 15A09 · 15A29 · 15A18 · 41A27 · 86A22

✉ Armand Abordán
   gfaa@uni-miskolc.hu

1   Department of Geophysics, University of Miskolc, Miskolc-Egyetemváros 3515, Hungary

2   MTA-ME Geoengineering Research Group, University of Miskolc,
    Miskolc-Egyetemváros 3515, Hungary

1 Introduction

Geophysics provides several methods for the quantitative evaluation of subsurface geological structures. The physical properties of subsurface formations can be studied by surveying techniques such as magnetics, gravity and seismics. By interpreting these physical properties, we can acquire detailed information of the subsurface geology, which can lead us to potential mineral resources including hydrocarbon-bearing formations. To properly assess the economic value of a hydrocarbon reservoir, wireline logging methods must be applied as well. Well logging is an efficient and frequently used tool for the in situ characterization of rock formations (Serra 1984). Usually, electrical and sonic properties are measured along with nuclear and dimensional (technical, well diagnostics and production logging) measurements in the wellbore. Then the acquired data is processed through deterministic modeling or some inversion procedure to estimate the petrophysical properties of the investigated formations, such as shale volume, porosity and water saturation. Several possible solutions are available in the literature for the wireline logging inverse problem (Alberty and Hashmy 1984; Ball et al. 1987; Jarzyna et al. 2002; Narayan and Yadav 2006). Conventionally, inversion of wireline logging data is done in a local manner, meaning that data measured at a given depth point is jointly inverted to estimate the petrophysical parameters at that same depth point (Drahos 2005; Mayer and Sibbit 1980). This usually leads to a marginally overdetermined inverse problem, since we have slightly more logging tools than unknowns, including shale volume, porosity and water saturation in the invaded zone and in the virgin zone. Although this can be done very quickly and delivers adequate results, the low data-to-unknowns ratio has a negative effect on the estimation accuracy of parameters. A possible solution, interval inversion was developed for increasing the data-to-unknowns ratio of the well logging inverse problem (Dobróka and Szabó 2001). This approach provides a significant improvement in the estimation error of model parameters relative to local inversion (Dobróka et al. 2016). In the interval inversion method, petrophysical parameters are assumed to be the functions of depth, therefore depth-dependent response functions are introduced to relate the measured data to the unknown physical properties of geological formations of longer intervals. Then the model parameters are discretized by series expansion using Legendre polynomials. This way, inversion can be carried out simultaneously for a longer interval rather than just in a specific depth point. The number of observed data does not increase, but the simultaneous processing of several depth points greatly increases the relative number of data compared to series expansion coefficients as unknowns of the inverse problem.

Generally, the overdetermination ratio in evaluating conventional (shaly sand) formations is somewhere around 1.5 in case of local inversion, this is increased to 6 or more depending on the length of the inverted dataset and the desired resolution that is controlled by the number of series expansion coefficients used. A further increase of the overdetermination ratio of the interval inversion method can be made by properly decreasing the number of unknowns. Some parameters

are available from independent (reliable) sources which may be integrated into the joint inverse problem. By the improvement of the overdetermination ratio, one can reduce the estimation errors and maintain the vertical resolution of the estimated model parameters. We leave the number of series expansion coefficients unchanged, however, we estimate the shale volume of the investigated formation prior to inversion by factor analysis (Szabó 2011). Then we incorporate the resultant shale volume log into the interval inversion procedure, thus the number of known data is increased and the number of petrophysical parameters to be discretized by series expansion is decreased from 4 to 3. A similar approach was developed for direct push logging data, where factor analysis is used to derive the water content of the investigated shallow subsurface formations to reduce the number of unknowns of the inversion procedure afterwards (Abordán and Szabó 2019a). Factor analysis and similar dimension reduction methods are widely applied in geosciences to support data interpretation, e.g., principal component analysis for analyzing multi- and hyperspectral image datasets (Benedetto et al. 2013), for the analysis of processes affecting the water quality of a sub-section of a river (Tanos et al. 2011) or as an aid for characterizing a hydrocarbon reservoir of Late Miocene (Grund and Geiger 2011).

In this study, we focus on studying the accuracy improvement of the interval inversion result. It is shown that the estimation error of model parameters can be decreased considerably by involving the information of factor analysis. To overcome the shortcomings of the conventionally applied Damped Least Squares (DLSQ) method (Marquardt 1959), i.e., high derivative and starting model dependence, for the solution of both the geophysical inverse problem and factor analysis, we utilize the particle swarm optimization (PSO) technique (Kennedy and Eberhart 1995). The algorithm of PSO utilizes swarm intelligence to solve optimization problems. This metaheuristic method has already been applied to solve geoscience-related problems numerous times, e.g., array lateral logging inversion (Zhu et al. 2019), identifying groundwater contaminant sources (Guneshwor et al. 2018), inversion of resistivity and IP sounding data (Cui et al. 2017) and for optimizing well placement (Ding et al. 2014).

2 Local inversion of wireline logging data

In the course of local inversion, petrophysical parameters are estimated in each measured depth point separately by inverting the measured well logging data recorded at the same depth. The wireline logging dataset inverted in this paper includes the natural gamma-ray intensity (GR), neutron porosity (Φ_N), true resistivity (R_t), acoustic interval time (Δt), bulk density (ρ_b) and potassium concentration (K) logs. The response equations used to relate the measured data to the model parameters are as follows (Wyllie et al. 1956; Alberty and Hashmy 1984; Baker Atlas 1996)

$$GR = \rho_b^{-1}(V_{sh}GR_{sh}\rho_{sh} + V_{sd}GR_{sd}\rho_{sd}), \tag{1}$$

$$\Phi_N = \left\{ \begin{array}{c} \Phi_{N,mf} - (1 - S_{x0})C_{cor} - 2\Phi(1 - S_{x0})S_{hrf}(1 - 2.2\rho_{hc}) \\ \cdot \left[1 - (1 - S_{x0})(1 - 2.2\rho_{hc}) \right] \end{array} \right\} + V_{sh}\Phi_{N,sh} + V_{sd}\Phi_{N,sd}, \tag{2}$$

$$R_t = \left[\frac{\Phi^m S_w^n}{aR_w(1 - V_{sh})} + \frac{V_{sh}S_w}{R_{sh}} \right]^{-1}, \tag{3}$$

$$\Delta t = \Phi \left[\Delta t_{mf} S_{x0} + (1 - S_{x0})\Delta t_{hc} \right] c_p + V_{sh}\Delta t_{sh} + V_{sd}\Delta t_{sd}, \tag{4}$$

$$\rho_b = \Phi \left[\rho_{mf} - 1.07(1 - S_{x0})(\alpha_0 \rho_{mf} - 1.24\rho_{hc}) \right] + V_{sh}\rho_{sh} + V_{sd}\rho_{sd}, \tag{5}$$

$$K = \rho_b^{-1} \left(\Phi S_{x0} K_{mf}\rho_{mf} + V_{sh}K_{sh}\rho_{sh} + V_{sd}K_{sd}\rho_{sd} \right), \tag{6}$$

where the model parameters are porosity (Φ), shale volume (V_{sh}), sand volume (V_{sd}), water saturation in the invaded zone (S_{x0}) and water saturation in the virgin zone (S_w). All other parameters in the response equations are taken as constants during the inversion procedure, including the physical properties of the mud filtrate (mf), shale (sh), sand (sd) and hydrocarbon (hc), the textural parameters, cementation exponent (m), saturation exponent (n) and tortuosity factor (a). As well as the mud filtrate coefficients (α_0, C_{cor}), the compaction factor (c_p) and the residual hydrocarbon coefficient (S_{hrf}).

Then both the measured$^{(m)}$ and calculated$^{(c)}$ well logs are collected into column vectors

$$\mathbf{d}^{(m)} = \left[GR^{(m)}, \Phi_N^{(m)}, R_t^{(m)}, \Delta t^{(m)}, \rho_b^{(m)}, K^{(m)} \right]^T \tag{7}$$

$$\mathbf{d}^{(c)} = \left[GR^{(c)}, \Phi_N^{(c)}, R_t^{(c)}, \Delta t^{(c)}, \rho_b^{(c)}, K^{(c)} \right]^T \tag{8}$$

where superscript T means transpose. Model parameters are estimated by minimizing the deviation between the vectors defined in Eqs. (7), (8). This is conventionally solved by some linearized method (Menke 1984). In this study, we utilize the metaheuristic method of PSO to minimize the squared difference between the observed and calculated data vectors as defined in Eq. (9)

$$E = \sqrt{\frac{1}{S} \sum_{s=1}^{S} \left(\frac{d_s^{(m)} - d_s^{(c)}}{d_s^{(m)}} \right)^2} \cdot 100[\%], \tag{9}$$

where S is the number of applied logging tools. This metaheuristic method is more and more frequently used to solve complex geophysical inverse problems (Shaw and Srivastava 2007). It is inspired by the movement of animals like birds or fish in search of food. PSO utilizes a swarm of particles randomly generated in the search space to find the optimal solution. In an n-dimensional search space, the position of

the ith particle can be written as $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{in})^T$ and similarly the velocity of the ith particle is $\mathbf{v}_i = (v_{i1}, v_{i2}, \ldots, v_{in})^T$, which defines both the direction and distance of movement of the particle in each iteration step.

The particles, as solution candidates, move around in the search space looking for the best solution defined by the objective function according to Eqs. (10), (11)

$$\mathbf{x}_i(t + 1) = \mathbf{x}_i(t) + \mathbf{v}_i(t + 1), \tag{10}$$

$$\mathbf{v}_i(t + 1) = w\mathbf{v}_i(t) + r_1 c_1(\mathbf{p}_i(t)) - \mathbf{x}_i(t)) + r_2 c_2(\mathbf{g}(t) - \mathbf{x}_i(t)), \tag{11}$$

where $i = 1, 2,\ldots, L$ and L is the size of the swarm, iteration steps are denoted by t and $t = 1,\ldots,t_{max}$ where the last iteration step is t_{max}. During the iteration steps the best position of each particle is stored and continuously updated in $\mathbf{p}_i = (p_{i1}, p_{i2}, \ldots, p_{in})^T$ and the very best position of the whole swarm is stored in vector $\mathbf{g}$. Control parameters c_1 and c_2 are set during initialization, c_1 controls to what extent the personal best position $\mathbf{p}_i$ affects the movement of the ith particle and c_2 defines the movement of particles in the direction of the best position found by the whole swarm. Here, both c_1 and c_2 are set to 2 as recommended in the literature (Kennedy and Eberhart 1995). As an alternative approach, an inversion method can be used to refine the above control parameters in an automated way (Abordán and Szabó 2019b). The so-called inertia weight w was introduced to better control the algorithm through the velocity of particles (Shi and Eberhart 1998). It is set in each iteration step according to $w_{new} = w_{old} \bullet w_{damp}$, where w is set at the start of the algorithm to 1 and w_{damp} is a damping factor set to 0.99. For randomizing the search to some extent, uniformly distributed random numbers r_1 and r_2 from the range of 0 to 1 are introduced to Eq. (11). In the local inversion procedure, each particle of the swarm represents a solution for the model parameters and the particle with the best cost after the last iteration step is accepted as the solution of the optimization problem.

3 Interval inversion of wireline logging data

In the course of interval inversion, the petrophysical parameters are assumed to be the functions of depth, which are estimated for a longer interval in one inversion procedure using all data recorded in that interval. Therefore, the local probe response functions are modified to be depth-dependent

$$\mathbf{d}_s^{(c)}(z) = g_s(m_1(z), \ldots, m_i(z), \ldots, m_M(z)), \tag{12}$$

where g_s denotes the response function of the sth logging tool ($s = 1, 2, \ldots, S$, where S is the number of logging instruments) and m_i is the ith petrophysical property (there is M number of petrophysical parameters). The model parameters in Eq. (12) can be discretized by series expansion (Dobróka et al. 2016)

$$m_i(z) = \sum_{q=1}^{Q^{(i)}} B_q^{(i)} P_{q-1}(z), \tag{13}$$

where B_q is the qth series expansion coefficient and the lth degree Legendre polynomial can be written using

$$P_{q-1}(z) = P_l(z) = (2^l l!)^{-1} \frac{d^l}{dz^l}(z^2 - 1)^l, \tag{14}$$

the value of the ith model parameter along the inverted interval is described by $Q^{(i)}$ number of series expansion coefficients. The coordinates of the well logging measurements are scaled in the interval -1 to 1 where the Legendre polynomials form an orthogonal set of functions. The main advantage of the described series expansion-based inversion is that the required number of expansion coefficients to describe the model parameters is reasonably smaller than the number of inverted data, which leads to a well-overdetermined procedure. The optimal values of the expansion coefficients are found by the minimization of the relative data distance of measured and calculated data as in Eq. (15). Calculated data is obtained by using Eqs. (12), (13)

$$D = \frac{1}{N} \sum_{n=1}^{N} \left(\sqrt{\frac{1}{S} \sum_{s=1}^{S} \left(\frac{d_s^{(m)} - d_s^{(c)}}{d_s^{(m)}} \right)^2} \right) \cdot 100 \, [\%], \tag{15}$$

where N denotes the number of measured depth points along the investigated interval and S is the number of applied logging tools. Equation (15) serves as the objective function to be minimized. In the optimization problem, the normalized overall deviation between the measured and calculated data is reduced by the iterative algorithm of the Damped Least Squares (DLSQ) method (Marquardt 1959). When input data accuracies are not known, weights can be calculated automatically (Drahos et al. 2011) which could be adopted for interval inversion as well.

To overcome the starting model dependence of the procedure, the optimal values of expansion coefficients describing the petrophysical parameters are first approximated by PSO, once the solution is adequately close to the optimum, the inversion procedure is switched to a linearized optimization phase, where the final values of expansion coefficients are found by the DLSQ method and then are substituted into Eq. (13) to calculate the model parameters for the whole inverted interval. This greatly reduces the runtime of the procedure and allows for the calculation of estimation errors, since the covariance matrix of the model parameters estimated by the DLSQ method relates to the data covariance matrix that also includes the variances of measured data (Menke 1984). For the series expansion coefficients the covariance matrix is

$$\text{cov}\mathbf{B} = \mathbf{G}^{-g}\text{cov}\mathbf{d}^{(m)}(\mathbf{G}^{-g})^T, \tag{16}$$

where $\mathbf{d}^{(m)}$ is the vector of measured data and $\mathbf{G}^{-g}$ is the general inverse matrix of the linearized inversion. To calculate the uncertainty of estimated petrophysical

parameters, we have to introduce the depth-dependent model covariance matrix of the model parameters (Dobróka et al. 2016)

$$[\text{cov}\mathbf{m}(z)]_{ij} = \sum_{n=1}^{Q^{(i)}} \sum_{m=1}^{Q^{(j)}} P_{n-1}(z)(\text{cov}\mathbf{B})_{hh'} P_{m-1}(z), \tag{17}$$

where $\mathbf{B}$ is the vector of series expansion coefficients, indices are ($i = 1, 2, \ldots, M$; $j = 1, 2, \ldots, M$; $h = n + Q^1 + Q^2 + \ldots, + Q^{i-1}$; $h' = m + Q^1 + Q^2 + \ldots, + Q^{j-1}$). Using Eq. (17) one can derive the estimation error of model parameters as

$$\sigma\left[m_i(z)\right] = \left[\text{cov}\mathbf{m}_{ii}(z)\right]^{1/2}. \tag{18}$$

4 Factor analysis

This multivariate statistical tool is used to reduce the number of variables in a dataset to a smaller number of uncorrelated variables to help the data interpretation and to reveal hidden information (Lawley and Maxwell 1962). In wireline logging applications, these new variables are called factor logs, which can be related to petrophysical parameters such as shale volume or permeability through regression analysis (Szabó and Dobróka 2018). Other applications include lithological classification (Asfahani 2014) or the recognition of potential shale gas deposits (Puskarczyk et al. 2019).

To start the statistical procedure, first we have to standardize the S number of measured well logs and put them into the matrix $\mathbf{D}$, where each column contains the data measured by a different logging tool and there is N number of rows representing the measured depth points along the borehole. Then $\mathbf{D}$ is decomposed as

$$\mathbf{D} = \mathbf{F}\mathbf{L}^{\mathrm{T}} + \mathbf{E}, \tag{19}$$

where $\mathbf{F}$ denotes the N-by-R matrix of factor scores and R is the number of extracted factors. $\mathbf{L}$ is the S-by-R matrix of factor loadings and $\mathbf{E}$ denotes the N-by-S error matrix. Based on Eq. (19), the measured well logs are derived as the linear combination of the extracted factors. The factor loadings quantify the correlation relationship between the measured data and the computed factors. Most of the data variance is represented by the first-factor log, which is the first column of the matrix $\mathbf{F}$. The factor loadings are estimated by a non-iterative estimation method (Jöreskog 2007)

$$\mathbf{L} = (diag\mathbf{\Sigma}^{-1})^{-1/2}\mathbf{\Omega}(\mathbf{\Gamma} - \theta\mathbf{I})^{1/2}\mathbf{U}, \tag{20}$$

where $\mathbf{\Gamma}$ is the diagonal matrix of the first R number of sorted eigenvalues of the sample covariance matrix $\mathbf{\Sigma}$, and the first R number of eigenvectors are in matrix $\mathbf{\Omega}$ and $\mathbf{U}$ denotes an arbitrarily chosen R-by-R orthogonal matrix. In this paper, for finding the optimal values of factor scores, the previously presented algorithm of PSO is used (Abordán and Szabó 2018). Therefore Eq. (17) is rearranged to

$$\mathbf{d} = \tilde{\mathbf{L}}\mathbf{f} + \mathbf{e}, \tag{21}$$

where $\mathbf{d}$ denotes the *SN* length vector of measured well logging (standardized) data, $\tilde{\mathbf{L}}$ is the *NS*-by-*NR* matrix of factor loadings, $\mathbf{f}$ is the *RN* length vector of factor scores and $\mathbf{e}$ is the *SN* length vector of errors. Starting the procedure, $\tilde{\mathbf{L}}$ is estimated by Eq. (20) and then rotated with the varimax algorithm (Kaiser 1958) for getting more meaningful factors. Then PSO is applied to find the optimal values of factor scores $\mathbf{f}$. For this optimization problem, we choose the objective function based on the square of the L_2 norm as

$$E = \sqrt{\frac{1}{NS} \sum_{i=1}^{NS} (d_i^{(m)} - d_i^{(c)})^2} \cdot 100 \, [\%], \tag{22}$$

where $\mathbf{d}^{(m)}$ and $\mathbf{d}^{(c)}$ denote the measured and calculated (standardized) well-logging data vectors, respectively. Once Eq. (22) is minimized, the first extracted factor F_1 can be related to the shale volume of the investigated formation using regression analysis.

5 Field example

5.1 Interval inversion procedure

The wireline logging dataset used in this study was recorded in a hydrocarbon exploratory well drilled in the Pannonian Basin in Hungary. The processed interval is 19.2 m long and is built up of gas-bearing unconsolidated shaly sand layers of Pliocene age. In the interval inversion procedure the natural gamma-ray intensity (*GR*), neutron porosity (Φ_N), true resistivity (R_t), acoustic interval time (Δt), potassium concentration (K) and bulk density (ρ_b) logs are processed to estimate porosity (Φ or POR), shale volume (V_{sh}), water saturation in the flushed zone (S_{x0}) and water saturation in the virgin zone (S_w). Using the material balance equation $\Phi + V_{sh} + V_{sd} = 1$, we can calculate sand volume without increasing the number of unknowns. We use only the potassium concentration log from spectral gamma-ray measurements, because of the type of the clay minerals present in the studied formation. Logging was done with a sampling rate of 0.1 m, thus the inverted dataset including 6 types of logs contains 1158 data points. By discretizing the 4 model parameters according to Eq. (13) with Legendre polynomials of degree 44, the number of unknowns is 180, because the required number of series expansion coefficients for each petrophysical parameter is the maximum degree of Legendre polynomials plus one. Thus the data-to-unknowns ratio is 6.4. The optimal values of expansion coefficients are found by minimizing Eq. (15) with the algorithm of PSO. At the start of the inversion procedure, 45 particles are initialized within PSO, each representing a possible solution for the expansion coefficients describing the petrophysical parameters. The only constraint is that the *0*th degree Legendre polynomials are initialized from the ranges of $0 \le B_0^{\Phi} \le 0.4, 0 \le B_0^{V_{sh}} \le 1, 0 \le B_0^{S_w} \le 1, 0 \le B_0^{S_{x0}} \le 1$, however these
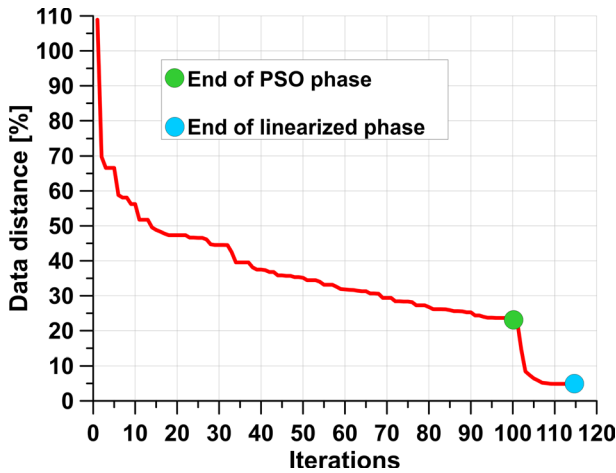
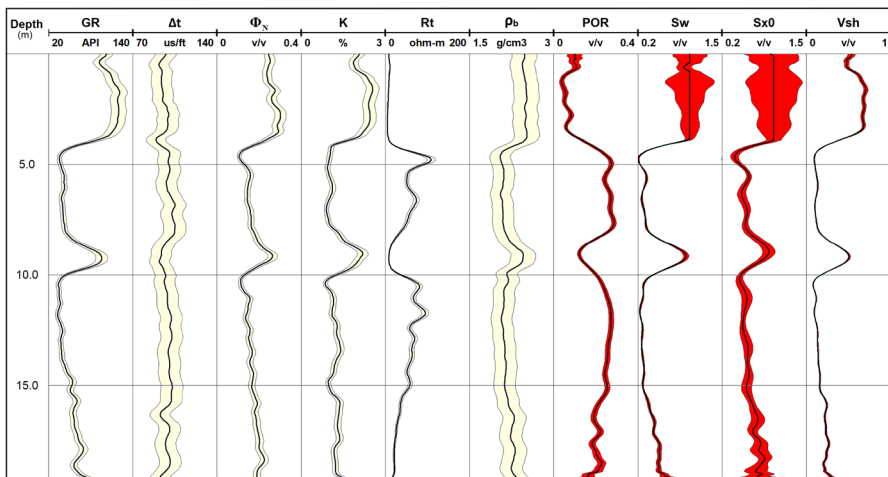**Fig. 1** Convergence of data distance to the optimum having 4 unknown petrophysical properties

Fig. 2 The measured well logs in the first 6 tracks with their assumed uncertainties (shaded area) and the resultant petrophysical parameters and their estimation errors (shaded area) in the last 4 track

values cover the full possible range of estimated petrophysical parameters. All the rest expansion coefficients are randomly generated from the range of -0.2 and 0.2. Thus, the initialized set of particles have an average data distance of $2.72 \times 10^7 \%$, a standard deviation of data distance of $1.61 \times 10^8 \%$ and the maximum data distance of all particles is $1.08 \times 10^9 \%$. Then the set of particles move in the search space according to Eqs. (10), (11) to find the optimal solution where the squared data distance between the measured and calculated data is minimal. PSO is run for 100 iteration steps where it reaches a data distance of 23.51%, then the procedure is followed with a linearized inversion phase using the DLSQ method for 15 more steps,

the final data distance is 4.87% (Fig. 1). The whole iteration process takes 30 s on a quad core workstation.

For checking the quality of the inversion estimates, first, measured data standard deviations are assumed to be $\sigma_{GR} = 0.08, \sigma_{\Phi_N} = 0.09, \sigma_{R_t} = 0.06, \sigma_{\Delta t} = 0.06, \sigma_{\rho_b} = 0.05, \sigma_K = 0.07$. Then by Eq. (18), we can calculate the estimation error of the resultant petrophysical parameters by using the uncertainties of the input well logs (Fig. 2).

The average estimation error of the resultant petrophysical parameters are $\Phi = 0.026v/v, S_w = 0.112v/v, S_{x0} = 0.245v/v, V_{sh} = 0.024v/v$.

5.2 PSO-based factor analysis

To assess how the overdetermination ratio affects the suggested PSO-based interval inversion method, first, we have to derive one of the petrophysical parameters from some other source and then incorporate it into the inversion procedure. In this study, we choose factor analysis to estimate shale volume of the investigated formation (Szabó and Dobróka 2017). First, well logging data including the natural gamma-ray intensity, neutron porosity, true resistivity, potassium concentration and bulk density logs of the investigated interval is standardized and put into the data vector **d** defined in Eq. (21). Factor loadings representing the correlation relationship between the extracted factor and measured logs are calculated according to Eq. (20). We extract only one factor from the dataset, the calculated factor loadings are $L^{(GR)} = 0.99, L^{(\Phi_N)} = 0.96, L^{(R_t)} = -0.88, L^{(\rho_b)} = 0.98$ and $L^{(K)} = 0.94$, which are fixed for the next phase of the procedure to save CPU time. Then PSO finds the optimal values of the factor scores by minimizing Eq. (22) utilizing 90 particles as solution candidates in 1000 iteration steps in 20 s on a quad core workstation (Fig. 3a). Based
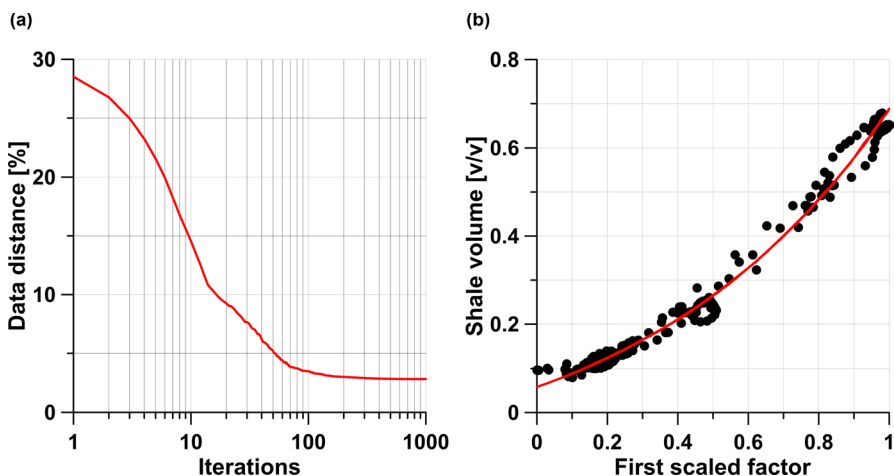

Fig. 3 The convergence of the PSO-based factor analysis procedure (**a**). The functional relation of the first extracted factor and shale volume (**b**): exponential regression model (red line) and shale volume estimated by local inversion (dots) (color figure online)

on the calculated loadings, all logs included in factor analysis correlate well with the first extracted factor. First, we have to scale it into the range of 0 to 1 so that it is comparable to shale volume. Then by regression analysis, the relation can be found between the extracted factor and shale volume of the investigated formation. In this study, we choose the local inversion derived shale volume for regression analysis that is detailed in chapter 2. Other options would be a deterministic approach based on the natural gamma-ray intensity log (Larionov 1969) or core analysis. The relationship between the first factor and shale volume typically takes the form of

$$V_{sh} = a e^{bF_1} + c,$$
(23)

where the regression coefficients in this case with 95% confidence bounds are found to be $a = 0.197$ [$a_{min} = 0.1598$, $a_{max} = 0.2344$], $b = 1.434$ [$b_{min} = 1.291$, $b_{max} = 1.577$] and $c = -0.139$ [$c_{min} = -0.1826$, $c_{max} = -0.0949$] (Fig. 3b).

5.3 Interval inversion improved by factor analysis

Interval inversion is run again with the same parameters and constraints as detailed in chapter 5.1, however, shale volume is now considered as a known parameter along the inverted interval derived from factor analysis. Thus the number of inverted data is increased to 1351 (7×193) and the number of unknowns is decreased to 135 because only 3 model parameters need to be discretized this time by series expansion using Legendre polynomials of degree 44. This results in an overdetermination ratio of 10, which is a 56.3% increase compared to the case detailed in chapter 5.1 where 4 model parameters were estimated by interval inversion. The convergence of data distance both in the global and linearized phases of the inversion procedure is quite steady (Fig. 4).
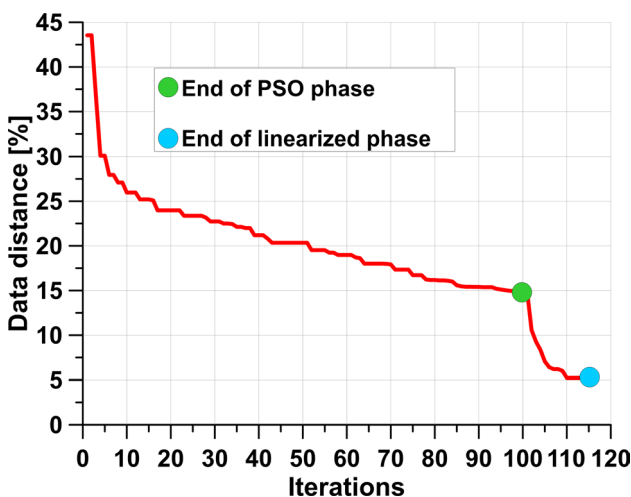

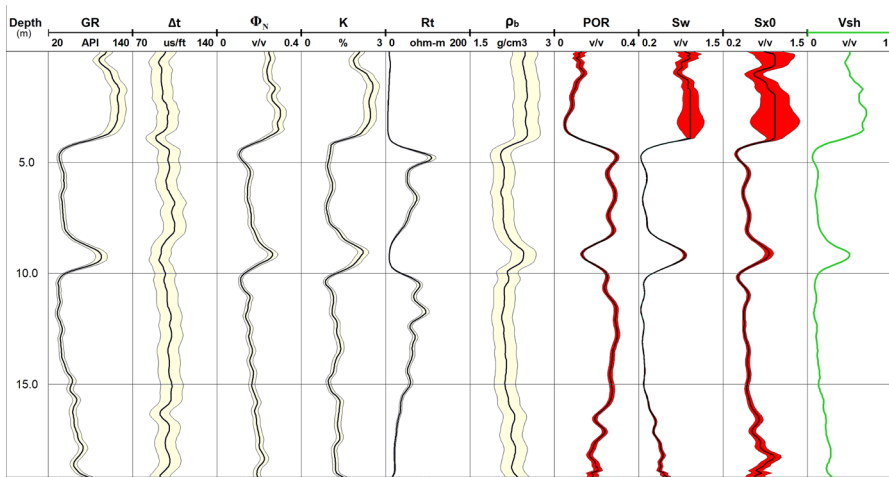**Fig. 4** Convergence of data distance for the hybrid interval inversion procedure

Fig. 5 The measured well logs in the first 6 tracks with their assumed uncertainties (shaded area) and the resultant petrophysical parameters and their estimation errors (shaded area) in tracks 7–9 and the shale volume derived from factor analysis in the last track

After 100 iteration steps, PSO reaches a data distance of 14.69% by minimizing Eq. (15). Then the DLSQ method further decreases data distance to 5.22% in 15 iteration steps. Including the runtime of factor analysis, this takes 40 s on a quad core workstation. Due to the increased data-to-unknowns ratio, the estimation error of the resultant petrophysical parameters is decreased (Fig. 5), the shale volume derived by factor analysis based on Eq. (23) can be seen in the last track.

The average estimation errors of the resultant petrophysical parameters calculated by Eq. (18) are decreased to $\Phi = 0.024v/v$, $S_w = 0.074v/v$, $S_{x0} = 0.170v/v$
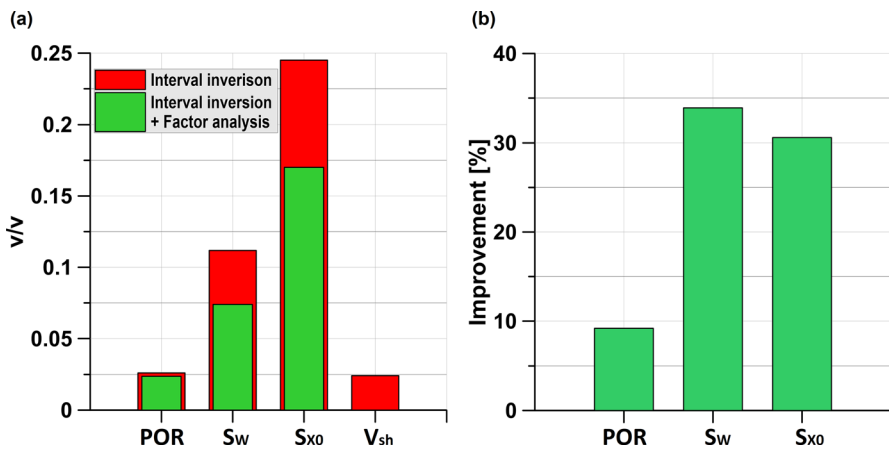
Fig. 6 The decrease in average estimation errors due to the increased data-to-unknowns ratio of the interval inversion method (**a**). The improvement of estimation in percent (**b**)

(Fig. 6a), which is quite an improvement as indicated in percentage (Fig. 6b). However, the estimation errors in the impermeable sections (high shale volume) are still fairly high compared to those in permeable intervals, especially in case of water saturations. This is due to the strong correlation between the two model parameters, which could be possibly resolved by the redefinition of the probe response functions of the forward problem.

The accurate estimation of these parameters is especially important since the irreducible and movable hydrocarbon volumes can be calculated as $V_{hc,irr} = \Phi(1 - S_{x0})$ and $V_{hc,mov} = \Phi(S_{x0} - S_w)$, respectively. The increased overdetermination ratio would also allow for the estimation of additional parameters within the inversion procedure, such as zone parameters or other constants found in the response equations of logging tools in Eqs. (1)–(6).

6 Discussion

To extract more information from a geophysical dataset, data acquisition is often followed by some sort of inverse modeling to get a detailed picture of the physical properties of the investigated structures. Inverse problems are conventionally solved by linearized (known as gradient-based) methods (Tarantola 2005). These methods have several drawbacks, e.g., the result of the inversion procedure is greatly sensitive to the chosen starting model. During the search in the parameter space, the inversion algorithm often gets stuck in a local minimum of the objective function near the defined starting model and a global optimum is impossible to be found. However, if adequate a priori information (initial model) is available about the investigated structure, these linear optimization techniques can provide good solutions effectively and fast. To overcome such problems of the linearized inversion methods, optimization techniques utilizing random search have been developed in the past decades. Some of the most commonly used global optimization methods in geophysics are simulated annealing and the genetic algorithm (Sen and Stoffa 2013; Wilkosz and Wawrzyniak-Guz 2019). Particle swarm optimization used in this study can also eliminate the starting model dependence of the inverse problem. It is capable to search through a greater extent of the possible solutions than the linearized inversion methods without trapping in a local minimum (Fernández-Martínez et al. 2008). To check the stability of the proposed PSO-based inversion method, 10 independent runs are performed for the case where the results of factor analysis are used to further increase the overdetermination-ratio of the procedure. First, 45 particles are initialized in the search space, each representing a possible solution of the 135 series expansion coefficients. The initial solutions generated by PSO are very diverse (Fig. 7a–c).

Then each solution candidate is refined according to Eqs. (10), (11) in 100 iteration steps and then the algorithm is switched to a linearized optimization phase to find the solution of the inverse problem in 15 more iteration steps. The data distance converges to the optimal value in all 10 independent runs (Fig. 8), which concurs with the findings that metaheuristic methods can be effectively applied to eliminate the starting model dependence of inverse problems (Pace et al. 2019).
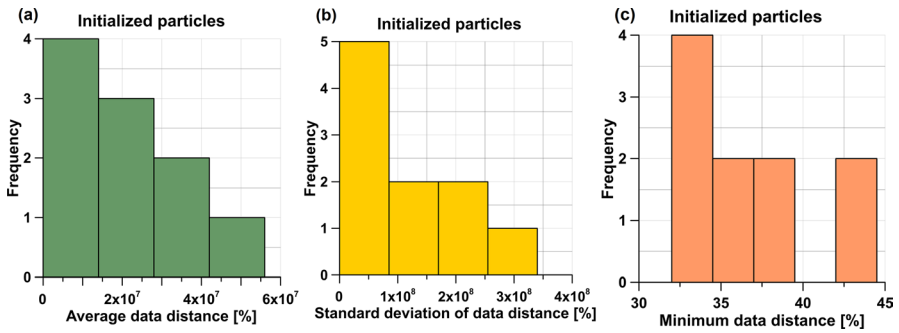
Fig. 7 Statistical distribution of the randomly generated starting models by PSO. The average data distance (**a**), the standard deviation (**b**) and the minimum data distance (**c**) of the solution candidates at initialization
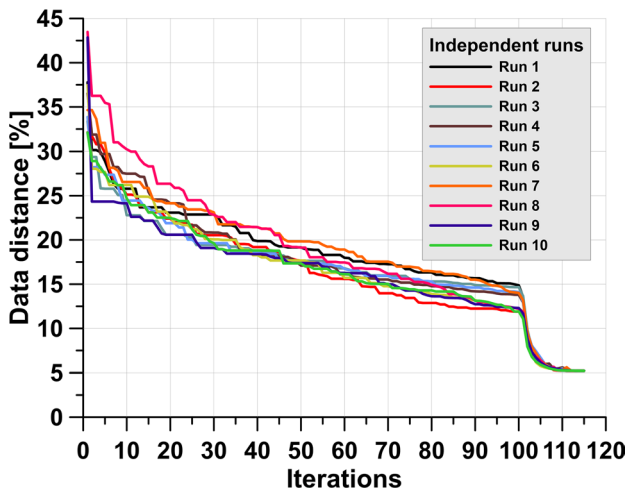

Fig. 8 Convergence plots of the PSO-initialized interval inversion procedure for 10 independent program runs

It is our experience that once the data distance is adequately small ($\sim 15\%$), the optimization can be continued by a faster linearized algorithm to find the final solution of the series expansion coefficients without trapping in a local minimum. Here the average data distance reached at the end of the procedure is 5.22% with a standard deviation of only 0.0027%. This proves the applicability and the effectiveness of the suggested hybrid method for solving the wireline logging inverse problem. However, since PSO is a metaheuristic, the presented method could be applied to a wide range of optimization problems where setting an initial model is problematic, e.g., a similar two-step process was effectively applied for full-waveform inversion, where very fast simulated annealing was combined with a conventional gradient-based method (Datta and Sen 2016).

7 Conclusions

The suggested PSO-based metaheuristic approach for solving the series expansion-based interval inversion of well log data proves to be quick and effective. The starting model dependence of the procedure is virtually eliminated by PSO and the switch to the linearized DLSQ method near the optimum greatly reduces the runtime of the inversion and allows for the calculation of estimation errors. It is shown that the factor analysis derived shale volume can be successfully incorporated into the interval inversion method to increase its data-to-unknowns ratio, and thus improving the estimation accuracy of the estimated petrophysical parameters. Factor analysis was also improved by the use of the highly adaptive PSO method. A 56.3% increase in the overdetermination ratio results in a 9.2% improvement in the estimation accuracy of porosity, 33.9% improvement of water saturation in the virgin zone and 30.6% improvement of the water saturation in the flushed zone. These parameters are the basis for calculating movable hydrocarbons, therefore their most reliable estimation is of crucial importance. The increased overdetermination ratio also gives the possibility to automatically estimate the value of some zone parameters within inversion, while still maintaining a good level of resolution of the estimated model parameters. The suggested procedure to increase the overdetermination of the well logging inverse problem might be very effectively used for unconventional reservoirs where multi-mineral models need to be built and therefore the number of unknowns is considerably higher than in case of conventional reservoirs.

Author contributions AA developed the code for the optimization task, ran the tests and prepared the manuscript. NPSZ prepared the code for the forward problem of the inversion and contributed to the manuscript. The authors applied the SDC approach for the sequence of authors.

References

Abordán, A., Szabó, N.P.: Particle swarm optimization assisted factor analysis for shale volume estimation in groundwater formations. Geosci. Eng. **6**(9), 87–97 (2018)

Abordán, A., Szabó, N.P.: Particle swarm optimization based interval inversion of direct push logging data. In: MultiScience—XXXIII. microCAD International Multidisciplinary Scientific Conference University of Miskolc, 23–24 May (2019a)

Abordán, A., Szabó, N.P.: Selecting control parameters for the particle swarm optimization based factor analysis. Műszaki Földtudományi Közlemények, **88** (1), 134–140. (2019b)

Alberty, M., Hashmy, K.: Application of ULTRA to log analysis. In: SPWLA Symposium Transactions. Paper Z, pp. 1–17 (1984)

Asfahani, J.: Statistical factor analysis technique for characterizing basalt through interpreting nuclear and electrical well logging data (case study from Southern Syria). Appl. Radiat. Isot. **84**, 33–39 (2014)

Atlas, B.: OPTIMA: eXpress Reference Manual: Baker Atlas, Western Atlas International, Inc. (1996)

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, 61–85 (1987)

Benedetto, J.J., Czaja, W., Ehler, M.: Wavelet packets for time-frequency analysis of multispectral imagery. Int. J. Geomath. **4**, 137–154 (2013)

Cui, Y., Chen, Z., Zhu, X., et al.: Sequential and simultaneous joint inversion of resistivity and IP sounding data using particle swarm optimization. Earth Sci. **28**, 709 (2017)

Datta, D., Sen, M.K.: Estimating a starting model for full-waveform inversion using a global optimization method. Geophysics **81**(4), R211–R223 (2016)

Ding, S., Jiang, H., Li, J., et al.: Optimization of well placement by combination of a modified particle swarm optimization algorithm and quality map method. Comput. Geosci. **18**, 747 (2014)

Dobróka, M., Szabó, N.P.: The inversion of well log data using simulated annealing method: geosciences, Publications of the University of Miskolc, Series A (Mining), 59, 115–137 (2001)

Dobróka, M., Szabó, N.P., Tóth, J., Vass, P.: Interval inversion approach for an improved interpretation of well logs. Geophysics **81**, D155–D167 (2016)

Drahos, D.: Inversion of engineering geophysical penetration sounding logs measured along a profile. Acta Geod. Geophys. Hung. **40**, 193–202 (2005)

Drahos, D., Gyulai, Á., Ormos, T., et al.: Automated weighting joint inversion of geoelectric data over a two dimensional geologic structure. Acta Geod. Geophys. Hung. **46**, 309–316 (2011)

Fernández-Martínez, J.L., Fernández-Álvarez, J.P., García-Gonzalo, M.E., Ménendez-Pérez, C.O., Kuzma, H.A.: Particle swarm optimization (PSO): a simple and powerful algorithm family for geophysical inversion. In: SEG Technical Program Expanded Abstracts, pp. 3568–3571 (2008)

Grund, S.Z., Geiger, J.: Sedimentologic modelling of the Ap-13 hydrocarbon reservoir. Cent. Eur. Geol. **54**(4), 327–344 (2011)

Guneshwor, L., Eldho, T.I., Vinod Kumar, A.: Identification of groundwater contamination sources using meshfree RPCM simulation and particle swarm optimization. Water Resour. Manag. **32**, 1517 (2018)

Jarzyna, J., Bala, M., Cichy, A., Gądek, W., Gąsior, I., Karczewski, J., Marzencki, K., Stadtmüller, M., Twaróg, W., Zorski, T.: GeoWin—System for processing and interpretation of well logging. Proceedings of the Eighth Conference and Exhibition of EEGS-European Section, Aveiro, Portugal, September 8–12 (2002) (CD only)

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. Erlbaum, Mahwah (2007)

Kaiser, H.F.: The varimax criterion for analytical rotation in factor analysis. Psychometrika **23**, 187–200 (1958)

Kennedy, J., Eberhart, R.C.: Particle swarm optimization. Proc. IEEE Int. Conf. Neural Netw. **4**, 1942–1948 (1995)

Larionov, V.V.: Radiometry of Boreholes. Nedra, Moscow (1969). **(in Russian)**

Lawley, D.N., Maxwell, A.E.: Factor analysis as a statistical method. Stat. **12**, 209–229 (1962)

Marquardt, D.W.: Solution of non-linear chemical engineering models. Chem. Eng. Prog. **55**, 65–70 (1959)

Mayer C., Sibbit, A.M.: GLOBAL, a new approach to computer-processed log interpretation. In: Proceedings of the 55th SPE Annual Fall Technical Conference and Exhibition, paper 9341, pp. 1–14 (1980)

Menke, W.: Geophysical Data Analysis: Discrete Inverse Theory. Academic Press Inc., Cambridge (1984)

Narayan J.P., Yadav L.: Application of adaptive processing technique for the inversion of open hole logs recorded in oil fields of Indian basins. In: Proceedings of the SPG 6th International Conference and Exposition on Petroleum Geophysics, pp. 505–512 (2006)

Pace, F., Santilano, A., Godio, A.: Particle swarm optimization of 2D magnetotelluric data. Geophysics **84**(3), E125–E141 (2019)

Puskarczyk, E., Jarzyna, J.A., Wawrzyniak-Guz, K., et al.: Improved recognition of rock formation on the basis of well logging and laboratory experiments results using factor analysis. Acta Geophys. **67**, 1809–1822 (2019)

Sen, M.K., Stoffa, P.L.: Global Optimization Methods in Geophysical Inversion. Cambridge University Press, Cambridge (2013)

Serra, O.: Fundamentals of Well-Log Interpretation 1: The Acquisition of Logging Data: Developments in Petroleum Science, vol. 15. Elsevier, Amsteram (1984)

Shaw, R., Srivastava, S.: Particle swarm optimization: a new tool to invert geophysical data. Geophysics **72**(2), F75–F83 (2007)

Shi, Y., Eberhart, R.A.: modified particle swarm optimizer, evolutionary computation proceedings. In: The 1998 IEEE International Conference on IEEE World Congress on Computational Intelligence, pp. 69–73 (1998)

Szabó, N.P.: Shale volume estimation based on the factor analysis of well-logging data. Acta Geophys. **59**, 935–953 (2011)

Szabó, N.P., Dobróka, M.: Robust estimation of reservoir shaliness by iteratively reweighted factor analysis. Geophysics **82**(2), D69–D83 (2017)

Szabó, N.P., Dobróka, M.: Exploratory factor analysis of wireline logs using a float-encoded genetic algorithm. Math. Geosci. **50**, 317 (2018)

Tanos, P., Kovács, J., Kovácsné, Székely I., Hatvani, I.G.: Exploratory data analysis on the Upper-Tisza section using single and multi-variate data analysis methods. Cent. Eur. Geol. **54**(4), 345–356 (2011)

Tarantola, A.: Inverse Problem Theory and Methods for Model Parameter Estimation. Society of Industrial and Applied Mathematics (SIAM), Philadelphia (2005)

Wilkosz, M., Wawrzyniak-Guz, K.: An iterative inversion of Dual Induction Tool logs from thin-bedded sandy–shaly formations of the Carpathian Foredeep using a modified simulated annealing method. Acta Geophys. **67**, 1865 (2019)

Wyllie, M.R.J., Gregory, A.R., Gardner, L.W.: Elastic wave velocities in heterogeneous and porous media. Geophysics **21**, 41–70 (1956)

Zhu, P., Li, Z., Chen, M., et al.: Study on forward and inversion modeling of array laterolog logging in a horizontal/highly deviated well. Acta Geophys. **67**, 1307–1318 (2019)