DOI: 10.14750/ME.2020.007

UNIVERSITY OF MISKOLC

Faculty of Earth Science and Engineering

Department of Geophysics

GLOBAL OPTIMIZATION-BASED DATA PROCESSING METHODS FOR ADVANCED WELL LOGGING APPLICATIONS

PhD THESIS

by

Armand Abordán

Scientific supervisor:

Prof. Dr. Norbert Péter Szabó

MIKOVINY SÁMUEL DOCTORAL SCHOOL OF EARTH SCIENCES
Head of the Doctoral School: Prof. Dr. Péter Szűcs

Miskolc, 2020

HUNGARY

DOI: 10.14750/ME.2020.007

ADVISOR'S FOREWORD

for the (PhD) thesis

„GLOBAL OPTIMIZATION-BASED DATA PROCESSING METHODS FOR ADVANCED WELL LOGGING APPLICATIONS"

by Armand Abordán

The topic of the Candidate's thesis - inversion based geophysical data processing - is in focus of international research. The new method developments introduced by the Candidate in the thesis belong to the range of modern data processing tools of applied geophysics. The suggested global optimization based procedures are capable to derive the petrophysical parameters of geological formations in a highly accurate and reliable way, which is of high importance in today's petroleum exploration and several other fields of geosciences, too.

For advanced well logging applications, the author explores new opportunities for the further development of factor analysis and interval inversion. For improving the results of these multi-log interpretation procedures, he chooses the tool of global optimization. From the range of available techniques, he selects simulated annealing and particle swarm optimization, and uses them to enhance the before mentioned data processing methods. For improving the mathematical treatment of factor analysis, he develops alternative methods where the factor scores are estimated by the Metropolis simulated annealing and particle swarm optimization techniques. By using the results of these unique applications of global optimization tools, he establishes a regression model between the extracted first factor and the permeability of hydrocarbon-bearing formations. To validate the suggested regression relationship, the Candidate tests the method on Hungarian hydrocarbon bearing formations using in situ data. The permeability derived by the newly developed method is compared and successfully verified by deterministic modeling and core laboratory measurements.

In a hyperparameter estimation approach, he generalizes the developed particle swarm optimization based factor analysis in regard of some of its control parameters by automatically selecting them in an iterative procedure by simulated annealing. Thus, there is no need to set them manually in the initialization phase. Moreover, to provide a fully optimized solution for factor analysis, the Candidate suggest the simultaneous optimization of both the factor scores and factor loadings in an iterative procedure using particle swarm

DOI: 10.14750/ME.2020.007

optimization to even further decrease the misfit between the observed and calculated data. By processing in situ data, he proves the feasibility of the suggested methods in Hungarian groundwater formations.

An essential part of his thesis is connected to interval inversion, where he uses particle swarm optimization to eliminate the starting model dependence of the inversion procedure and to find the final solution he utilizes the damped least squares method, which enables the quantification of parameter estimation accuracy. To increase the overdetermination ratio of the inverse problem, he estimates shale volume prior to inversion by the developed factor analysis and then uses it as a known (fixed) parameter within inversion and thus significantly decreases the estimation error of the interval inversion method. This may have an impact in hydrocarbon reserve calculations, from the point of view of oilfield practice.

His continuous efforts towards scientific research, his creativity, and the results presented in this thesis prove the scientific knowledge and the suitability of the Candidate for independent research. In my opinion, the Candidate's results, especially those of related machine learning tools assisted inversion approaches are worth to be published in ranked international journals of applied geophysics.

I certify that this dissertation contains only valid data and the presented results are representing the Candidate's own work. In my opinion, it is fully adequate in scope and quality required by the Mikoviny Sámuel Doctoral School of Earth Sciences. Based on the above, I support and recommend the public defense of the thesis and the award of the PhD title.

07/05/2020, Miskolc

Dr. Norbert Péter Szabó
University full professor

DOI: 10.14750/ME.2020.007

CONTENT

DOI: 10.14750/ME.2020.007

INTRODUCTION

Applied geophysics provides several methods for the quantitative evaluation of subsurface geological structures and finding mineral resources. 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 aquifers and hydrocarbon reservoirs. To properly assess the economic value of these formations, wireline logging methods must be applied (Serra 1984). Usually, electrical and elastic rock properties are measured along with nuclear and dimensional measurements (i.e., caliper log) in the wellbore. Then the recorded well logs can be used to derive the geometry (e.g., structural dip, layer thickness) and the petrophysical properties of the rock formations (e.g., porosity, shale volume, water saturation and matrix volumes), which can enable the quantitative assessment of hydrocarbon and mineral resources.

Wireline logging began in 1927 when the Schlumberger brothers recorded the first resistivity log in France's Pechelbronn field. Shortly after the spontaneous potential and natural gamma-ray logs were introduced as well. Plotting these measurements versus depth, the separation of permeable hydrocarbon bearing layers from non-productive ones became possible. This was a simple qualitative, curve-shape recognition-based approach to log analysis, however due to its effectiveness it was quickly adopted by the hydrocarbon industry. A great advancement in wireline logging was achieved when Archie (1942) developed an empirical relationship for estimating the water saturation from porosity and resistivity data, which allowed the quantitative analysis of well logs for the first time. By the 1960s, apparent porosity was available from three independent sources (density, neutron and sonic logs), which were first interpreted using crossplot techniques (Asquith and Krygowski 2004). Then with the advances in computer technology, well logs were recorded digitally and processed by computers (e.g., the CYBERLOOK interpretation program by Schlumberger) based on the crossplot techniques and on simultaneous equation solutions. In the 1980s, new statistics-based log analysis approaches were established for the interpretation of well logs. Their computer implementations allowed the log analyst to provide the response equations and the geophysical model instead of still relying on deterministic approaches as earlier programs and now these could also quantify the uncertainties of derived petrophysical parameters. The well-known examples are Schlumberger's GLOBAL (Mayer and Sibbit 1980), OPTIMA by Baker Hughes (Ball et al.

1987) and Gearhart's ULTRA (Alberty and Hashmy 1984). These systems provide a so-called inversion based well log analysis which is superior to the conventional deterministic methods. The latter derive the petrophysical parameters separately from each other by single well log analysis, while in case of inversion based methods all the available set of measurements are jointly inverted to derive the model parameters simultaneously in a more accurate and reliable way. The petrophysical unknowns are derived in an iterative procedure by optimizing the misfit between the measured and calculated well logs.

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 sets a limit 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 probe 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 an arbitrary long 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 for local inversion, this can be increased to at least 6 or more depending on the length of the processed depth interval and the desired resolution that is controlled by the number of series expansion coefficients used.

Geophysical inverse problems are conventionally solved by linearized (known as gradient-based) methods (Menke 1984, Tarantola 2005). These methods have several

drawbacks, e.g., the result of the inversion procedure is greatly sensitive to the 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 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 optimal 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, which was made possible by the great advancement of computer performance. Some of the most commonly used global optimization methods in geophysics are simulated annealing, the genetic algorithm and the particle swarm optimization (Sen and Stoffa 2013, Holland 1975, Pace et al. 2019). These soft computing methods are capable to search through a greater extent of the possible solutions than the linearized inversion methods without trapping in a local minimum.

Inversion method development has been in the focus at the Department of Geophysics, University of Miskolc for decades. My aim in this thesis is to develop new methods for the advanced interpretation of well logs that rely on these previous findings. I intend to utilize the advantageous properties of the before mentioned global optimization techniques and incorporating them into factor analysis (Szabó 2011) and the so-called interval inversion method (Dobróka 1995) to develop new data processing tools for a more reliable estimate for petrophysical properties of geological formations. By introducing these new data processing methods, I offer new possibilities for the advanced analysis of wireline logging data.

In the first part of my thesis, I improve the mathematical treatment of factor analysis. It was previously shown that by factor analysis of well logging data, one can derive some of the petrophysical properties of subsurface formations (Szabó and Dobróka 2013, Szabó et al. 2012). Instead of the conventionally used maximum likelihood approach for the calculation of factor scores (Bartlett 1937), I suggest a method, where the values of factor scores are estimated in a globally optimized procedure by decreasing the misfit between the measured and calculated data. I test the feasibility of this newly developed method on measured well logging data.

With the help of the developed globally optimized factor analysis, I determine the correlation relationship between the first factor log extracted from well logging datasets and the decimal logarithm of intrinsic permeability of hydrocarbon-bearing formations. I test the

applicability of the suggested regression model on well logs measured in Hungarian hydrocarbon exploratory wells.

To eliminate the starting model dependence of the interval inversion method originally developed at the Department of Geophysics, I solve the inverse problem by a newly developed metaheuristic approach. Furthermore, for increasing the reliability and accuracy of the estimated model parameters, I derive the shale volume by the newly developed globally optimized factor analysis and then incorporate it into the interval inversion procedure to increase its overdetermination ratio and the accuracy of inversion estimation.

I further develop the suggested globally optimized method of factor analysis (inner loop) by the automated selection of some of its control parameters by simulated annealing in an outer iteration loop. In this way, I generalize the improved method of factor analysis in regard of some if its control parameters, so there is no need to set them empirically in the initialization phase based on suggestions from literature. This way they can be selected automatically for the given optimization task.

For further optimizing the results of the developed factor analysis, I suggest an improvement in which not just the factor scores are optimized by a global optimization technique but the factor loadings as well in the same procedure. This simultaneous optimization can reach a better fit between the measured and calculated well logs and thus offer a better solution overall. All the new method developments and improvements presented in this thesis are tested on measured well logging data and validated by core derived petrophysical parameters where available.

DOI: 10.14750/ME.2020.007

1. FACTOR ANALYSIS OF WIRELINE LOGGING DATA

This multivariate statistical tool is used to reduce the number of measured variables into a smaller number of uncorrelated parameters, while keeping most of the information contained in the original variables to help the data interpretation and to possibly reveal hidden information (Lawley and Maxwell 1962). Therefore, in well log applications it can be effectively applied because the acquired datasets are usually fairly large due to the numerous types of applied logging tools. In wireline logging applications, these new variables are called factor logs, which can be related to petrophysical parameters of the investigated geological formations through regression analysis (Szabó 2011). Thus giving independent estimations to these parameters that can be used to get a more detailed picture of the studied formations or to improve the results of further data processing methods. The added advantage of the factor analysis based well log processing is that it utilizes several type of logs to infer the desired petrophysical parameters, while e.g., simple deterministic methods rely only on one type of log to derive a given parameter. As a statistical method, it can also process such data, e.g., the caliper log, for which an exact mathematical relation with petrophysical properties is not available.

In this chapter, I suggest a global optimization based solution of factor analysis, in which the factor scores are estimated by simulated annealing by decreasing the misfit between the measured and calculated data. With the help of existing correlation relationships and by regression analysis, I relate the first factor extracted from wireline logging datasets to shale volume of different geological formations. By testing the developed method on actual well logging data, I show that the method is capable to provide an in situ estimate to shale content along arbitrary depth intervals, which may improve the geological model of investigated areas.

For initializing 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

DOI: 10.14750/ME.2020.007

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

Then **D** is decomposed as

$$\mathbf{D}=\mathbf{F}\mathbf{L}^{\mathrm{T}}+\mathbf{E}, \tag{2}$$

where **F** denotes the *N*-by-*R* matrix of factor scores and *R* is the number of extracted factors. Quantity **L** is the *S*-by-*R* matrix of factor loadings and **E** denotes the *N*-by-*S* matrix of approximation errors. Based on Eq. (2), 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 **F**. Given that the factors are linearly independent, the correlation matrix of the standardized original data can be written as

$$\mathbf{R}=N^{-1}\mathbf{D}^{\mathrm{T}}\mathbf{D}=\mathbf{L}\mathbf{L}^{\mathrm{T}}+\boldsymbol{\Psi}, \tag{3}$$

where $\boldsymbol{\Psi}$ denotes a diagonal matrix of error variances, which are independent of the common factors in matrix **F** normally explaining the largest part of data variance. Then the factor loadings can be estimated by a non-iterative estimation method (Jöreskog 2007)

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

where $\boldsymbol{\Gamma}$ is the diagonal matrix of the first *R* number of sorted eigenvalues of the sample covariance matrix $\boldsymbol{\Sigma}$, and the first *R* number of eigenvectors are in matrix $\boldsymbol{\Omega}$ and **U** denotes an arbitrarily chosen *R*-by-*R* orthogonal matrix. In traditional approaches, the factor scores are calculated by the maximum likelihood method. By assuming the hypothesis of linearity, the equation

$$\mathbf{F}^{\mathrm{T}}=\left(\mathbf{L}^{\mathrm{T}}\boldsymbol{\Psi}^{-1}\mathbf{L}\right)^{-1}\mathbf{L}^{\mathrm{T}}\boldsymbol{\Psi}^{-1}\mathbf{D}^{\mathrm{T}}, \tag{5}$$

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, it usually remains sensitive to the level and distribution of instrumental (data) noises and the localities of the objective function to be minimized.

DOI: 10.14750/ME.2020.007

1.1 Simulated annealing driven factor analysis (FA-SA)

I alter the traditional method of factor analysis outlined in Chapter 1 by using the classical simulated annealing (SA) algorithm (Metropolis et al. 1953) to calculate the factor scores. The combined method has been chosen to call as FA-SA (Abordán and Szabó 2018a). The workflow of the suggested well-log-analysis method is summarized in Fig. 1. Here, it should be noted that for regression analysis the shale volume (V_{sh}) of the investigated formations generally could be determined from other sources as well (e.g., core data, natural gamma-ray intensity log) not just the SP-log as in this case.

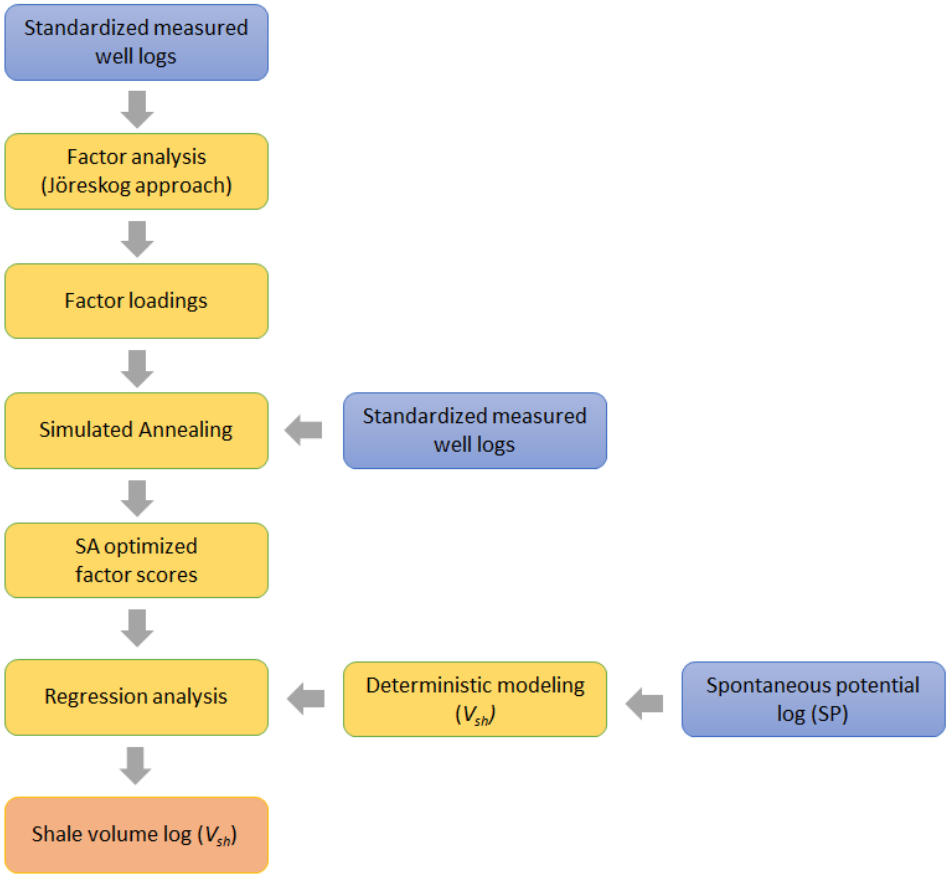

Fig. 1 The flowchart of the globally optimized factor analysis of well logs

In the first step of this optimization problem, the model of factor analysis defined in Eq. (2) is reformulated

$$\mathbf{d} = \tilde{\mathbf{L}}\mathbf{f} + \mathbf{e}, \tag{6}$$

where **d** denotes the *SN* length vector of measured well logging (standardized) data, $\tilde{\mathbf{L}}$ is the *NS*-by-*NR* matrix of factor loadings, **f** is the *RN* length vector of factor scores and **e** is the *SN* length vector of errors. Starting the procedure, $\tilde{\mathbf{L}}$ is estimated by Eq. (4) and then

rotated with the *varimax* algorithm (Kaiser 1958) for getting more meaningful factors. Then the vector of factor scores **f** is estimated by the algorithm of SA. To solve the inverse problem and estimate the factor scores, an objective function needs to be defined, the minimization of which finds the optimal solution. I choose the objective function, named energy function in the terminology, based on the L_2 norm as

$$E = \sqrt{\frac{1}{NS}\sum_{i=1}^{NS}(d_i^{(m)} - d_i^{(c)})^2}\,, \tag{7}$$

where $\mathbf{d}^{(m)}$ and $\mathbf{d}^{(c)}$ denote the measured and calculated (standardized) well-logging data vectors, respectively. In the modified approach of factor analysis, the term of $\tilde{\mathbf{L}}\mathbf{f}$ represents the calculated data and **d** denotes the measured data. The former term allows the estimation of the theoretical values of well logs, which can be considered as the solution of the forward problem. During the iterative procedure, the values of factor loadings are kept fixed to reduce the CPU time, and only the factor scores are updated. In each iteration, a randomly generated number (b) is added to any of the factor scores in vector **f**

$$\mathrm{f}_j^{(new)} = \mathrm{f}_j^{(old)} + b \qquad (j = 1,\ldots,RN)\,, \tag{8}$$

where parameter b is smaller or equal to the maximal perturbation ($b_{\max}$) that has to be specified in the initialization of the FA-SA algorithm. In the current procedure, I select the initial values of factor scores as zero. If the energy difference of factor models - estimated in two subsequent iterations (ΔE) according to Eq. (7) - is negative (i.e., better fit between the observed and calculated data), the new model is accepted and the process is continued with the new energy state. However, in the reverse case (if $\Delta E>0$), the probability of acceptance is given by the formula $P_a=\exp(-\Delta E/T)$, where T is the current temperature of the artificial system with no physical meaning. The new factor model is accepted only when a randomly generated number from the range of 0 and 1 is smaller than P_a. This acceptance rule for new energy states is referred to as the Metropolis criterion (Metropolis et al. 1953). This is a fundamental part of the FA-SA algorithm as it prevents the search from being stuck in a local minimum of function E in Eq. (7). During the annealing process, the temperature of the system is reduced iteratively according to Geman and Geman (1984) to guarantee that the global optimum is found

$$T^{(new)} = T_0 / \log_{10}(1+q)\,, \tag{9}$$

where q denotes the number of iterations already computed and T_0 is the initial temperature of the artificial system. The maximal perturbation term ($b_{\max}$) is also reduced according to $b_{\max}=b_{\max}\cdot\varepsilon$, where ε is an arbitrary chosen constant. These steps are repeated in each iteration

until the pre-defined maximal number of iteration steps is reached and then the value of factor scores in the last iteration step are accepted as the solution. The factors estimated by the FA-SA algorithm are directly used to reveal hidden petrophysical information from the well-logging datasets as first suggested by Szabó (2011). For this purpose, it is advantageous to scale the factor scores

$$F_1^{'} = F_{1,\min}^{'} + \frac{F_{1,\max}^{'} - F_{1,\min}^{'}}{F_{1,\max} - F_{1,\min}} (F_1 - F_{1,\min}), \tag{10}$$

where F_1 and $F_1^{'}$ are the estimated and scaled value of the first factor in a given depth point, $F_{1,\max}$ and $F_{1,\min}$ are the limits of the first factor log, and $F_{1,\max}^{'}$ and $F_{1,\min}^{'}$ are the new arbitrary limits of the first factor log. Szabó and Dobróka (2013) showed that the shale volume (in percent) correlates strongly to the first factor log (F_1) scaled into the range of 0 and 100, and their relationship can be well approximated by

$$V_{sh} = 2.76 e^{0.037 F_1}, \tag{11}$$

where V_{sh} is shale volume and the constant added for shifting the function equals zero. I verify the validity of Eq. (11) in Alaska costal sediments in the next sub-chapters. Szabó and Dobróka (2017) also confirmed the validity of Eq. (11) by a series expansion based interval inversion procedure. For such exponential relationship, the strength of correlation between the shale volume of the investigated interval and the first scaled factor can be quantified by the rank correlation coefficient (Spearman 1904).

1.1.1 Field test I. - Milky River Formation

I test the newly developed FA-SA method in two hydrocarbon exploratory wells drilled in Alaska, USA (Fig. 2). First, I investigate Well-1 that penetrated the Milky River Formation, which is mainly built up of conglomerates, sandstone and mudstone and was formed in a shallow marine environment in the Pliocene age (Wiley 1986). It has high porosity and high permeability. As the input of factor analysis, I utilized the natural gamma-ray intensity (GR), bulk density (RHOB), borehole caliper (CAL), deep induction resistivity (RILD), spontaneous potential (SP) and neutron-porosity (NPHIS) logs covering a depth interval of 266.5 ft with a sampling space of 0.5 ft. At the beginning of the statistical procedure, the factor loadings are calculated by Jöreskog's non-iterative approach using Eq. (4). Table 1 contains the resultant factor loadings representing the impact of the different well logs on the extracted factors for Well-1.

DOI: 10.14750/ME.2020.007

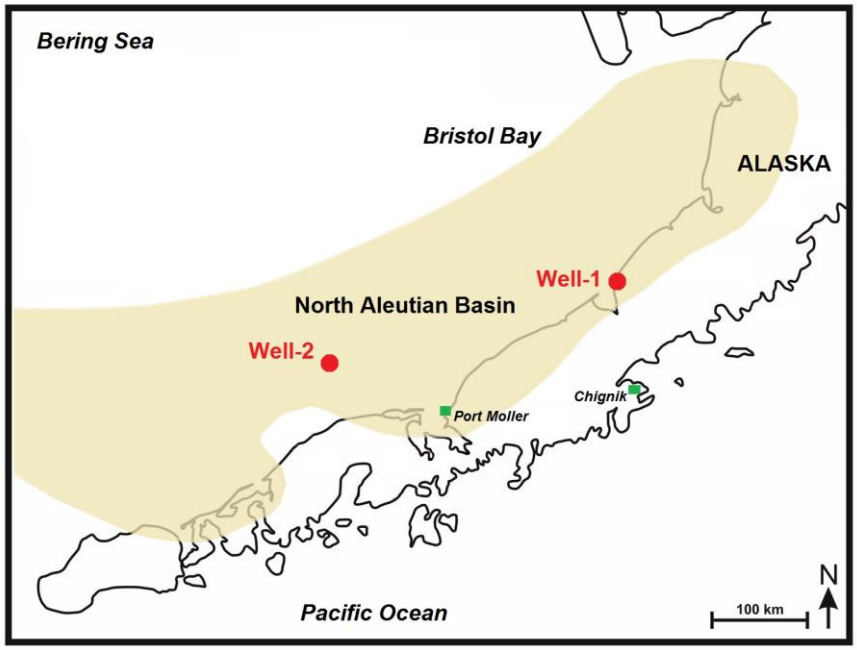

Fig. 2 The location of the investigated wells in the North Aleutian Basin of Alaska, and regional distribution of Tertiary-age sedimentary rocks (shaded area)

As Table 1 shows, the deep resistivity and the spontaneous potential logs have the highest loads on the first factor and unusually the load of the natural gamma-ray intensity is only –0.5213. The first factor log is considered as a lithological indicator that generally strongly correlates to the natural gamma-ray intensity log (Szabó 2011) therefore, we would expect the natural gamma-ray intensity to have a higher loading, but in this formation, the abundance of lithic detritus causes a mineralogical overlap between sandstones and mudstones. Therefore, the natural gamma-ray tool which response is mainly due to the radioactive mineral content recorded only little or no difference between sandstone and mudstone and this causes the relatively low factor loading of the natural gamma-ray intensity on the first extracted factor. The second factor log is in strong correlation with the bulk density log and in a strong negative correlation with the neutron porosity log, while the third extracted factor is mainly influenced by the caliper and natural gamma-ray logs.

Table 1 Rotated factor loadings estimated in Well-1 – Milky River Formation

Well logs	***Factor 1***	***Factor 2***	***Factor 3***
CAL	–0.3892	–0.1484	–0.5221
GR	–0.5213	–0.0205	0.4602
SP	0.9538	–0.0707	0.0320
RHOB	–0.0883	0.7609	0.0144
RILD	0.9386	–0.0780	0.0134
NPHIS	0.0969	–0.7397	–0.0612

DOI: 10.14750/ME.2020.007

Once the factor loadings are calculated and fixed, then the factor scores are estimated by the SA algorithm. For three factors, the number of unknowns to be estimated is 1,602 (3 factor × 534 measured depth points). In the step of initialization, I define the objective function according to Eq. (7). Based on preliminary runs, I set the initial temperature (T_0) to 10^{-7}, cooling schedule according to Eq. (9), maximal parameter perturbation (b_{max}) to 0.5, perturbation reduction parameter (ε) to 0.98 and the maximal number of iteration steps to 150,000. Figure 3 shows the decrease of the difference between the measured and calculated well logs by the iterations steps.

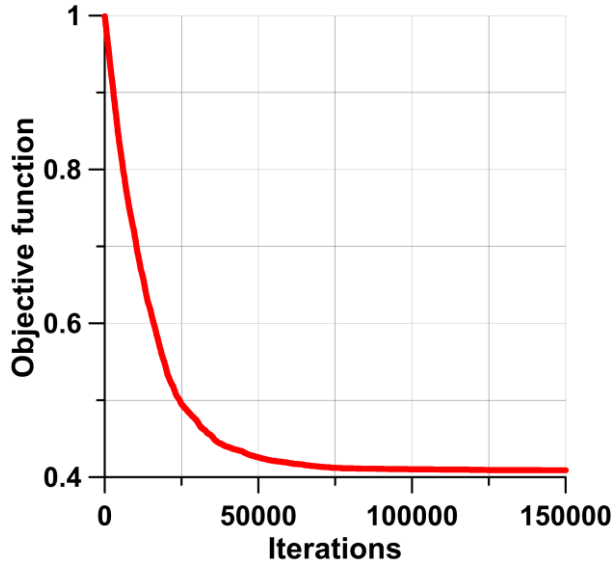

Fig. 3 Development of convergence of the FA-SA procedure for Well-1

The difference between the measured and calculated data reached the minimum at about eighty thousand iterations, which took less than 10 seconds on a quad-core based workstation. The continuous decrease of the value of the objective function indicates the highly stable nature of the FA-SA method. Figure 4 presents the relation between the first factor and the shale volume estimated by the FA-SA method utilizing Eq. (11) (where a better fit could be reached by refining the regression coefficients). The Spearman's rank correlation coefficient of 0.96 between the first factor and the shale volume indicates a strong non-linear relationship between the variables.

DOI: 10.14750/ME.2020.007

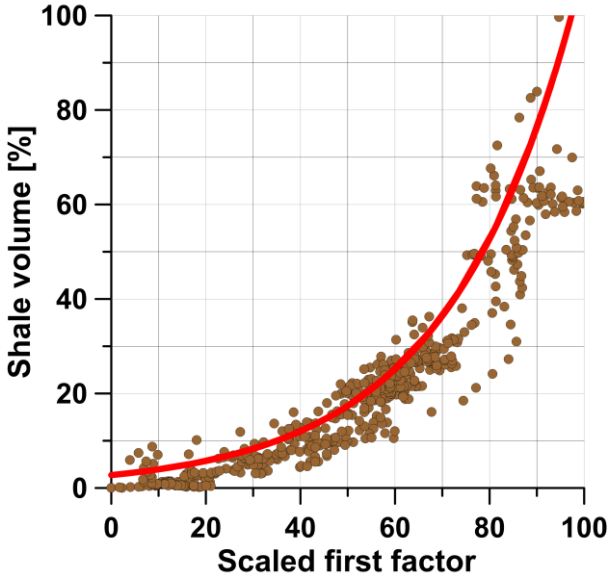

Fig. 4 Regression relation (red line) between the scaled first factor and the shale volume for Well-1. Dots represent the SP log derived shale volumes

The regression coefficients of the suggested exponential relation agree well with earlier studies (Szabó and Dobróka 2013), the same coefficients in several formations show consistent results and confirm the applicability of the method at different areas. Shale volume derived from the FA-SA method is shown in Fig. 5. On tracks 1-6, the standardized (input) well logs (black solid line) and the calculated logs (red dashed line) can be seen. Track 7 is the scaled first factor log, while next to it on the right the shale volume calculated from the first factor log (red solid line) is compared to that of calculated by deterministic modeling using the SP log (purple dashed line).

DOI: 10.14750/ME.2020.007

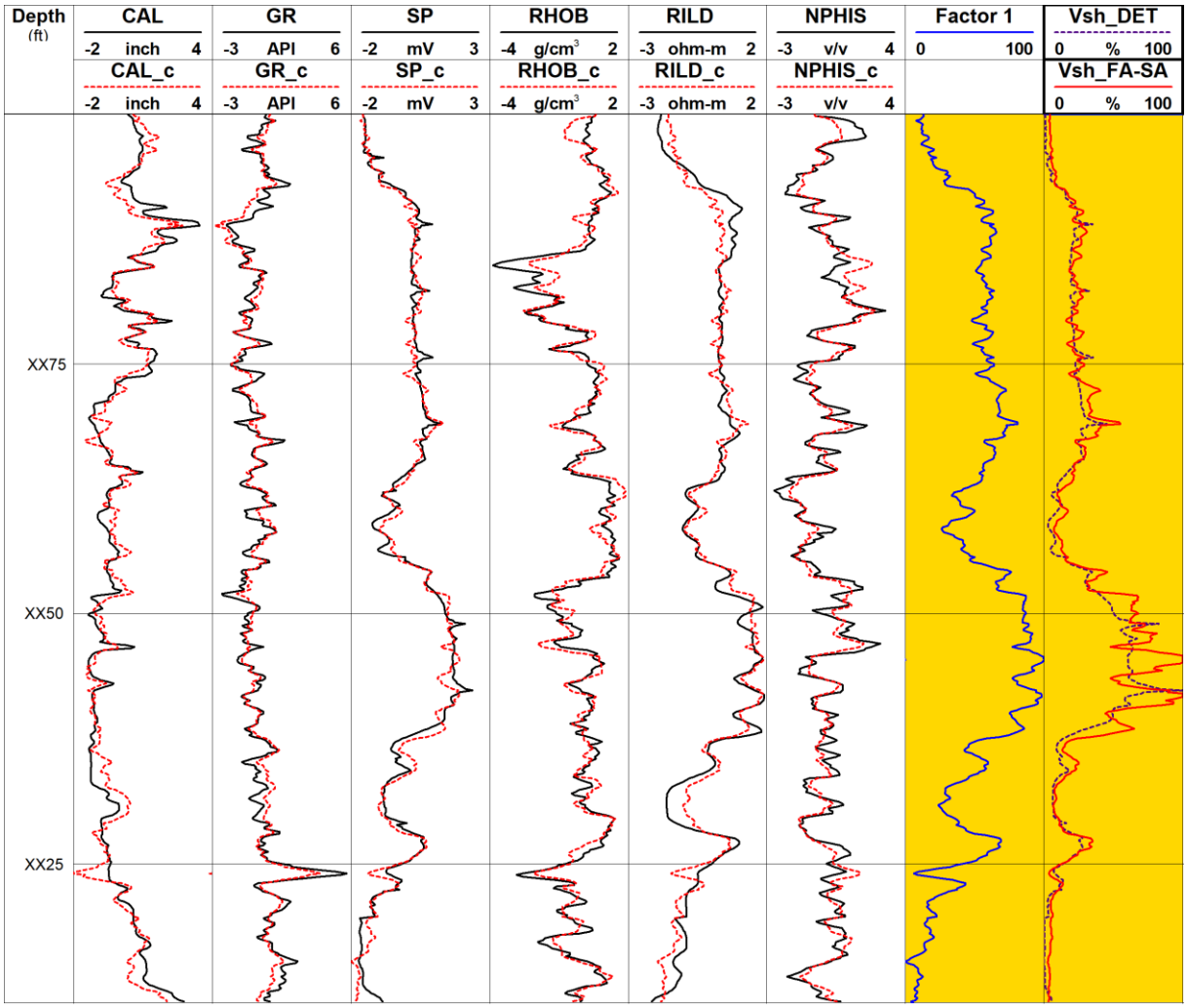

Fig. 5 The results of FA-SA method for Well-1. Measured (black) and calculated well logs (red dashed line) in tracks 1-6, scaled first factor log in track 7 and the resultant shale volumes in track 8

The fit between the observed and calculated well logs is quite good. A better fit would be possible by increasing the number of extracted factors. Theoretical well logs represented by red dashed lines were directly calculated as the multiplication of the rotated factor loadings and factor scores. The shale volume calculated by the FA-SA method also agrees well with the results of the SP-log based deterministic analysis. These results indicate the applicability of the FA-SA method for a more detailed shale volume estimation in sedimentary formations.

1.1.2 Field test II. - Bear Lake Formation

In case of Well-2, the penetrated horizon is the Bear Lake Formation, which is of Miocene age, and consists mainly of sandstones, conglomerates and thin mudstones (Finzel

et al. 2009). It formed in neritic to tidal flat environments and it can be characterized by high porosity and moderate permeability. The inputs of the factor analysis are the borehole caliper (CAL), sonic interval transit-time (DT), natural gamma-ray intensity (GR), deep induction resistivity (RILD), neutron-porosity (NPHIS), bulk density (RHOB) and spontaneous potential (SP) logs. The investigated interval is 130 ft long with a sample spacing of 0.5 ft. The FA-SA procedure is initialized by the same control parameters as in case of Well-1. Table 2 contains the factor loadings related to three extracted factors in case of Well-2.

Table 2 Rotated factor loadings calculated in Well-2 – Bear Lake Formation

Well logs	***Factor 1***	***Factor 2***	***Factor 3***
CAL	–0.6222	–0.6081	–0.1344
GR	0.6529	0.5218	0.2648
SP	0.9338	–0.0429	0.0081
RHOB	–0.1935	0.5788	0.4545
RILD	0.8967	–0.0588	–0.1758
NPHIS	0.2424	–0.8167	–0.1616
DT	–0.2733	–0.7565	0.0344

In this case, not just the deep resistivity and the spontaneous potential logs have high loadings on the first factor, but both the gamma-ray intensity and the caliper logs, too. I also implemented the caliper log into the procedure, because washouts and the thickening of the mudcake might have strong relation to lithology. As we can see in this example, the caliper log affects highly the lithology-sensitive first factor log. The second factor log again correlates well with the bulk density and the neutron porosity logs as in case of the Milky River Formation along with the acoustic, natural gamma-ray and caliper logs. The third factor only correlates moderately with the density log. Figure 6 shows the decrease of the value of energy function by the optimization of the 783 unknowns (3 factors × 261 measured depth points). It reached a minimum at about sixty thousand iterations in less than 5 seconds. In this case, the FA-SA method again proves to be very stable in the iteration process.

DOI: 10.14750/ME.2020.007

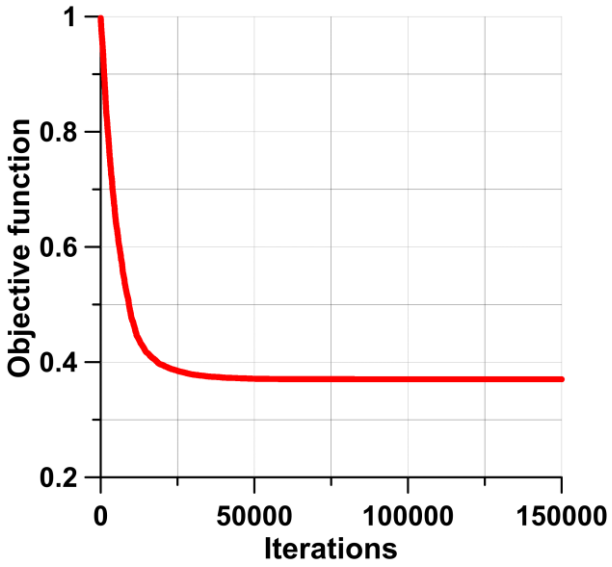

Fig. 6 The decrease of model energy by the iteration steps for Well-2

The regression function of the first factor and the shale volume estimated by the FA-SA method is illustrated in Fig. 7, which refers to a strong relation also for Well-2, although a few data are outlying from the model.

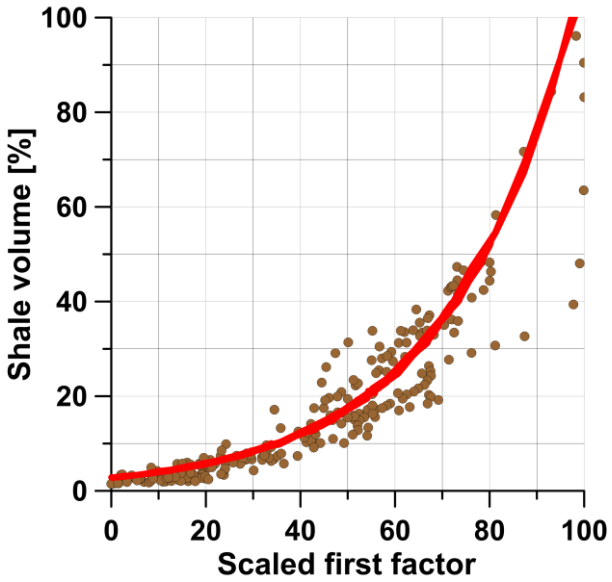

Fig. 7 Regression relation (red line) between the first factor and the shale volume for Well-2. Dots represent the SP log derived shale volumes

The rank correlation coefficient is 0.95 between the extracted factor and that of shale volume, which again proves the validity of the exponential relationship defined in Eq. (11). The interpretation results of the FA-SA method applied to Well-2 is shown in Fig. 8. The first seven columns from the left represents the input well logs (black solid line) and the calculated logs (red dashed line), the next is the scaled first factor log (blue solid line), and

DOI: 10.14750/ME.2020.007

in the last track the shale volume calculated from the first factor log (red solid line) can be compared to the shale volume calculated by deterministic modeling (purple dashed line). Here the calculated logs again fit the measured data acceptably well. In conclusion, the shale volume calculated by the FA-SA method is consistent with the shale volume calculated by deterministic modeling.

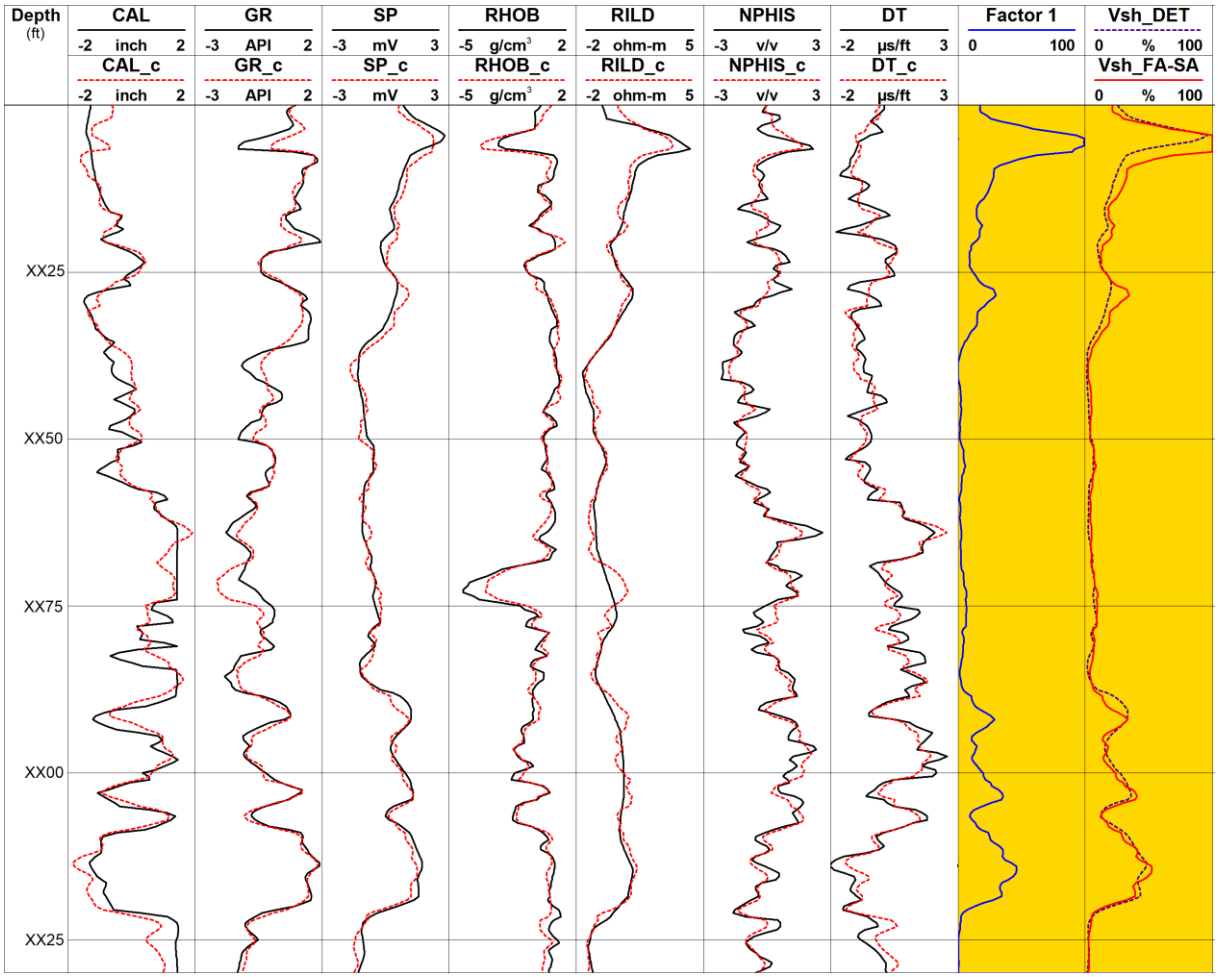

Fig. 8 The results of the FA-SA procedure for Well-2. Measured (black) and calculated well logs (red dashed line) in tracks 1-7, scaled first factor log in track 8 and the resultant shale volumes in track 9

1.2 Particle swarm optimization based factor analysis (FA-PSO)

For finding the optimal values of factor scores, I have combined the particle swarm optimization (PSO) algorithm with factor analysis (Abordán and Szabó 2018b). It is fundamentally different from the algorithm of SA in a sense that it improves a set of solution candidates in the iteration process rather than just one solution. Hence a more reliable (and

stable) solution is expected with the expense of somewhat higher CPU times. The workflow of the FA-PSO well-log-analysis method is summarized in Fig. 9.

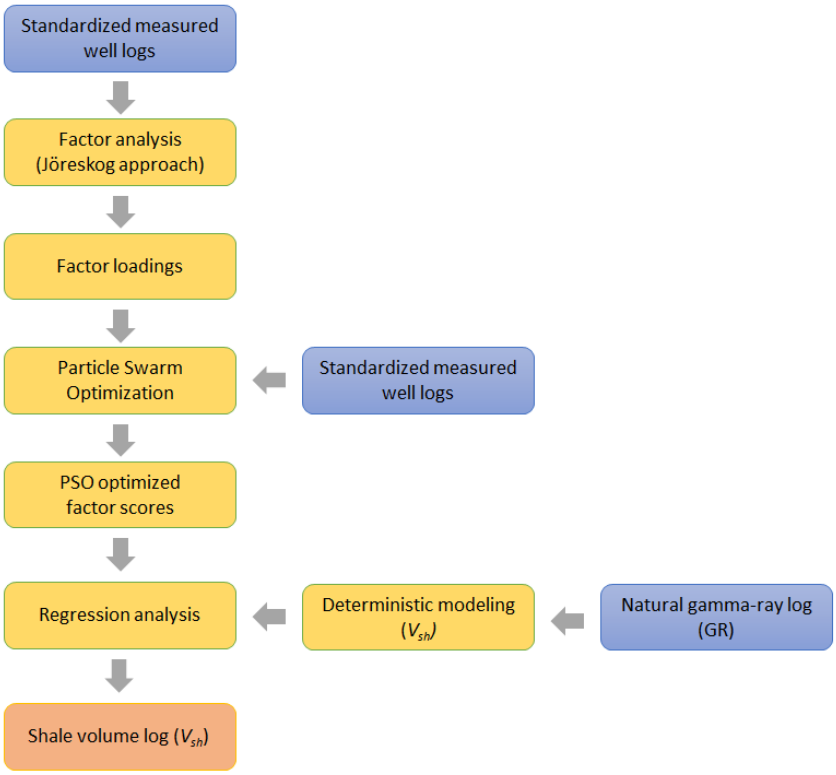

Fig. 9 The flowchart of the particle swarm optimization based factor analysis of well logs

PSO is a metaheuristic technique that is inspired by the social behavior of bird flocking and fish schooling. The basic method was originally developed by Kennedy and Eberhart (1995). It is widely used for its easier implementation and in some cases higher efficiency compared to other metaheuristic optimization approaches such as the genetic algorithm (Holland 1975). PSO can be effectively used for non-linear inverse problems with large search domains. It utilizes a swarm of particles randomly generated within 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 some objective function according to Eqs. (12)-(13)

$$\mathbf{x}_i(t+1) = \mathbf{x}_i(t) + \mathbf{v}_i(t+1)\,, \tag{12}$$

$$\mathbf{v}_i(t+1) = w\mathbf{v}_i(t) + r_1c_1(\mathbf{p}_i(t) - \mathbf{x}_i(t)) + r_2c_2(\mathbf{g}(t) - \mathbf{x}_i(t))\,, \tag{13}$$

where i=1,2,...,L and L is the size of the swarm, iteration steps are denoted by t and t=1,...,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 **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 *i*th 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). In Eq. (13), parameters r_1 and r_2 are uniformly distributed random numbers from the range of 0 and 1. In the same equation, w represents an inertia weight that was introduced by Shi and Eberhart (1998) to control the optimization procedure more effectively. Figure 10 shows how particle $\mathbf{x}_i(t)$ gets to its new position $\mathbf{x}_i(t+1)$ by utilizing its own personal best position $\mathbf{p}_i(t)$, the best position found by the whole swarm $\mathbf{g}(t)$ and its velocity $\mathbf{v}_i(t)$.

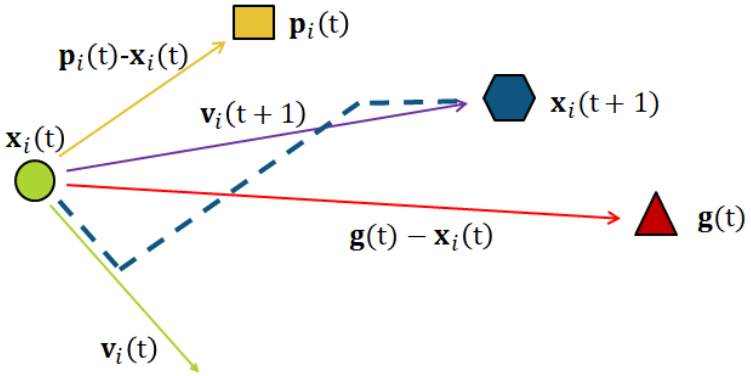

Fig. 10 Illustration of the searching mechanism of the PSO algorithm

The algorithm of PSO is utilized to find the optimal value of factor scores in a similar manner as SA in the previous subchapter. In the first step of finding the optimal values of the factor scores **f,** a random population of 90 particles with uniform distribution is generated within the search space, which can be defined by solving Eq. (5) for the factor scores. In this case, the boundaries of factor scores are found between -6 and 6. In Eq. (13) both c_1 and c_2 are set as 2 in the whole iteration process. The inertia weight w is set in each iteration step according to Feng et al. (2007). They suggested the use of chaotic descending inertia weight, which is calculated in 3 steps: first a random number z is generated in the range between 0 to 1. Then by logistic mapping, z is set according to $z=4z(1–z)$ and finally

$$w = (w_1 - w_2)(q_{\max} - q) / q_{\max} + w_2 z\,, \tag{14}$$

where w_1 and w_2 are the initial and the final value of inertia weight respectively (w_1=0.3 and w_2=0.08), q_{max} and q are the maximal number of iteration steps and current iteration step respectively, in this case q_{max} =3,000. Then in each iteration step, the positions of the

particles are updated according to Eqs. (12)–(13), and the previously defined objective function in Eq. (7) is recalculated with the new values of factors **f**. The particles are updated in each iteration step until a pre-defined number of iteration steps is reached and then the final values of the factor scores are accepted as the solution. The factor scores calculated by the newly developed FA-PSO algorithm can also be directly used to reveal hidden petrophysical information from the wireline logging data set. It was previously shown that for shorter intervals (max. 100–150 m), shale volume V_{sh} (given in percent) often correlates linearly to the first factor log (F_1) scaled into the range of 0 and 100 (Szabó and Dobróka 2011, Szabó 2011) as

$$V_{sh} = aF_1 + b\,, \tag{15}$$

where a is the slope of the regression line and b is the intercept.

1.2.1 Field test – Baktalórántháza

The FA-PSO method is tested in a thermal water well (Well-A) in Baktalórántháza, north-east Hungary (Fig. 11). The well is 1,197 m deep and in the upper portion of it upper Pleistocene aquifers with varying grain sizes can be found. The mainly horizontal porous layers are bordered by shales. In the range of 100 to 160 m, mostly sandy layers are located, below it 10 to 15 m thick coarse-grained beds were deposited. The investigated interval is 93.1 m long, from 100 m to 193.1 m. As the input of factor analysis, the natural gamma-ray intensity (GR), gamma-gamma intensity (GG), shallow resistivity (RS), spontaneous potential (SP) and neutron-neutron intensity (NN) logs are utilized.

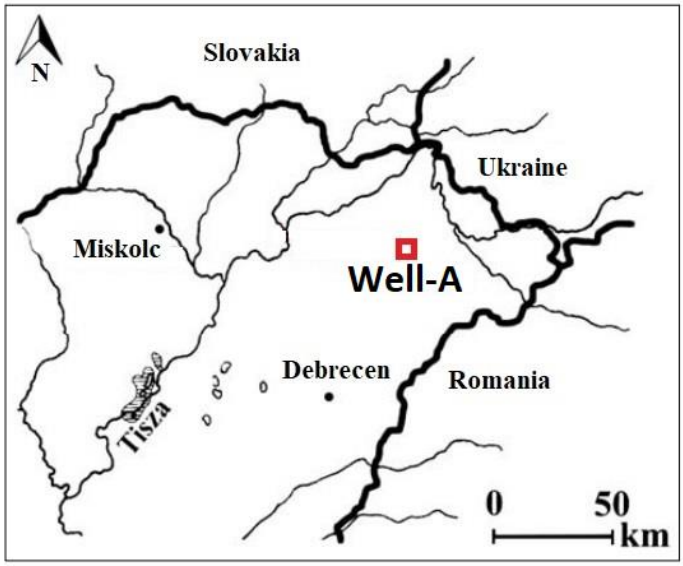

Fig. 11 Location of the water-producing well used for testing the FA-PSO procedure

First, the factor loadings are calculated by Jöreskog's non-iterative approach for three factors using Eq. (4). Table 3 contains the resultant factor loadings representing the correlation between the different well logs and the extracted factors. By calculating three factors for this 93.1 m long section, which has 932 measured depth points, the vector of factors **f** is composed of 2,796 scores. This means that the algorithm of PSO has to find the global optimum by finding the optimal values of 2,796 variables along the borehole.

Table 3 Rotated factor loadings estimated by the FA-PSO method in Well-A

Well logs	***Factor 1***	***Factor 2***	***Factor 3***
GR	0.7515	−0.1836	0.0127
SP	−0.0380	−0.7396	−0.0893
NN	0.0580	−0.1184	−0.5976
GG	−0.5930	−0.0948	0.3062
RS	−0.4955	0.6537	−0.0081

As Table 3 indicates, the natural gamma-ray log has the highest loading and both the shallow resistivity and gamma-gamma logs have moderate loadings on the first factor and the loadings of the spontaneous potential and neutron-neutron logs are negligible. However, it should be noted that here gamma-gamma intensity and neutron-neutron intensity logs are processed, which would require further calibration to represent bulk density and neutron-porosity and that could change their loadings on the extracted factors. It can be concluded that the first factor is mainly influenced by lithological properties of groundwater formations. While the second factor log correlates with the spontaneous potential and shallow resistivity logs. The third factor only correlates negatively with the neutron-neutron log.

In the next step, the factor scores are estimated by the particle swarm optimization algorithm as detailed above. Figure 12 shows the continuous decrease of the objective function values that is defined in Eq. (7). It reaches a minimum at about fifteen hundred iterations, which took approximately thirty seconds on a quad-core workstation. The steady decrease of the objective function indicates that the method is highly stable.

DOI: 10.14750/ME.2020.007

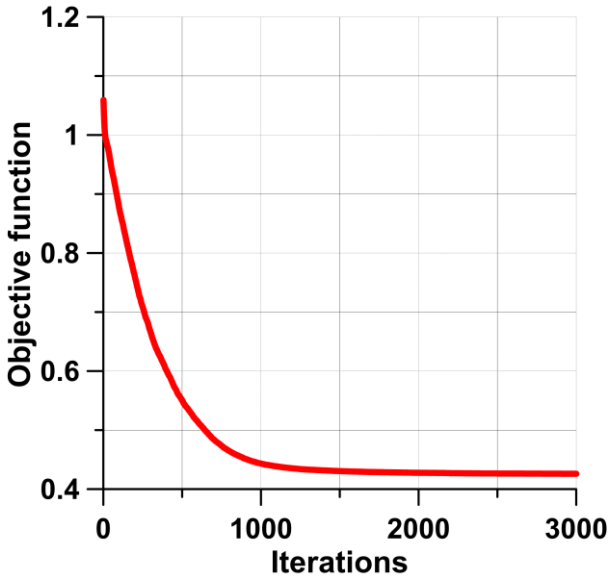

Fig. 12 Development of convergence of the FA-PSO procedure in Well-A

Figure 13 presents the linear regression relation between the scaled first factor and the shale volume of the investigated groundwater formation based on Eq. (15). In the given example, I have found the regression coefficients with 95% confidence bounds to be a=0.611 [a_{min}=0.5863, a_{max}=0.6353] and b=-1.299 [b_{min}= -2.435, b_{max}= -0.1632].

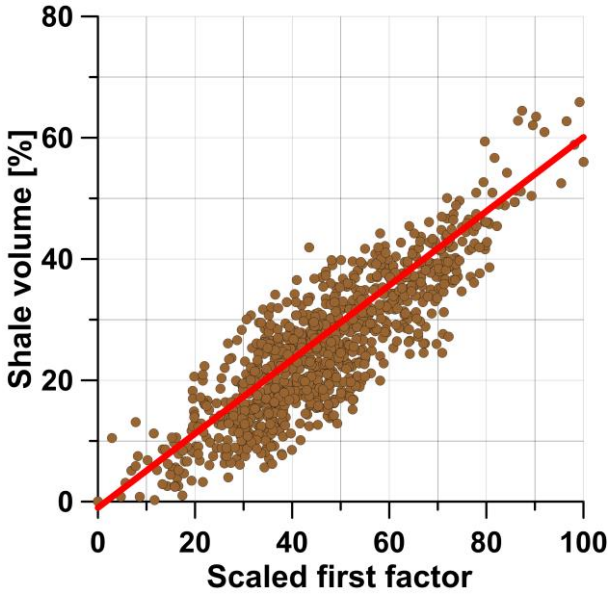

Fig. 13 Regression relation (red line) between the first factor and shale volume in Well-A. Dots represent the shale volumes derived by GR log based deterministic modeling

The Pearson's correlation coefficient for the first factor and the shale volume of the investigated interval is 0.86, which indicates a strong linear relationship and verifies the applicability of Eq. (15) in the well. The observed well logs and those calculated by the newly developed FA-PSO method are shown in Fig. 14. On tracks 1 to 5, the standardized

(input) well logs (black solid line) and the calculated logs (red dashed line) can be seen. Track 6 represents the extracted first (scaled) factor log (blue).

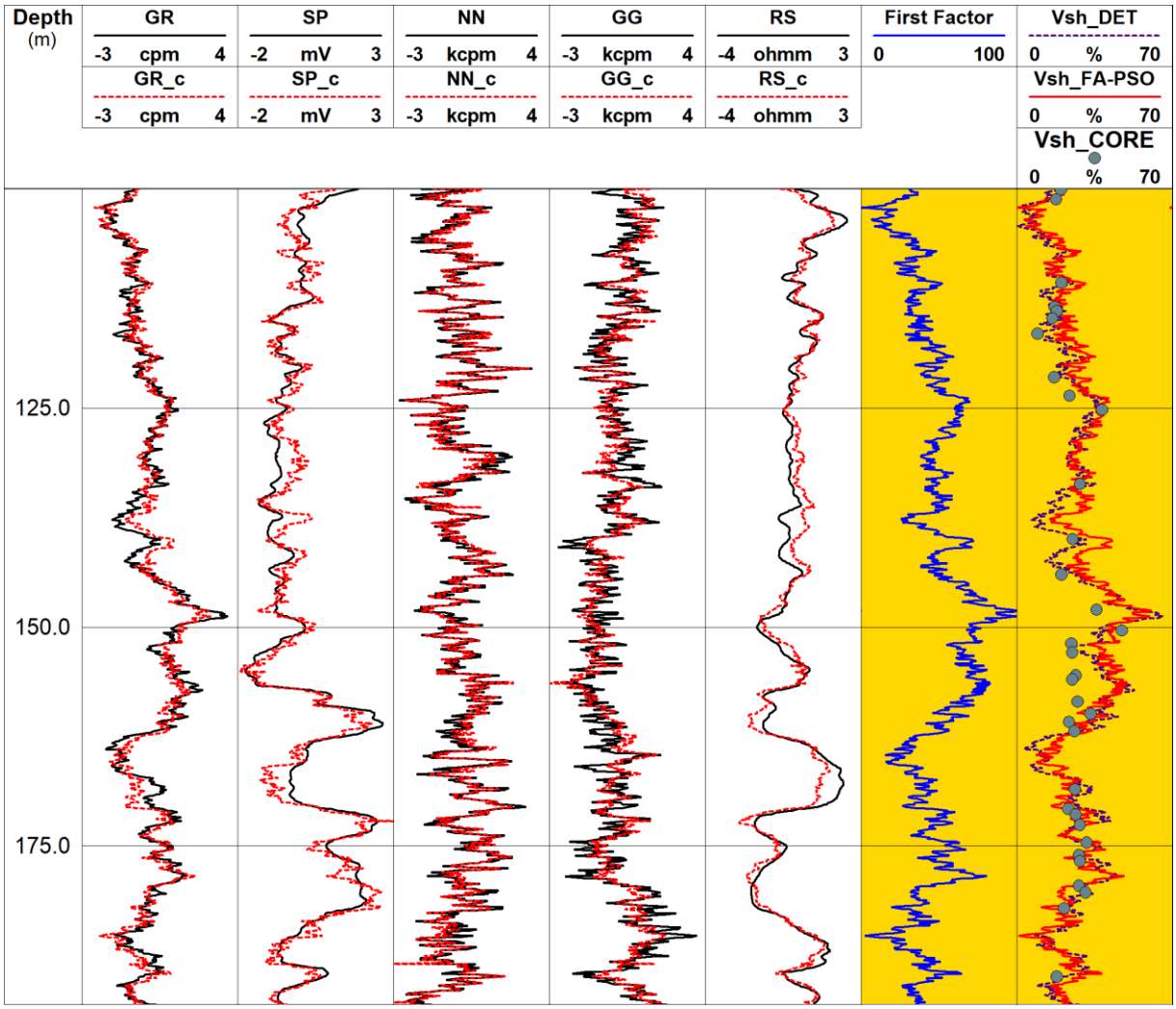

Fig. 14 The results of the FA-PSO procedure in Well-A. The measured (black) and calculated well logs (red dashed line) in track 1-5, the scaled first factor in track 6 and the resultant shale volumes in track 7

The fit between the measured and calculated data is quite good. The theoretical well logs drawn by red dashed lines are directly calculated as the multiplication of the rotated factor loadings and factor scores from Eq. (6). The result of traditional factor analysis was earlier confirmed by core laboratory measurements in the same well (Szabó et al. 2014). On track 7, the shale volume calculated from the PSO-derived first factor log (solid red line) is compared to core data (dots) and to that calculated by deterministic modeling suggested by Larionov (1969) using the GR log (purple dashed line), where first the gamma ray index is calculated as

DOI: 10.14750/ME.2020.007

$$I_{GR} = \frac{GR - GR_{MIN}}{GR_{MAX} - GR_{MIN}}, \tag{16}$$

where GR is the gamma-ray reading of the formation, GR_{MIN} is the gamma ray reading in clean sand and GR_{MAX} is the reading in shale. Then shale volume is calculated as

$$V_{sh} = 0{,}083\,(2^{3{,}7 I_{GR}} - 1). \tag{17}$$

It can be seen that the shale volume calculated by the developed statistical method also agrees well with the results of deterministic analysis and core data. These results confirm the applicability of the FA-PSO method for shale volume estimation in water bearing formations.

1.3 Summary

The globally optimized solution for the factor scores was first suggested by Szabó (2016) with the computationally expensive genetic algorithm. For finding the optimal solution in a faster procedure, I have developed the simulated annealing based factor analysis (FA-SA) and an efficient population-based solution by particle swarm optimization (FA-PSO), both of which is shown to work effectively. Furthermore, another population based optimization method, the so-called invasive weed optimization (Mehrabian and Lucas 2006) was earlier shown to effectively solve the problem of factor analysis (Abordán 2018). Here, it should be noted that by implementing the very fast simulated annealing (VFSA) algorithm (Ingber 1989) into FA-SA, the CPU time of the developed method could be further reduced.

With a global optimization approach, the measured well logs are transformed into factor logs in a reliable way. One finds that the first factor log calculated by the suggested methods strongly correlate with the independently calculated shale volume of the investigated formations. However, it should be noted, that as it is a general method, it can be used to evaluate more types of lithologies than shown in my thesis. The implementation of the suggested metaheuristic methods allows also for the estimation of the theoretical values of well logs, which neglects the preliminary knowledge of zone parameters and other petrophysical information. Both of the applied metaheuristic proves to be very stable in the iteration process and delivers the results within a minute for sections of around a hundred meters. Therefore, the newly developed FA-SA and FA-PSO statistical methods can provide a reliable (independent well-log-analysis based) estimate to shale volume in practice. Once a regression relationship is found between the first factor extracted from a well logging

DOI: 10.14750/ME.2020.007

dataset and the shale volume of a given formation in a specific area, the relationship might be used in neighboring wells as well. This might reduce the operating costs, such as the need for taking more core samples to determine shale volume.

Thesis 1.

I have developed global optimization based solutions of factor analysis that are capable to estimate the factor scores by means of simulated annealing and particle swarm optimization, separately. Factor analysis is solved as an inverse problem, the optimal values of factor scores are estimated by finding the best fit between the measured and calculated data. By the above manner, I have developed the FA-SA and FA-PSO algorithms. The two developed methods are both capable to derive the factor scores from wireline logging datasets in a reliable way, while factor loadings are fixed during the procedure. By processing in situ datasets, I proved the feasibility of the suggested methods in different measurement sites, and estimated the shale volume directly from the global optimization-derived factor logs.

DOI: 10.14750/ME.2020.007

2. ESTIMATING PERMEABILITY FROM THE FACTOR SCORES DERIVED BY FA-PSO

Petrophysical parameters are usually estimated in formation evaluation from several different sources for a better reservoir modeling. As a new alternative, I utilize well logging data and derive the permeability of different hydrocarbon formations directly from the factor scores estimated by the improved particle swarm optimization assisted factor analysis (FA-PSO). The flowchart of the suggested well-log-analysis method is shown in Fig 15.

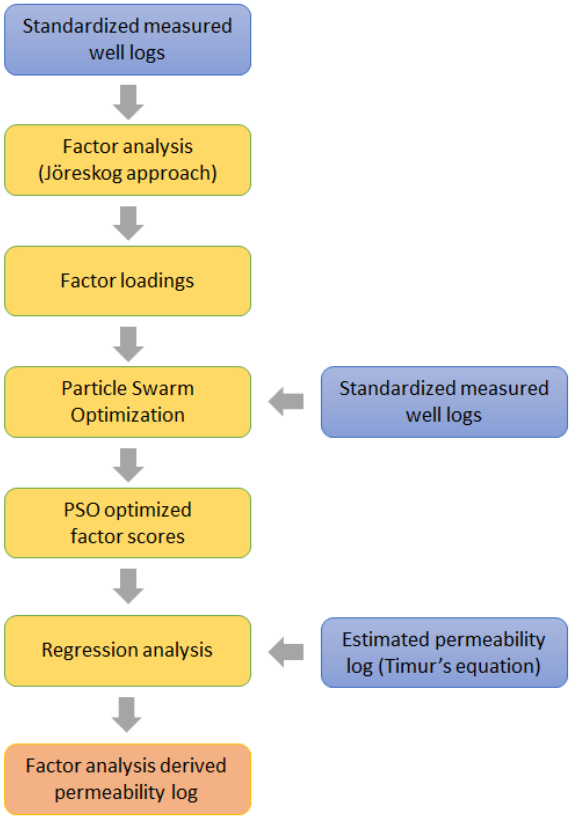

Fig. 15 The flowchart of the FA-PSO based permeability estimation

As it was shown, the first extracted factor correlates well with the shale volume of different formations. In this chapter, a strong correlation is shown between the decimal logarithm of permeability and the first factor log, which is partly due to the fact that permeability is to some extent inversely proportional to shale volume in clastic formations (Schön and Georgi 2003, Revil and Cathles 1999). The method is tested in Hungarian hydrocarbon wells. For estimating permeability (K) from wireline logging dataset, usually deterministic approaches are used in practice which rely on porosity (Φ) and irreducible water saturation ($S_{w,irr}$) data. The most frequently applied method uses the formula of Timur (1968)

$$K = 0.136 \frac{\Phi^{4.4}}{S_{w,irr}^2}. \tag{18}$$

The suggested statistical method is tested in two wells drilled in the Pannonian basin in Hungary. It is shown that the first factor log derived by the developed FA-PSO method scaled into the range 0 to 1 can be directly used to estimate permeability along the investigated intervals. In my study, the non-linear relationship between the first factor log extracted from the wireline logging dataset and decimal logarithm of permeability is found at different Hungarian oilfield wells in the form of

$$\lg(K) = a(1 - F_1^b) + c. \tag{19}$$

I seek the regression coefficients *a*, *b* and *c* by regression analysis. For Well-I, the relationship is found by utilizing the permeability data deterministically estimated by the EXPRESS-CLASS system used by MOL Plc. (Baker Atlas 1996). For Well-II the available permeability data measured on core samples is used for determining the regression relationship between the first extracted factor and the decimal logarithm of permeability along the investigated formation.

2.1 Hungarian field test I.

The applicability of the improved method of factor analysis is first tested in a hydrocarbon well (Well-I) drilled in the Pannonian basin, Hungary. The processed well logs by the FA-PSO method include the borehole caliper (CAL), natural gamma-ray intensity (GR), neutron-porosity (NPHIS), bulk density (DEN), deep resistivity (RD), potassium-thorium product (KTH), acoustic interval time (DT) and spontaneous potential (SP) logs. The processed interval is 290 m long and the sampling interval is 0.2 m. At the start of the procedure, the input well logs are standardized and collected into the column vector **d** 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 resultant rotated loadings are collected in Table 4.

DOI: 10.14750/ME.2020.007

Table 4 Rotated factor loadings estimated in Well-I

Well logs	*Factor 1*	*Factor 2*	*Factor 3*
CAL	0.2708	−0.1205	0.2009
NPHIS	0.3806	−0.6440	−0.3984
DT	0.1183	−0.8691	−0.0374
GR	0.8671	−0.4167	−0.0653
KTH	0.7252	−0.2522	0.0295
RD	−0.1203	0.3440	0.5603
SP	0.4543	−0.2669	0.1949
DEN	0.9412	0.0402	−0.0639

The first factor is strongly related to the natural gamma-ray, potassium-thorium index and bulk density logs that make it a good lithological indicator, which is consistent with my previous applications (Table 1, 2, 3). The second factor log correlates with the acoustic and neutron-porosity logs, while the third factor is mainly influenced by the deep resistivity log. Once the loadings are determined, the factor scores can be optimized by the algorithm of PSO as detailed in Chapter 1.2. First, an initial solution is given by solving Eq. (5) for the factor scores, thus the search space of the PSO process is defined as [-12 to 12]. Since three factors were calculated, the number of unknowns is 4,353 (3 factor × 1451 depth points). For estimating their optimal values, 90 particles are initialized within the pre-defined search space. They search for the solution by moving around in the search domain according to Eq. (12) and (13) to minimize the objective function in 5,000 iteration steps earlier defined in Eq. (7). The other control parameters (c_1, c_2, w) of the algorithm are unchanged from the ones used in Chapter 1.2. At the end of the optimization phase, the data distance reached 0.53 (Fig. 16) within a hundred seconds.

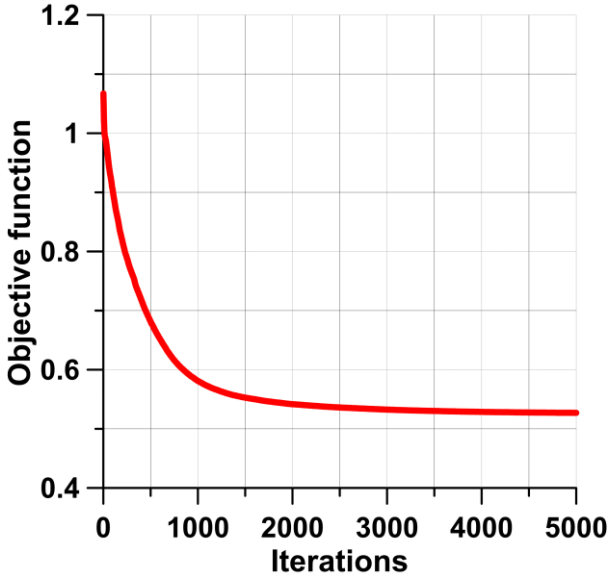

Fig. 16 Convergence of the FA-PSO procedure in Well-I

DOI: 10.14750/ME.2020.007

To find the regression relationship between the first factor log found by the developed FA-PSO method and the decimal logarithm of permeability, first it needs to be scaled in the range of 0 to 1 using Eq. (10). For regression analysis, the deterministically (Timur equation) derived permeability data is used.

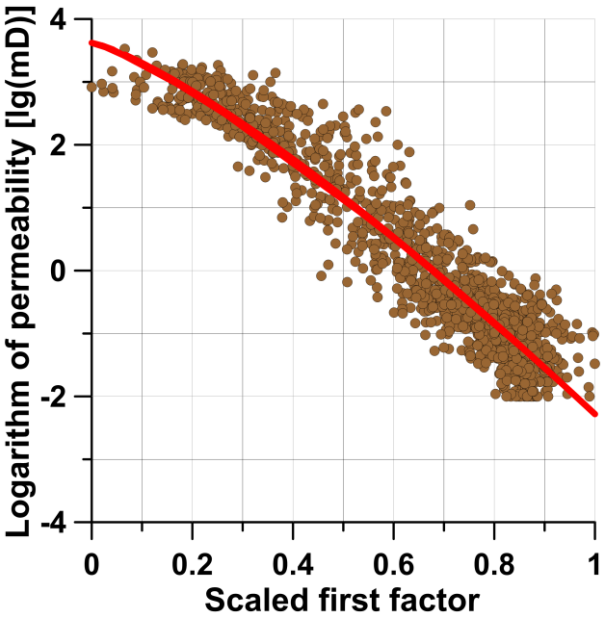

Fig. 17 Regression relation between the scaled first factor and the decimal logarithm of permeability for Well-I

In this example, the regression coefficients for Eq. (19) with 95% confidence bounds are found to be a=5.939 [a_{min}=5.807, a_{max}=6.071], b=1.267 [b_{min}= 1.182, b_{max}= 1.352] and c=2.321 [c_{min}=2.241, c_{max}=2.402]. The correlation relation between the scaled first factor and the decimal logarithm of permeability is plotted in Fig. 17. The rank correlation coefficient of -0.77 between the first factor and the decimal logarithm of permeability indicates a strong inverse non-linear relationship.

The input well logs of the procedure are plotted in Fig. 18 in the first 8 tracks, and as the last track indicates the fit between the FA-PSO method derived permeability and that derived by the Baker Atlas CLASS-EXPRESS software routinely used by MOL Plc. is fairly consistent along the investigated interval.

DOI: 10.14750/ME.2020.007

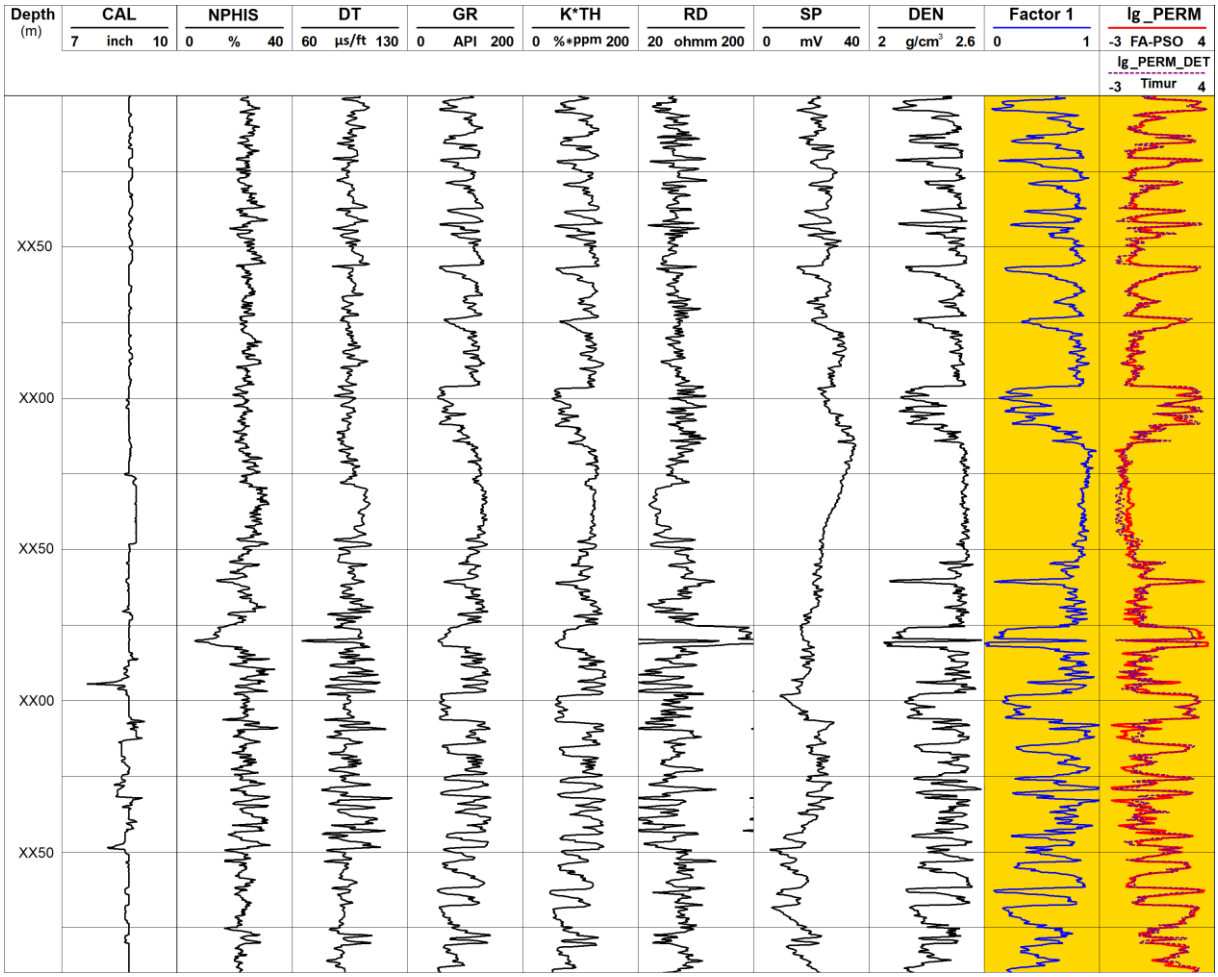

Fig. 18 The input logs of FA-PSO (track 1-8), the scaled first factor log (track 9), the derived permeability log and the CLASS-EXPRESS derived permeability logs (track 10)

2.2 Hungarian field test II.

I test the applicability of the FA-PSO based permeability estimation method using core plug data as well. Here the processed well logs are from another Hungarian hydrocarbon well (Well-II) drilled in the Pannonian basin, Hungary, which include the natural gamma-ray intensity (GR), neutron-porosity (NPHIS), bulk density (DEN), deep resistivity (RD), potassium-thorium product (KTH), acoustic interval time (DT) and spontaneous potential (SP) logs. From core plugs, the equivalent liquid permeability (K_L) is available, which is derived from the measured gas permeability (K_G) as

$$K_L = \frac{K_G}{1+(b/P_m)}, \tag{20}$$

where b is a constant for a particular gas in a given rock type and P_m is the mean pressure (Klinkenberg 1941). Here, it should be noted that the laboratory measurements on core plugs

and wireline logging are carried out under two very different conditions. The temperature and pressure of the environment and the water saturation of rock all might be different. Core plugs can provide only local information about permeability, while logging data is influenced by a larger extent of the rock formation, therefore the misfit between the FA-PSO derived permeability and that of core data can be significant. The reliability of the method also relies on the accuracy of the documented core plug locations. To improve the correlation coefficient for relationship (19), I tend to concentrate the observed information into one factor. I carried out the regression analysis for finding the relationship between the first factor and the decimal logarithm of permeability on the locations from where the core plugs were taken. In the investigated interval, core data is available at 54 locations, therefore logging data is collected from these locations and standardized to serve as the input of factor analysis. The processed interval is 18 m long. Here again, factor loadings are estimated by Eq. (4) only for one factor. The calculated and rotated loadings are shown in Table 5.

Table 5 Rotated factor loadings estimated in Well-II

Well logs	***Factor 1***
NPHIS	−0.0178
DT	0.8697
GR	0.8734
KTH	0.9825
RD	0.9595
SP	−0.3745
DEN	0.3058

The extracted factor from the wireline logging dataset shows a strong correlation with the acoustic, natural gamma-ray, potassium-thorium and deep resistivity logs. After the loadings are calculated, the factor scores are optimized by the algorithm of PSO in the same manner as detailed for Well-I. First, an initial solution is given by solving Eq. (5) for the factor scores, which defines the search space of the FA-PSO method as [-3 to 3]. Since only one factor is calculated, the number of unknowns adds up to only 54 (1 factor × 53 depth points). For estimating their optimal values, 45 particles are initialized in the pre-defined search space. They search for the solution by moving around in the search domain according to Eq. (12) and (13) to minimize the objective function in 200 iteration steps earlier defined in Eq. (7). The other control parameters (c_1, c_2, w) of the algorithm are unchanged from the ones used in Chapter 1.2. At the end of the optimization, the data distance reached 0.66 (Fig. 19) in less than a second.

DOI: 10.14750/ME.2020.007

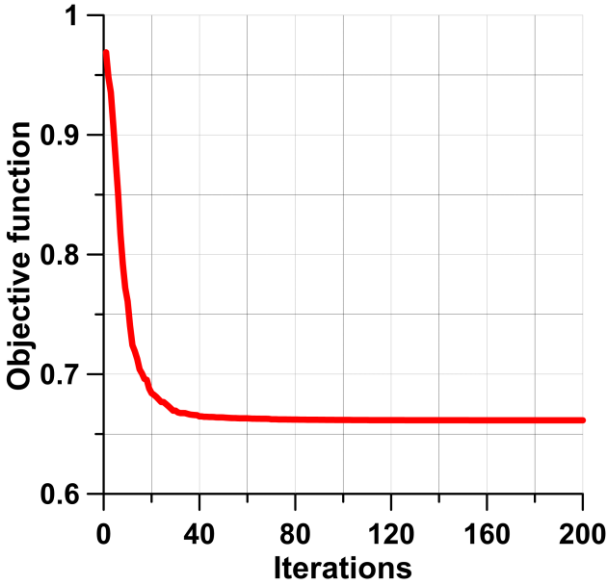

Fig. 19 Convergence of the FA-PSO procedure in Well-II

For establishing the regression relationship between the first factor and the decimal logarithm of permeability, first it needs to be scaled into the range of 0 to 1 according to Eq. (10). For regression analysis, the equivalent liquid permeability data is used. In this example, the regression coefficients for Eq. (19) with 95% confidence bounds are found to be a=2.818 [a_{min}=2.364, a_{max}=3.271], b=2.696 [b_{min}= 1.804, b_{max}= 3.588] and c=0.1259 [c_{min}=-0.331, c_{max}=0.583]. The non-linear relationship between the first scaled factor log and the decimal logarithm of permeability is plotted in Fig. 20.

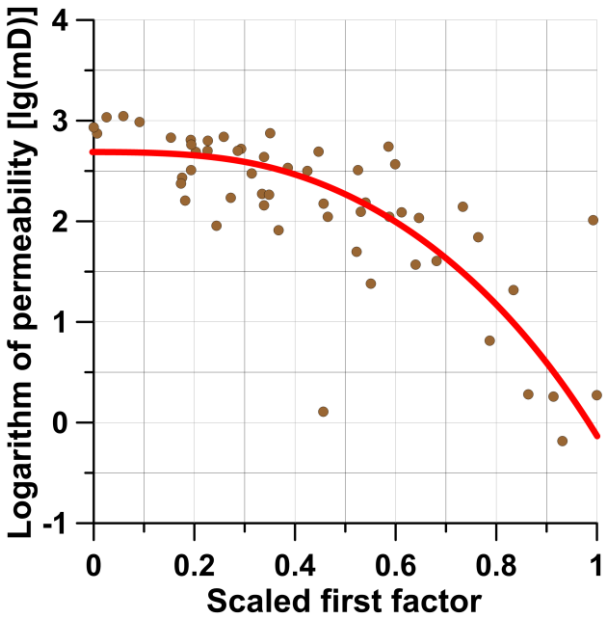

Fig. 20 Regression relation between the scaled first factor and the decimal logarithm of permeability measured on core plugs in Well-II

DOI: 10.14750/ME.2020.007

The rank correlation coefficient of -0.78 between the first scaled factor and the decimal logarithm of permeability indicates a strong negative non-linear relationship. As it was expected, the fit between the permeability derived from well log data by the improved method of factor analysis and that from core plugs is only moderate, however the trend and changes in permeability in the investigated interval is well followed by the FA-PSO based permeability estimation (Fig. 21). The input well logs of factor analysis are plotted in the first 7 tracks, the scaled first factor log is in track 8, and the last track contains the decimal logarithm of permeability derived by the developed FA-PSO method using Eq. (19) and the core plug data.

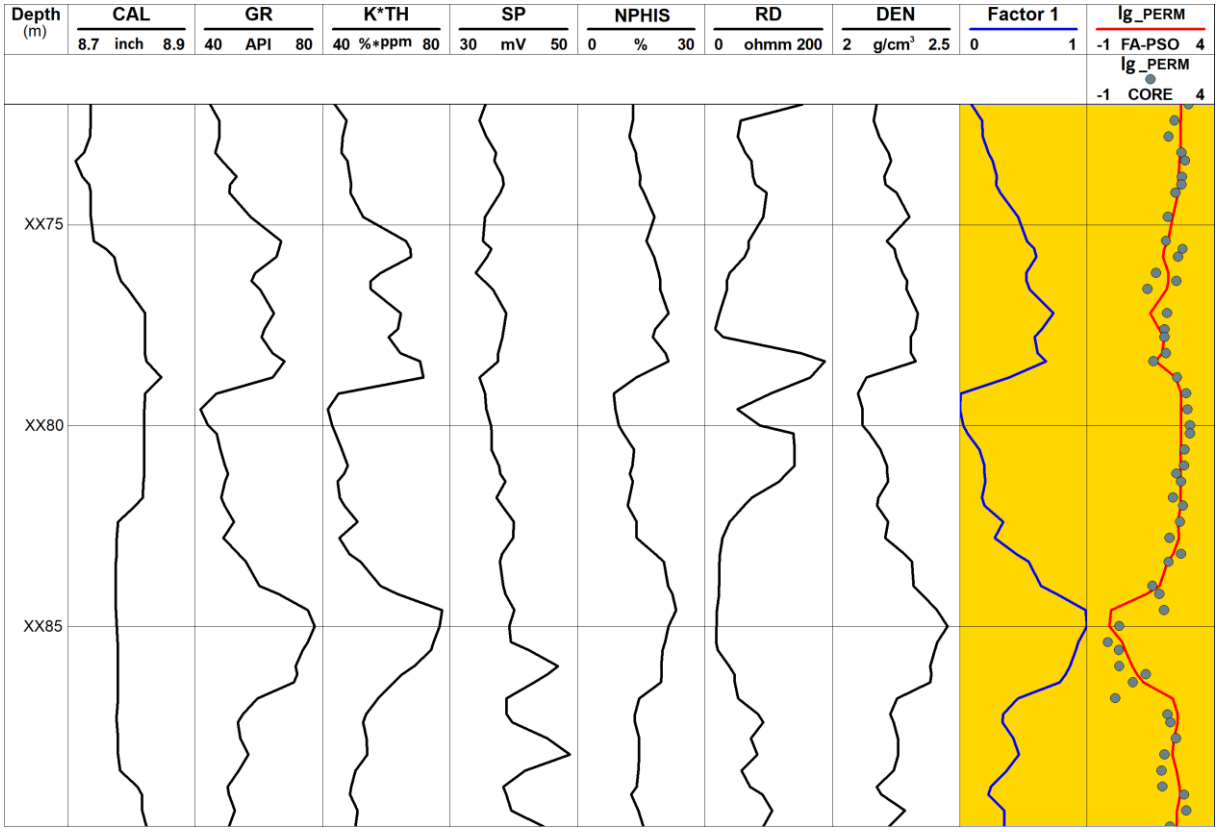

Fig. 21 The input logs of FA-PSO (track 1-7), the scaled first factor log (track 8), the derived permeability log and the permeability derived from core plugs (track 9) in Well-II

2.3 Summary

In this section, I proved that there is a strong non-linear relationship between the first factor log derived from well logging data by the newly developed FA-PSO method and the decimal logarithm of permeability of the studied hydrocarbon formations. The permeability estimations are compared with both deterministic and core derived permeability data and

DOI: 10.14750/ME.2020.007

they show a good correlation. Thus the developed method might serve as an independent source of permeability in oil field applications.

Thesis 2.

I have established a regression model between the first factor log estimated by the particle swarm optimization based factor analysis (FA-PSO) and the permeability of hydrocarbon-bearing formations. This allows the estimation of permeability from an independent well-log-analysis method, which is based on the comprehensive interpretation of all available wireline logs. The developed method has been tested on Hungarian hydrocarbon bearing formations using in situ data. The factor analysis derived permeability is compared and successfully verified by deterministic modeling and laboratory based measurements.

DOI: 10.14750/ME.2020.007

3. UNCERTAINTY REDUCTION OF INTERVAL INVERSION ESTIMATION RESULTS USING A FACTOR ANALYSIS APPROACH

Inversion methods are widely used in applied geophysics (Oldenburg and Li 2005). The increase of accuracy and reliability of inversion estimation is in focus of scientific researches, which has a great practical impact on geophysical data processing especially in well logging applications. In this chapter, I aim 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 (Dobróka and Szabó 2001), 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. I apply a metaheuristic particle swarm optimization approach as a first phase of inversion for eliminating 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, the accuracy of the model parameters is quantified.

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 inversion problem. By the improvement of the overdetermination ratio, one can reduce the estimation errors and maintain the vertical resolution of the estimated model parameters. For this purpose, I leave the number of series expansion coefficients unchanged, however, I estimate the shale volume of the investigated formation prior to inversion by the developed FA-PSO method (Fig. 22). Then I incorporate the resultant shale volume log into the interval inversion procedure as a known parameter, 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 (Abordán and Szabó 2020). This procedure significantly increases the overdetermination ratio of the interval inversion method and thus decreases the uncertainty of the remaining model parameters allowing for a more reliable calculation of hydrocarbon content.

A similar approach was developed for direct push logging data (Szabó et al. 2018), where factor analysis is used to derive the water content of the investigated shallow

DOI: 10.14750/ME.2020.007

subsurface formations to reduce the number of unknowns of the inversion procedure afterwards (Abordán and Szabó 2019a).

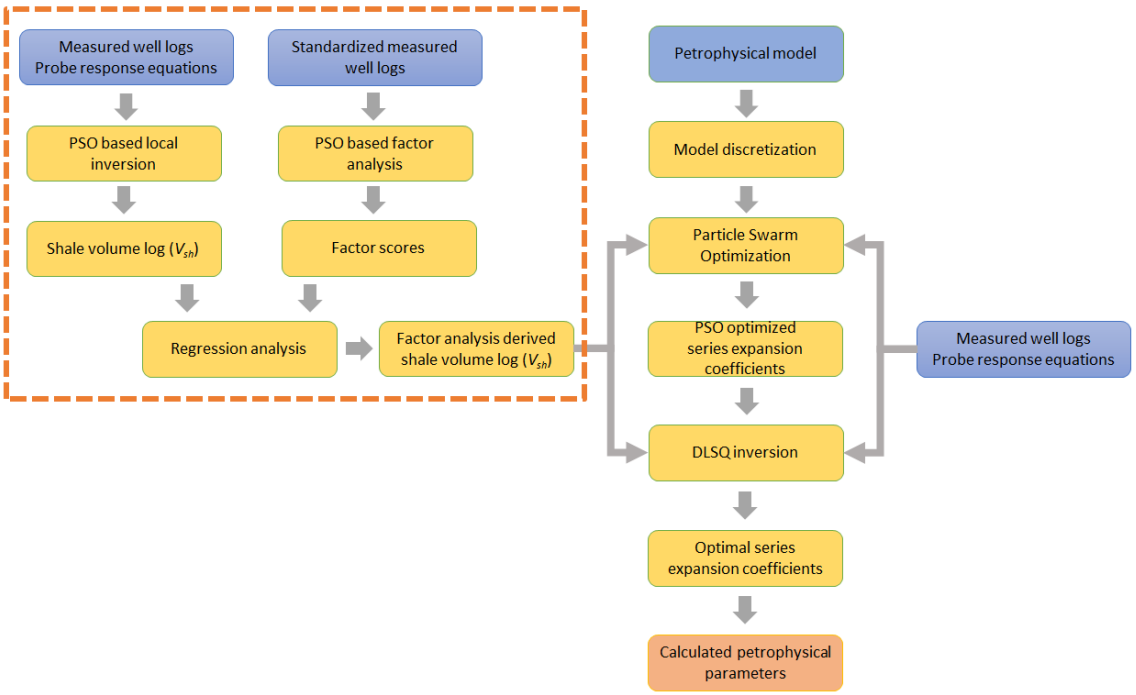

Fig. 22 The flowchart of the PSO based interval inversion procedure aided by the FA-PSO method

3.1 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 used in this section was recorded in a hydrocarbon exploratory well drilled in the Pannonian Basin in Hungary (Well-H). The processed interval is 19.2 m long and is built up of gas-bearing unconsolidated shaly sand layers of Pliocene age. In the inversion procedure the natural gamma-ray intensity (GR), true resistivity (R_t), neutron porosity (Φ_N), 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 $\Phi + V_{sh} + V_{sd} = 1$, sand volume can be calculated without increasing the number of unknowns. I 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 wireline logs contains 1158 data points. The response equations used to relate the measured data to the model parameters are as follows (Wyllie et al. 1956, Alberty and Hashmy 1984, Schlumberger 1989, Baker Atlas 1996, Szabó and Dobróka 2019)

$$GR = \rho_b^{-1}(V_{sh}GR_{sh}\rho_{sh} + V_{sd}GR_{sd}\rho_{sd}), \tag{21}$$

$$\Phi_N = \left\{ \begin{matrix} \Phi_{N,mf} - (1-S_{x0})C_{cor} - 2\Phi(1-S_{x0})S_{hrf}(1-2.2\rho_{hc}) \\ \cdot\left[1-(1-S_{x0})(1-2.2\rho_{hc})\right] \end{matrix} \right\} + V_{sh}\Phi_{N,sh} + V_{sd}\Phi_{N,sd}, \tag{22}$$

$$R_t = \left[\frac{\Phi^m S_w^n}{aR_w(1-V_{sh})} + \frac{V_{sh}S_w}{R_{sh}} \right]^{-1}, \tag{23}$$

$$\Delta t = \Phi\left[\Delta t_{mf}S_{x0} + (1-S_{x0})\Delta t_{hc}\right] c_p + V_{sh}\Delta t_{sh} + V_{sd}\Delta t_{sd}, \tag{24}$$

$$\rho_b = \Phi\left[\rho_{mf} - 1.07(1-S_{x0})(\alpha_0\rho_{mf} - 1.24\rho_{hc})\right] + V_{sh}\rho_{sh} + V_{sd}\rho_{sd}, \tag{25}$$

$$K = \rho_b^{-1}\left(\Phi S_{x0}K_{mf}\rho_{mf} + V_{sh}K_{sh}\rho_{sh} + V_{sd}K_{sd}\rho_{sd}\right), \tag{26}$$

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

$$\mathbf{d}^{(m)} = \left[GR^{(m)}, \Phi_N{}^{(m)}, R_t{}^{(m)}, \Delta t^{(m)}, \rho_b{}^{(m)}, K^{(m)}\right]^T, \tag{27}$$

$$\mathbf{d}^{(c)} = \left[GR^{(c)}, \Phi_N{}^{(c)}, R_t{}^{(c)}, \Delta t^{(c)}, \rho_b{}^{(c)}, K^{(c)}\right]^T, \tag{28}$$

where superscript T means transpose operator. Model parameters are estimated by minimizing the deviation between the vectors defined in Eqs. (27)-(28). This is conventionally solved by some linearized method (Menke 1984). Here, I utilize the metaheuristic method of PSO to develop a new variant of the local and interval inversion method, respectively. We minimize the root mean squared error between the (normalized) observed and calculated data as follows

$$E = \sqrt{\frac{1}{S}\sum_{s=1}^{S}\left(\frac{d_s^{(m)} - d_s^{(c)}}{d_s^{(m)}}\right)^2} \cdot 100[\%], \tag{29}$$

where S is the number of applied logging tools.

DOI: 10.14750/ME.2020.007

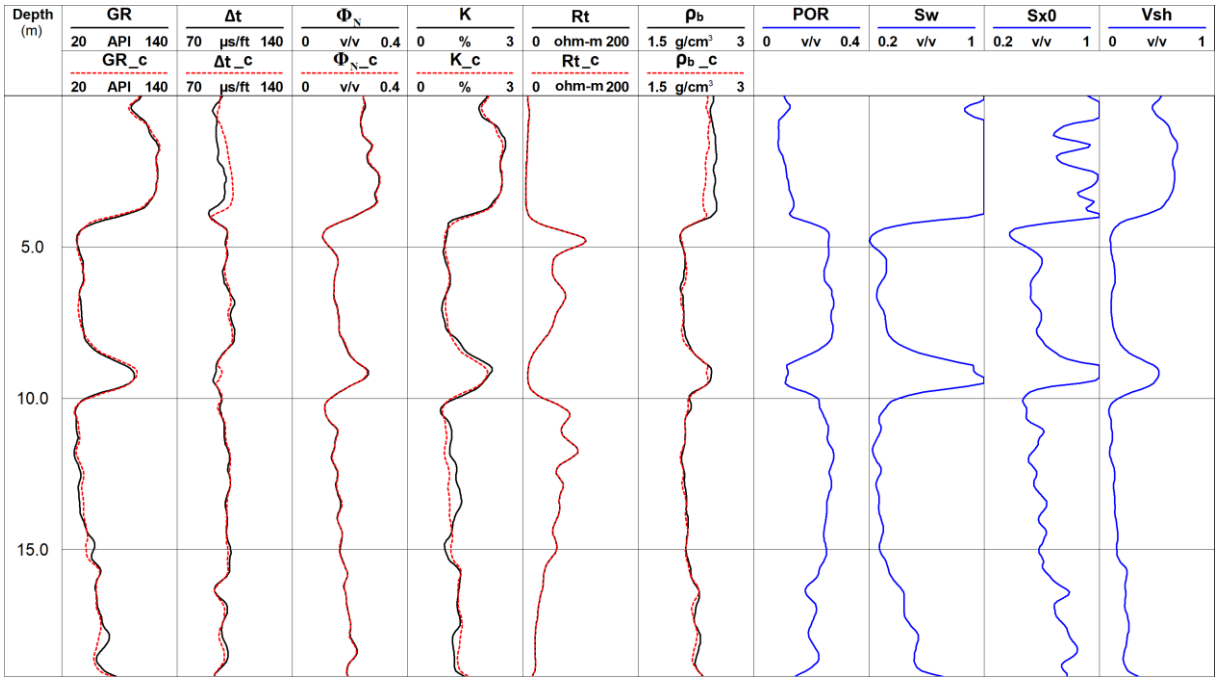

Fig. 23 Results of the PSO-based local inversion of well logging data in Well-H, the measured (black) and calculated (red dashed line) in the first 6 track, the estimated petrophysical parameters (blue) in the last 4 tracks

This metaheuristic PSO method is more and more frequently used to solve complex geophysical inverse problems (Shaw and Srivastava 2007). 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 solution. For the local inversion (Fig. 23), PSO was run one by one for the 193 measured depth points by utilizing 30 particles. The other control parameters (c_1, c_2, w) of the PSO algorithm are unchanged from the ones used in Chapter 1.2. In each depth point the four unknowns were derived from the six available logs and the response equations defined in Eqs. (21-26). The mean of the data distance defined in Eq. (29) for all the 193 depth points decreased to 3.93% by running the PSO-based inversion for 100 iteration steps in each depth point separately, overall it took twenty seconds. However, it should be noted that local inversion of wireline logging data has its drawbacks due to its high noise sensitivity, e.g., in this example S_{x0} considerably deviates from 1 in the shale layer at the top of the investigated interval (0-4 m).

3.2 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

$$\mathbf{d}_s^{(c)}(z) = g_s(m_1(z),..., m_i(z),..., m_M(z)), \tag{30}$$

where g_s denotes the response function of the sth logging tool (s=1,2,…,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. (30) are discretized by series expansion after Dobróka et al. (2016)

$$m_i(z) = \sum_{q=1}^{Q^{(i)}} B_q^{(i)} P_{q-1}(z), \tag{31}$$

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

$$P_{q-1}(z) = P_l(z) = (2^l l!)^{-1} \frac{d^l}{dz^l}(z^2-1)^l, \tag{32}$$

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 appropriately overdetermined inversion 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. (33). Calculated data is obtained by using Eqs. (30)-(31) and the data distance modifies to

$$D = \frac{1}{N}\sum_{n=1}^{N}\left(\sqrt{\frac{1}{S}\sum_{s=1}^{S}\left(\frac{d_s^{(\mathrm{m})}-d_s^{(\mathrm{c})}}{d_s^{(\mathrm{m})}}\right)^2}\right)\cdot 100\,[\%], \tag{33}$$

where N denotes the number of measured depth points along the processed interval and S is the number of applied logging tools. Equation (33) 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 inversion 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. (31) 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

$$\mathrm{cov}\mathbf{B} = \mathbf{G}^{-g}\,\mathrm{cov}\mathbf{d}^{(\mathrm{m})}(\mathbf{G}^{-g})^{\mathrm{T}}, \tag{34}$$

where $\mathbf{d}^{(\mathrm{m})}$ is the vector of measured data and $\mathbf{G}^{-g}$ is the generalized inverse of the problem. To calculate the uncertainty of estimated petrophysical parameters, the depth-dependent model covariance matrix of the model parameters needs to be introduced (Dobróka et al. 2016)

$$\left[\mathrm{cov}\mathbf{m}(z)\right]_{ij} = \sum_{n=1}^{Q^{(i)}}\sum_{m=1}^{Q^{(j)}} P_{n-1}(z)\left(\mathrm{cov}\mathbf{B}\right)_{hh'}P_{m-1}(z), \tag{35}$$

where **B** is the vector of series expansion coefficients, indices are (i=1,2,...,M; j=1,2,...,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

$$\sigma\left[m_i(z)\right] = \left[cov\,\mathrm{m}_{ii}(z)\right]^{1/2}. \tag{36}$$

3.3 Field test of interval inversion using Hungarian oilfield data

Interval inversion is performed on the same dataset and petrophysical model used for local inversion in Chapter 3.1. The workflow of the inversion procedure presented in this subchapter can be seen in Fig. 22, neglecting the steps bordered by the dashed line. By discretizing the 4 model parameters according to Eq. (31) 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 (1158 data points / 180) 6.4. The values of expansion coefficients are initialized by minimizing Eq. (33) with the algorithm of PSO. At the start of the inversion procedure, 45 particles are initialized by PSO, each representing a possible solution for the 180 expansion coefficients describing the petrophysical parameters. The only constraint is that the zeroth Legendre polynomials are initialized from their possible physical 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 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 2.72×10^{7}%, a standard deviation of data distance of 1.61×10^{8}% and the maximum data distance of all particles is 1.08×10^{9}%. Then the set of particles move in the search space according to Eqs. (12)-(13) 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 inversion procedure is followed with a linearized phase using the DLSQ method for 15 more steps, the final data distance is 4.87% (Fig. 24). The whole iteration process takes 30 seconds on a quad core workstation.

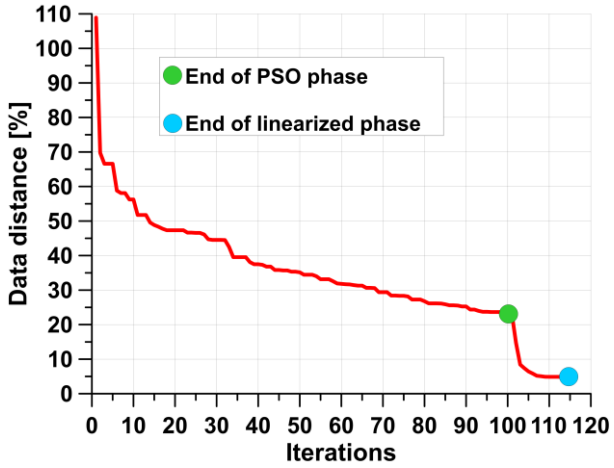

Fig. 24 Convergence of data distance to the optimum having 4 unknown petrophysical properties

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. (36), we can calculate the estimation error of the resultant petrophysical parameters by using the uncertainties of the observed parameters (Fig. 25).

DOI: 10.14750/ME.2020.007

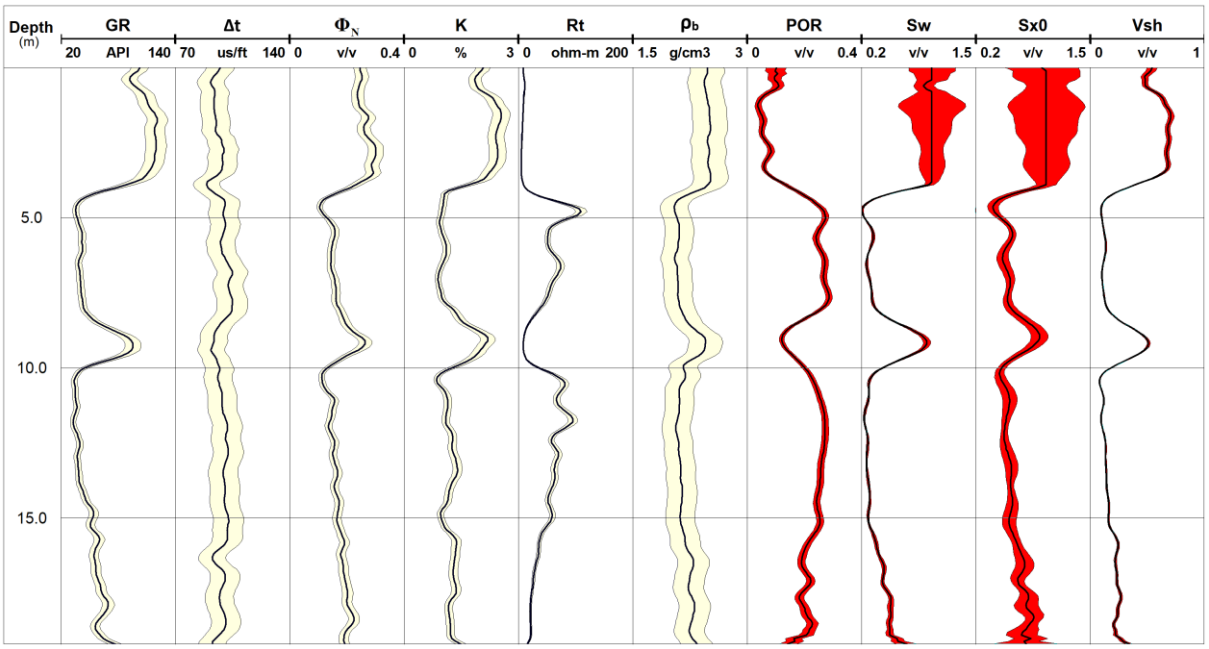

Fig. 25 The measured well logs in the first 6 tracks with their assumed uncertainties (shaded area) and the resultant petrophysical parameters and their estimation errors (shaded area) in the last 4 track in Well-H

The average estimation error of the resultant petrophysical parameters are $\Phi = 0.026\,v/v, S_w = 0.112\,v/v, S_{x0} = 0.245\,v/v, V_{sh} = 0.024\,v/v$. The estimation accuracy is considerably smaller for water saturation than for porosity or shale volume, respectively, because it shows the strongest correlation with the other model parameters.

3.3.1 FA-PSO derived shale volume

To assess how the increase of the overdetermination ratio affects the suggested PSO-based interval inversion method, first, I derive one of the petrophysical parameters from a different source and then incorporate it into the inversion procedure. For this purpose, I choose the newly developed method of factor analysis (FA-PSO) to estimate shale volume of the investigated formation. 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. (6). Factor loadings representing the correlation relationship between the extracted factor and measured logs are calculated according to Eq. (4). I extract only one factor from the dataset, the calculated factor loadings after rotation by the *varimax* algorithm 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 the developed FA-PSO method finds the optimal values

of the factor scores by minimizing Eq. (7) utilizing 90 particles as solution candidates in 1000 iteration steps in 20 seconds on a quad core workstation (Fig. 26a). Based on the calculated loadings, all logs included in factor analysis correlate well with the first extracted factor. First, the factor log needs to be scale it into the range of 0 to 1 according to Eq. (10) 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 (Fig. 26b). In this study, I choose the local inversion derived shale volume for regression analysis that is detailed in Chapter 3.1. 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 general form (Szabó 2011)

$$V_{sh} = ae^{bF_1} + c, \tag{37}$$

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. 26b).

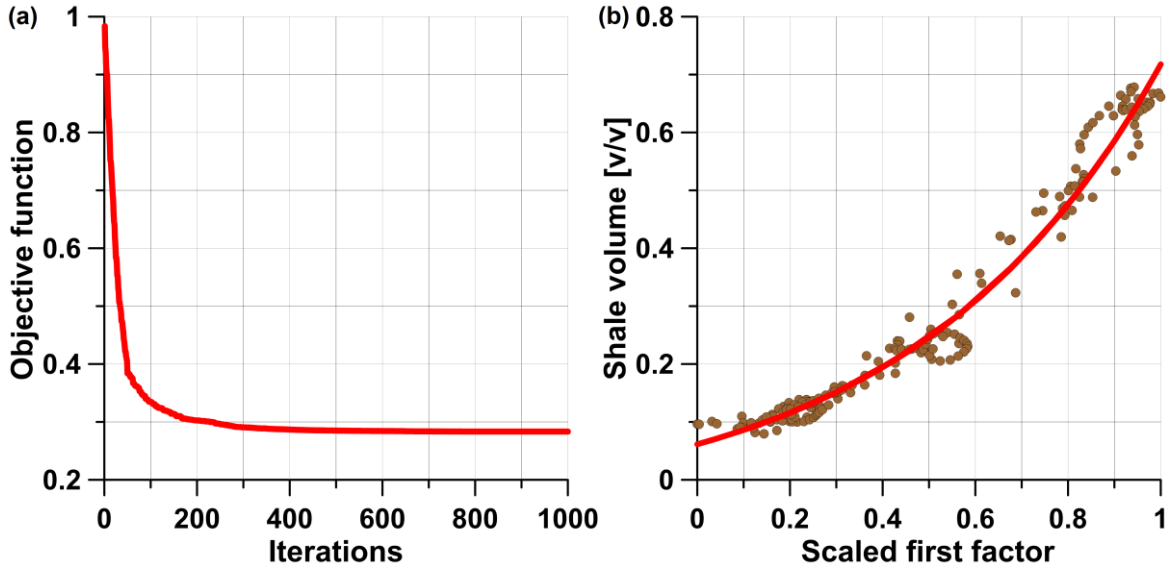

Fig. 26 The convergence of the PSO-based factor analysis procedure on the left, the relation of the first extracted factor and shale volume on the right: exponential regression model (red line) and shale volume estimated by local inversion (dots) in Well-H

The rank correlation coefficient of 0.99 practically indicates entire correlation and non-linear relationship between the scaled first factor and shale volume.

3.3.2 Interval inversion improved by factor analysis

Interval inversion is run again with the same parameters and constraints as detailed in Chapter 3.2. However, shale volume is now considered as a known parameter along the

inverted interval derived by the suggested FA-PSO method as it is shown in Fig. 22. 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 3.2 where 4 model parameters were estimated by interval inversion. The convergence of data distance to the optimum both in the global and linearized phases of the inversion procedure is quite steady (Fig. 27)

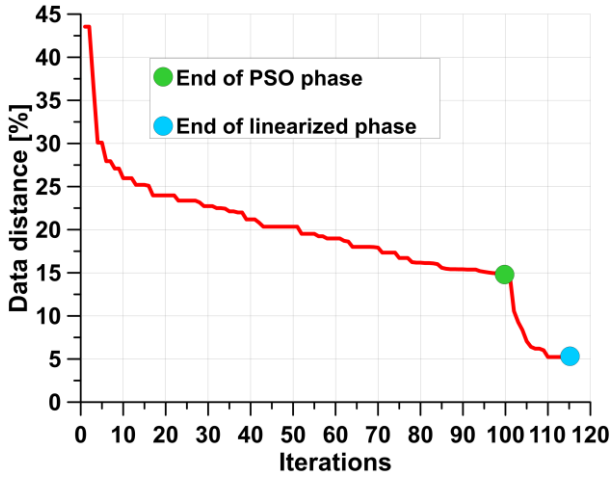

Fig. 27 Convergence of data distance for the hybrid interval inversion procedure

After 100 iteration steps, PSO reaches a data distance of 14.69% by minimizing Eq. (33). Then the DLSQ method further decreases data distance to 5.22% in 15 iteration steps. Including the runtime of factor analysis, this takes 40 seconds on a quad core workstation. Due to the increased data-to-unknowns ratio, the estimation error of the resultant petrophysical parameters is decreased (Fig. 28), the shale volume derived by the FA-PSO method based on Eq. (37) can be seen in the last track. The average estimation errors of the resultant petrophysical parameters calculated by Eq. (36) are decreased to $\Phi = 0.024\,v/v, S_w = 0.074\,v/v, S_{x0} = 0.170\,v/v$ (Fig. 29a), which is quite an improvement as indicated in percentage (Fig. 29b).

DOI: 10.14750/ME.2020.007

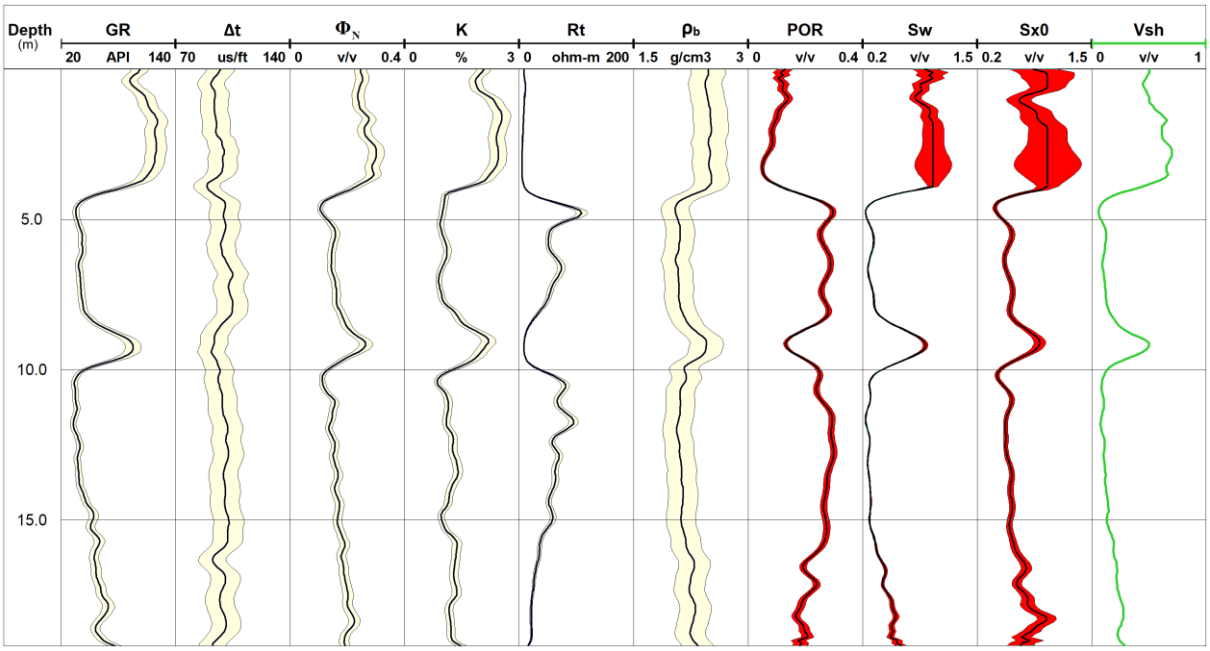

Fig. 28 The measured well logs in the first 6 tracks with their assumed uncertainties (shaded area) and the resultant petrophysical parameters and their estimation errors (shaded area) in tracks 7-9 and the shale volume derived from factor analysis in the last track in Well-H

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 or estimating one of them from another reliable sources.

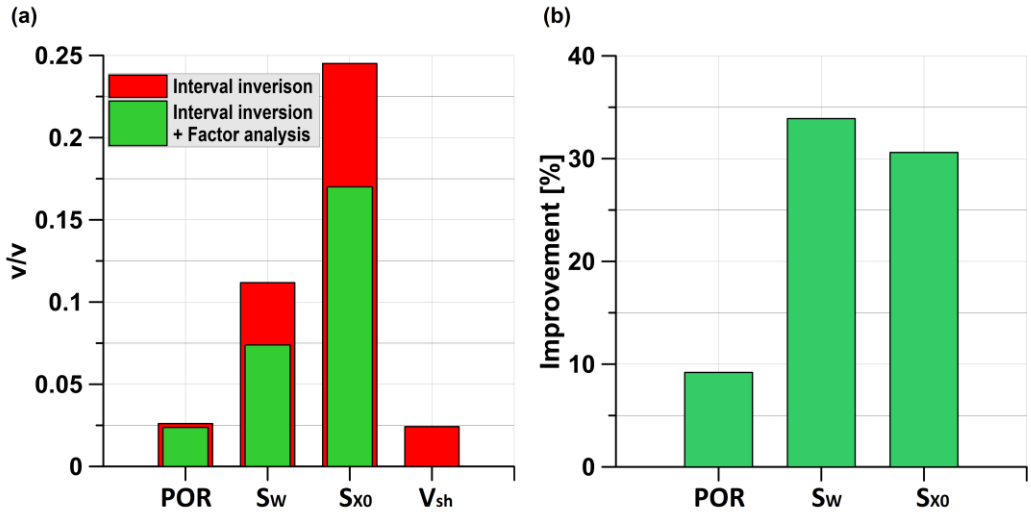

Fig. 29 The decrease in average estimation errors due to the increased data-to-unknowns ratio of the interval inversion method (a). The improvement of estimation accuracy in percent (b)

DOI: 10.14750/ME.2020.007

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 allows 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. (21)-(26).

3.4 Stability test of interval inversion procedure

To check the stability of the proposed PSO-based interval 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 (Chapter 3.3.2). First, 45 particles are initialized in the search space, each representing a possible solution of the 135 series expansion coefficients. The initial candidates generated by PSO are very diverse (Fig. 30a, b, c) and are of extremely high distance from the optimum.

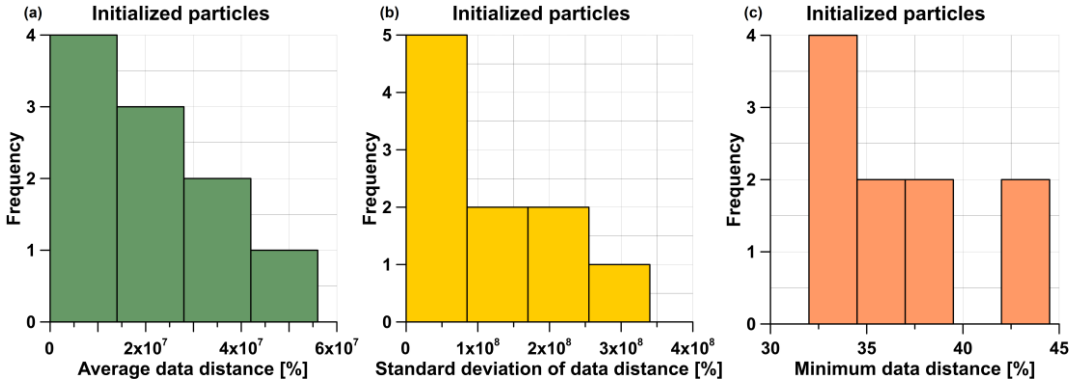

Fig. 30 Statistical distribution of the randomly generated starting models by PSO. The average data distance (a), the standard deviation (b) and the minimum data distance (c) of the solution candidates at initialization

Then each solution candidate is refined by PSO according to Eqs. (12)-(13) 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. 31), 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).

DOI: 10.14750/ME.2020.007

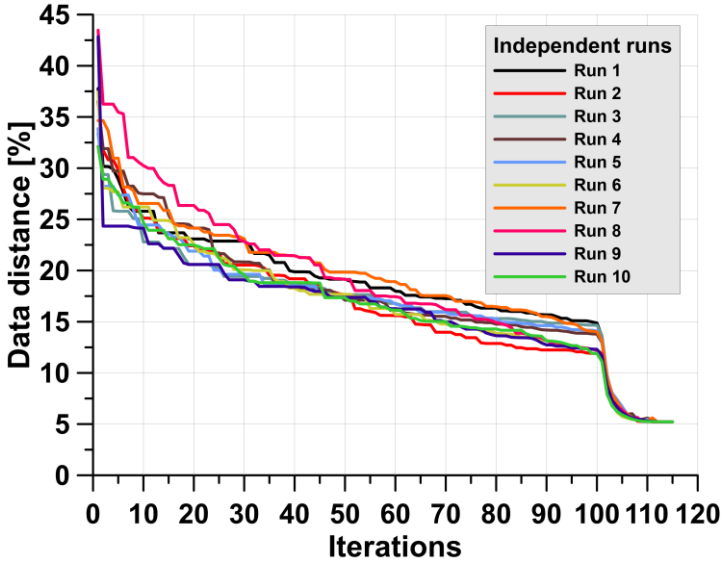

Fig. 31 Convergence plots of the PSO-initialized interval inversion procedure for 10 independent program runs

Once the data distance between the measured and calculated data is adequately small (~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).

3.5 Summary

The suggested new PSO-based metaheuristic inversion approach for estimating the series expansion coefficients 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 shale volume derived by factor analysis from a given dataset can be successfully incorporated into the interval inversion of the same dataset to increase its data-to-unknowns ratio, and thus improving the estimation

accuracy of the estimated petrophysical parameters. 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 the volume of movable hydrocarbons, therefore their most reliable estimation is of crucial importance. The increased overdetermination ratio would also give the possibility to automatically estimate the value of some zone parameters within inversion, while still maintaining a good level of spatial resolution of the estimated model parameters. The suggested inversion procedure to increase the overdetermination of the well logging inverse problem can be very effectively used for unconventional (shale gas) reservoirs where multi-mineral models need to be built and therefore the number of unknowns is considerably higher than in case of conventional reservoirs (Szabó and Dobróka 2020).

Thesis 3.

I have developed a particle swarm optimization based interval inversion method for estimating the petrophysical parameters of hydrocarbon formations from wireline logging data. The implementation of the metaheuristic approach highly reduces the starting model dependence of the inverse problem. I have increased the overdetermination ratio of the interval inversion method by incorporating the factor analysis derived shale volume log into the inversion procedure. Thus, by taking shale volume along the inverted interval as known (fixed) parameter, the number of unknown model parameters is decreased and the accuracy and reliability of the estimated petrophysical model is significantly improved. I performed detailed stability tests, and proved that the results of randomly initialized PSO-based interval inversion procedures are consistent.

DOI: 10.14750/ME.2020.007

4. THE AUTOMATED SELECTION OF CONTROL PARAMETERS FOR THE FA-PSO METHOD (FA-PSO-CC)

PSO as a metaheuristic approach, its output is highly effected by the control parameters that worth to be set as optimal as possible at the start of the searching mechanism. These parameters are usually chosen according to empirical suggestions from literature, which can make a problem ambiguous. In this section, I develop an algorithm to generalize the presented PSO-based factor analysis (FA-PSO) in regard of the c_1 and c_2 control parameters (see Eq. (13)). For this purpose, these control parameters are set by simulated annealing in an outer loop of the algorithm in an automated way and then in the inner loop of the procedure, FA-PSO is run with the newly estimated c_1 and c_2 control parameters (Abordán and Szabó 2019b). The newly developed method is named FA-PSO-CC.

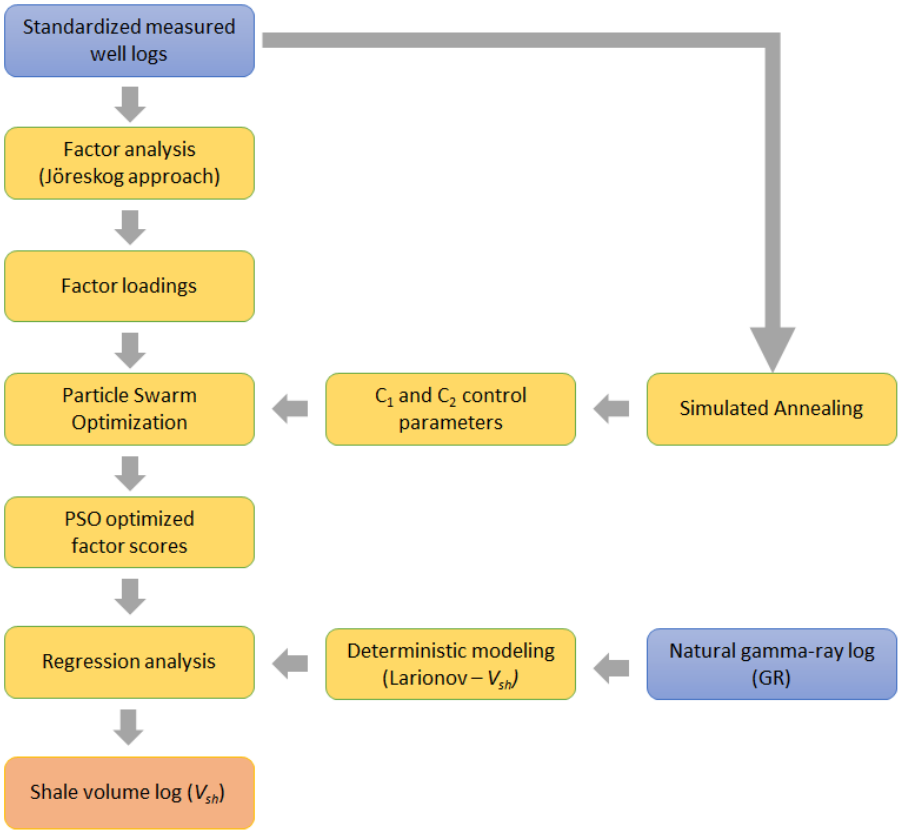

Fig. 32 Workflow of the suggested FA-PSO-CC well-log-analysis procedure

The suggested meta-algorithmic optimization approach is tested on the wireline logging dataset measured in the thermal water exploratory well in Baktalórántháza (Well-A), north-east Hungary (Fig. 11). The well is 1,197 m deep and the well logs from its upper Pleistocene section (100-193.1 m) were utilized for testing the FA-PSO method in Chapter

1.2.1. Here, a Pliocene interval is processed from 448.5 m to 486.4 m. Once the optimal values of the control parameters c_1 and c_2 of the FA-PSO method are found by SA (Metropolis et al. 1953), factor logs are extracted from the well logging dataset, the first of which is related to the shale volume of the investigated formation. Results of the method are also confirmed by core data.

4.1 Field test – Baktalórántháza

The utilized wireline logging dataset consists of natural gamma-ray intensity (GR), spontaneous potential (SP), shallow resistivity (RS), gamma-gamma intensity (GG) and neutron-neutron intensity (NN) logs. The investigated interval is from 448.5 m to 486.4 m, which consists mainly of shaly-sand, shaly marl and gravel layers. The workflow of the FA-PSO-CC procedure can be seen in Fig. 32. As a first step, factor loadings are calculated for three factors by Eq. (4). The resultant rotated factors are in Table 6.

Table 6 Rotated factor loadings estimated by the FA-PSO method in Well-A

Well logs	***Factor 1***	***Factor 2***	***Factor 3***
GR	0.7460	0.1005	0.1653
SP	0.5995	0.5136	0.3021
NN	−0.1311	−0.4922	−0.5660
GG	0.1371	0.7215	0.0329
RS	−0.3724	0.0077	−0.6280

As expected, the first extracted factor as a lithological indicator strongly correlates with the natural gamma-ray and spontaneous potential logs. The second factor is mainly related to the gamma-gamma, spontaneous potential and neutron-neutron logs, while the third factor is correlated negatively to the shallow resistivity and neutron-neutron logs. Then the objective function defined in Eq. (7) is minimized by PSO. In this example, 90 particles are generated in the search space, each representing a solution for the factor scores **f**, by extracting three factors, the number of unknowns to be estimated is (3 factors × 380 measured depth points) 1,140. For the PSO algorithm the inertia weight w is set according to Eq. (14). Then the values of c_1 and c_2 control parameters are optimized by SA. The usually recommended (default) setting is 2 for both parameters (Kennedy and Eberhart 1995). To test the presented method, at the beginning of the SA procedure, both c_1 and c_2 are set as 1 and in every iteration step their value is slightly altered by adding a small b perturbation parameter to both values. Then with these new control parameters, the PSO-based factor analysis is run for 3,000 iteration steps. If the energy difference

(ΔE) in two subsequent iteration steps of the SA algorithm in the outer loop of the algorithm according to Eq. (7) is negative, 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, then the probability of accepting the new control parameters is given by P_a=exp(-ΔE/T), where T is the current temperature of the system. During the iteration process, the temperature of the system is reduced logarithmically according to Eq. (9). The new control parameters are accepted only if a randomly generated number from the range of 0 and 1 is smaller than P_a. This is a fundamental feature of the SA algorithm; it prevents the search from being stuck in a local minimum near the starting model. SA is run for a 100 iteration steps to find the optimal values of c_1 and c_2. The newly generated control parameters are tested in each SA iteration step by running the FA-PSO method for 3,000 iterations for three consecutive times to eliminate its probabilistic nature, especially in the early stages, where data distances remain fairly far from the optimum because of the suboptimal c_1 and c_2 control parameters. The mean of the three consecutive FA-PSO runs for each SA iteration can be seen in Fig. 33. After 48 iteration loops SA finds the optimal values of the control parameters for the FA-PSO algorithm and data distance does not decrease any further.

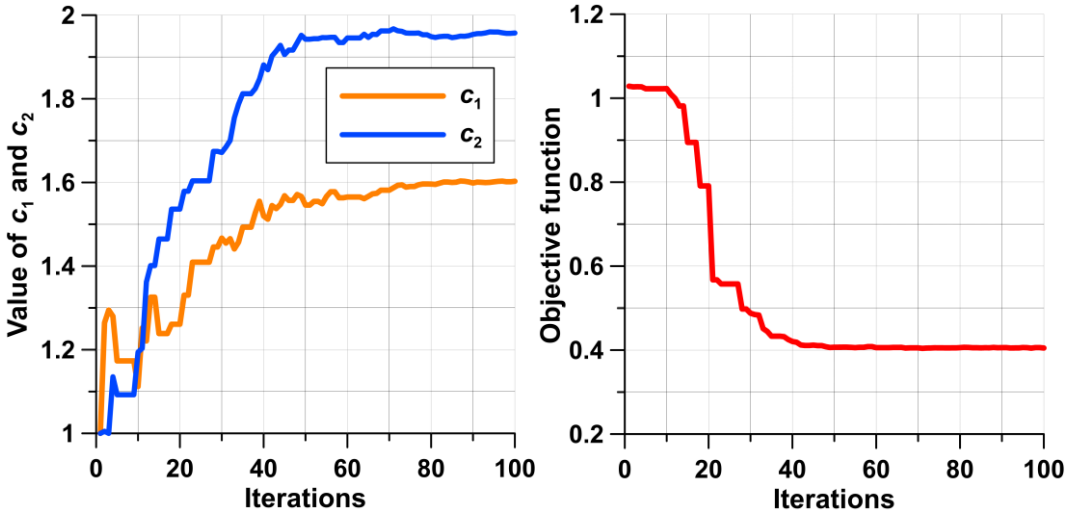

Fig. 33 Convergence of data distance (on the right) by altering control parameters c_1 and c_2 of the FA-PSO method (on the left)

In the last iteration step, the smallest data distance was found by using c_1=1.60 and c_2=1.96. Then the PSO-based factor analysis was run for 3,000 iteration steps with these parameters to find the optimal solution of the factor scores **f.** The convergence of data distance can be seen in Fig. 34 on the left. After 3,000 iterations the data distance decreased to 0.404. By regression analysis between the first extracted factor (F_1) scaled in the range of

0 to 100 according to Eq. (10) and the shale volume estimated by deterministic modeling resulted in a linear relationship (Fig. 34 on the right) in the form suggested by Szabó et al. (2014)

$$V_{sh} = aF_1 + b\,, \tag{38}$$

where the regression coefficients with 95% confidence bounds are calculated to be a=0.5005 [a_{min}=0.4771, a_{max}=0.5239] and b=30.41 [b_{min}= 29.39, b_{max}= 31.43].

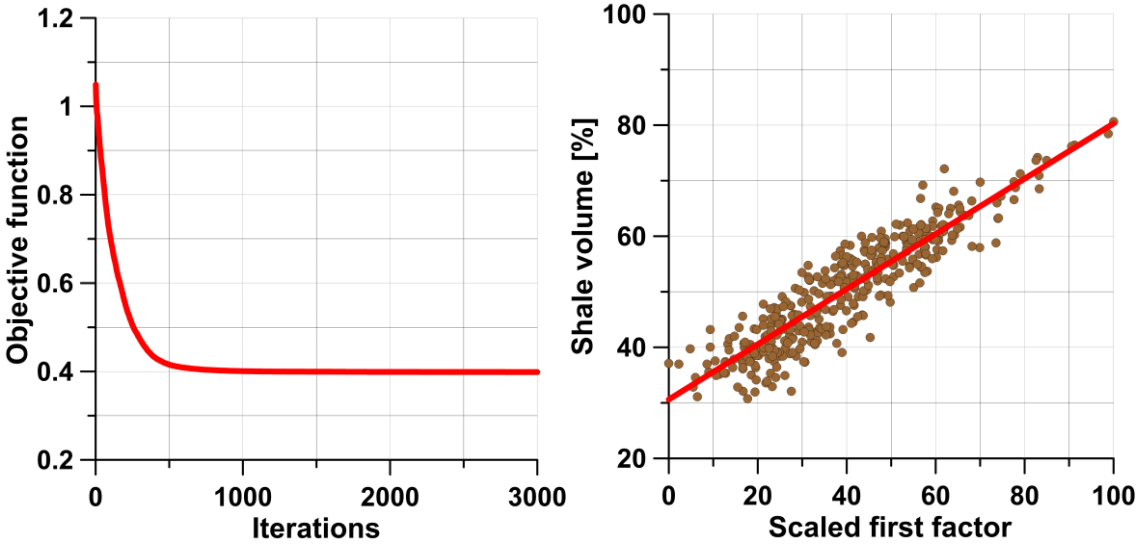

Fig. 34 Convergence of the energy function (left) for FA-PSO-CC and the relationship between the first factor and shale volume (right)

The Pearson's correlation coefficient of 0.91 for the first scaled factor and the shale volume of the investigated interval indicates a strong linear relationship. Figure 35 presents the results of the suggested FA-PSO-CC method. In the first five tracks, the standardized input well logs can be seen in black and the calculated logs from the factor model by red dashed line. The fit between the measurements and predictions is relatively good. In the sixth track, the scaled first factor log is shown in blue and the last track contains the resultant shale volumes. Shale volume estimated by deterministic modeling (Larionov 1969) is drawn by a purple dashed line and the one estimated by factor analysis is shown in red. They match really well, which verifies the applicability of the method. The latter is also confirmed by laboratory measurements for shale volumes, which are indicated by dots in the same track.

DOI: 10.14750/ME.2020.007

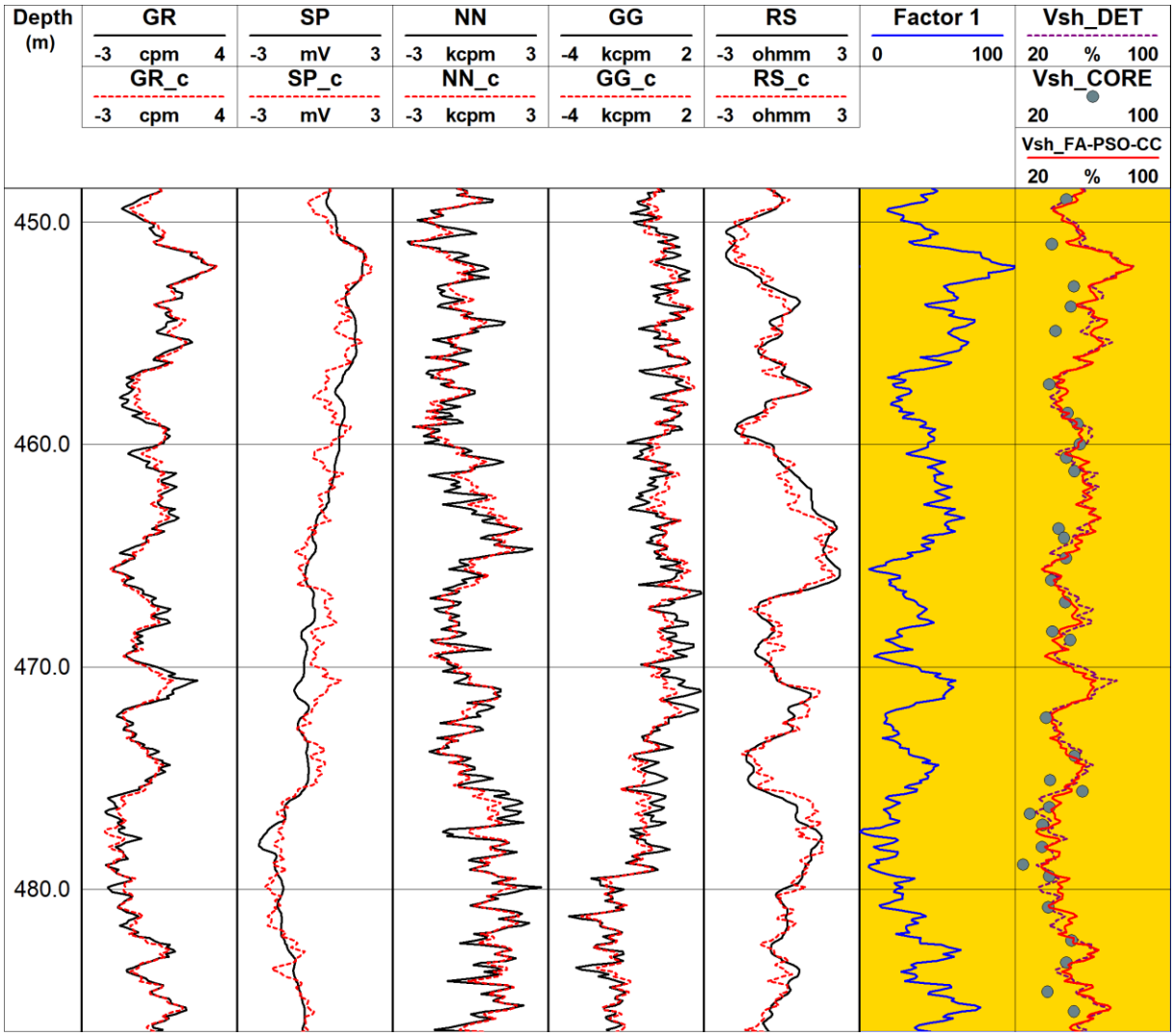

Fig. 35 The input (black) and calculated well logs (red dashed line) in tracks 1-5, extracted first factor log by the FA-PSO-CC method in tack 6, and the resultant shale volumes in track 7 in Well-A

4.2 Summary

For the generalization of the developed FA-PSO method, selection of c_1 and c_2 control parameters of the PSO algorithm was done by an automated simulated annealing-based iterative procedure known commonly as a hyperparameter estimation approach in the terminology of machine learning. By regression analysis, a linear relationship was found between the first extracted factor and shale volume, which forms a basis of the estimation of the shaliness of the investigated formation from well logs. The good fit between the measured and calculated well logs and the match between the estimated shale volumes by deterministic modeling, the suggested improved FA-PSO method and core analysis indicate the applicability of the method.

DOI: 10.14750/ME.2020.007

Thesis 4.

I have developed a hyperparameter estimation based data processing approach for the automated selection of control parameters c_1 and c_2 for the particle swarm optimization assisted factor analysis (FA-PSO). The factor scores are estimated in the inner loop of FA-PSO-CC, while the optimal values of control parameters c_1 and c_2 are automatically determined in an outer iteration loop by simulated annealing. By processing in-situ well logging data, I proved the feasibility of the suggested method in Hungarian groundwater formations, and estimated the shale volume directly from the global optimization-derived factor scores. The result was confirmed also by core measurements.

DOI: 10.14750/ME.2020.007

5. SIMULTANEOUS OPTIMIZATION OF FACTOR LOADINGS AND FACTOR SCORES FOR THE FA-PSO METHOD (FA-PSO-FL)

In case of the suggested methods for the solution of factor analysis in Chapter 1 (FA-SA, FA-PSO), only the vector of factor scores **f** is estimated by either simulated annealing or the particle swarm optimization technique in the model of factor analysis $\mathbf{d} = \tilde{\mathbf{L}}\mathbf{f} + \mathbf{e}$. Prior to the approximation of factor scores, first the values of factor loadings $\tilde{\mathbf{L}}$ are calculated by Eq. (4) with the non-iterative approximate method suggested by Jöreskog (2007) and then are fixed for the remainder of the procedure. Thus the unknowns of the inverse problem are only the factor scores, which are estimated iteratively to decrease the misfit between the measured well logs collected in the column vector **d** and the theoretical well logs that are given by the multiplication $\tilde{\mathbf{L}}\mathbf{f}$ in the model of factor analysis Eq. (6). An iteratively re-weighted optimization solution for both the factor scores and factor loadings was first suggested by Szabó and Dobróka (2017) for oilfield applications. Then Szabó et al. (2017) applied a similar linearized optimization procedure for the interpretation of engineering geophysical sounding logs.

Here, I suggest the simultaneous optimization of both the factor scores and factor loadings in an iterative procedure by particle swarm optimization to give a fully automated solution and to even further decrease the misfit between the observed and calculated data. The developed method has been chosen to call as FA-PSO-FL. Once the input wireline logs are standardized and collected in vector **d**, the solution is initialized by solving Eq. (4) to estimate the values of factor loadings by the method of Jöreskog. As well as before, the loadings are rotated with the *varimax* algorithm for getting more meaningful factors. Then to test the capabilities of the suggested metaheuristic procedure (Fig.36), the estimated matrix of factor loadings is contaminated with 25% Gaussian distributed noise.

DOI: 10.14750/ME.2020.007

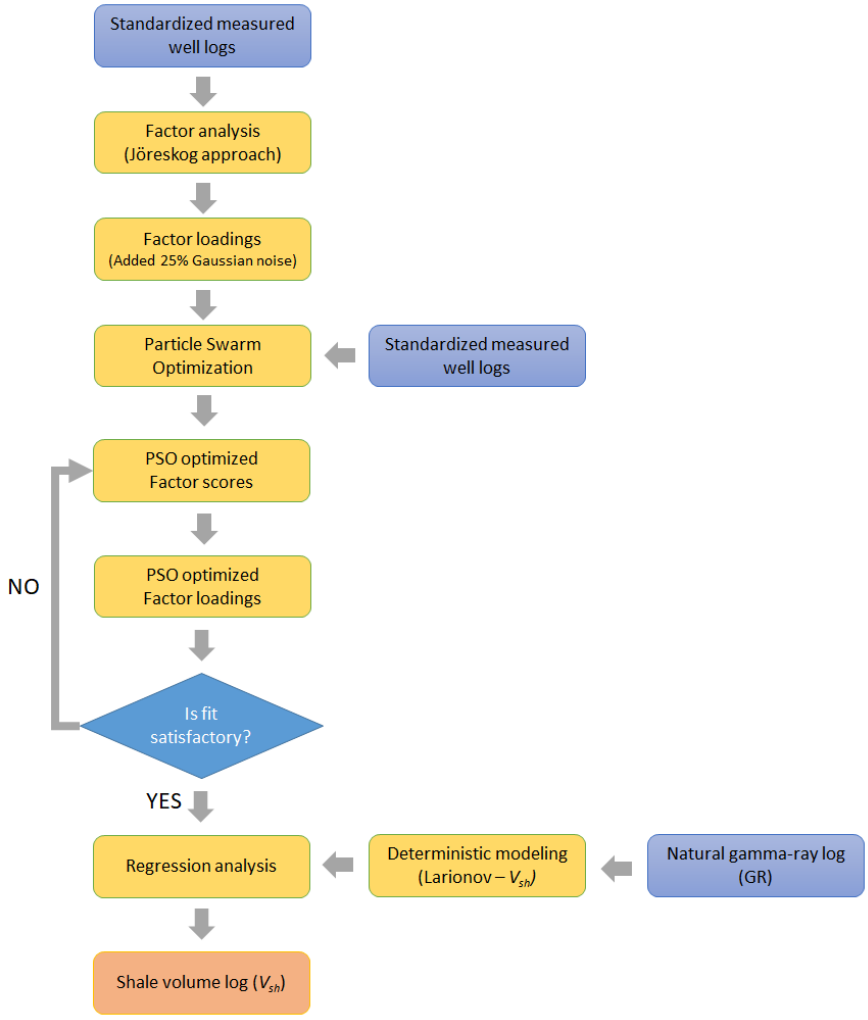

Fig. 36 Workflow of the suggested well-log-analysis procedure

As the next step, these loadings are fixed and by minimizing the difference between the measured data **d** and the calculated data $\tilde{\mathbf{L}}\mathbf{f}$, a first approximation is given for the factor scores by PSO. In all steps, the L_2-norm based objective function defined in Eq. (7) is being minimized. Once a first approximation is given for the column vector **f**, PSO is initialized to optimize the factor loadings, while the factor scores are kept fixed. Thus in each step, either the factor scores or the factor loadings are optimized while the other is kept constant until the fit between the measured and calculated data is minimized.

5.1 Feasibility test at Baktalórántháza site

This improved version of the developed FA-PSO method is tested on the wireline logging dataset measured in the thermal water well in Baktalórántháza, north-east Hungary (Fig. 11). Here, the same Pliocene interval is processed from 448.5 m to 486.4 m as in

Chapter 4.1, which consists mainly of shaly-sand, shaly marl and gravel layers. The initial vales of factor loadings given by solving Eq. (4) and then rotated by the *varimax* algorithm can be seen in Table 7.

Table 7 Rotated factor loadings estimated by Jöreskog's approach in Well-A

Well logs	***Factor 1***	***Factor 2***	***Factor 3***
GR	0.7460	0.1005	0.1653
SP	0.5995	0.5136	0.3021
NN	−0.1311	−0.4922	−0.5660
GG	0.1371	0.7215	0.0329
RS	−0.3724	0.0077	−0.6280

By extracting three factors, the number of unknowns to be estimated is (3 factors×380 measured depth points) 1,140. Therefore, the algorithm of PSO in the next phase has to find the global optimum by finding the optimal values of 1,140 variables along the borehole, i.e., factor scores. This is done in 3,000 iteration steps with 90 particles and all other control parameters of the algorithm identical to the ones detailed in Chapter 4.1. Then the PSO estimated vector of factor scores **f** is fixed, and the factor loadings are optimized by utilizing 60 particles in 300 iteration steps. In this phase of the inversion, the number of inversion unknowns is only 15 (3 factor logs × 5 factor loadings). That is the reason for the decreased number of necessary iteration steps and particles. In the phase of factor loadings optimization, the search space of PSO is set to -1 to 1. These PSO steps for the optimization for the factor scores and factor loadings are repeated five times as Fig. 37 indicates, which took 5 minutes.

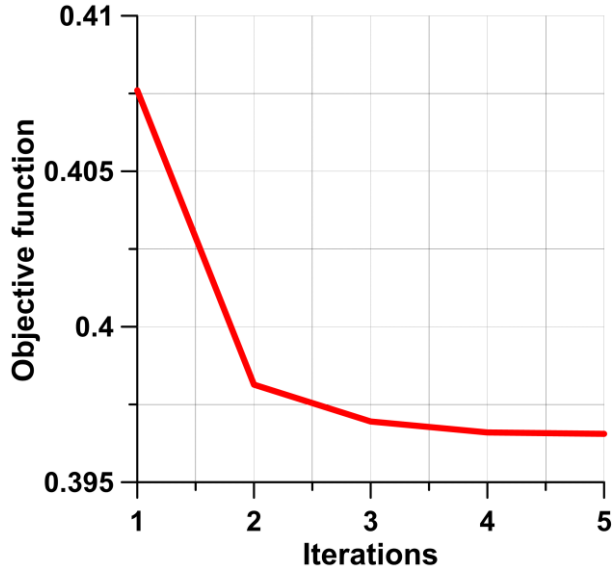

Fig. 37 The convergence of the fully automated PSO based factor analysis (FA-PSO-FL) method in Well-A

The first plotted iteration is the end of the first approximation of the vector of factor scores by PSO. This is slightly higher than it is in Chapter 4.1 for the same dataset with identical control parameters for the PSO algorithm. This is due to the fact that here the matrix of factor loadings is contaminated with 25% Gaussian noise. However, after five iteration loops, this further improved PSO-based factor analysis reaches an even lower misfit. The PSO estimated factor loadings are in Table 8.

Table 8 Factor loadings estimated by the FA-PSO-FL method in Well-A

Well logs	***Factor 1***	***Factor 2***	***Factor 3***
GR	0.7823	0.1119	0.1145
SP	0.5459	0.4262	0.2805
NN	−0.1274	−0.4302	−0.7134
GG	0.0503	0.6344	−0.1546
RS	−0.4399	−0.0073	−0.9903

Based on Table 8, one can see that the PSO optimized factor loadings are turned out to be very close to the original Jöreskog solution (Table 7) by Eq. (4). The new algorithm (FA-PSO-FL) managed to decrease the data distance even further found in Chapter 4.1 for the same dataset even though 25% Gaussian noise was added initially to the matrix of factor loadings.

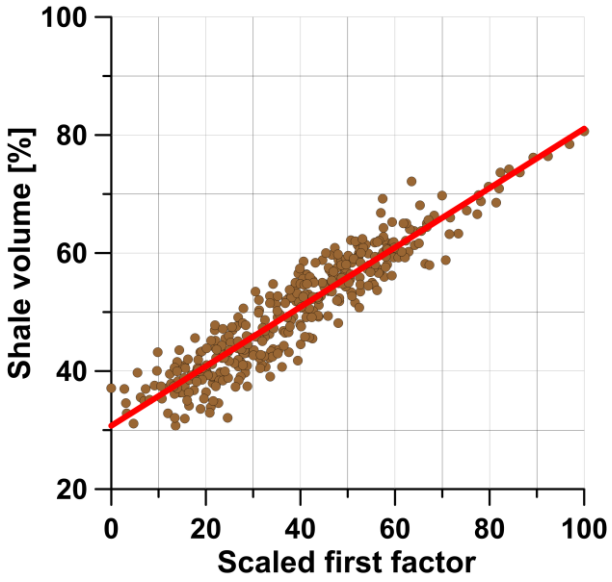

Fig. 38 Regression relationship between the first factor log estimated by the FA-PSO-FL method and shale volume derived from GR log based deterministic modeling in Well-A

For Eq. (38) the regression coefficients with 95% confidence bounds are calculated to be a=0.5035 [a_{min}=0.4833, a_{max}=0.5236] and b=30.73 [b_{min}= 29.87, b_{max}= 31.6]. The correlation coefficient of 0.93 shows a strong linear relationship between the first extracted

DOI: 10.14750/ME.2020.007

factor and shale volume. Compared to the results of the FA-PSO method in Chapter 4.1 (Fig. 34) it increased by 0.02. The input and resultant logs of the developed method are plotted in Fig 39. In the first five tracks, the measured (standardized) well logs are plotted in black and the calculated logs from the factor model by red dashed line.

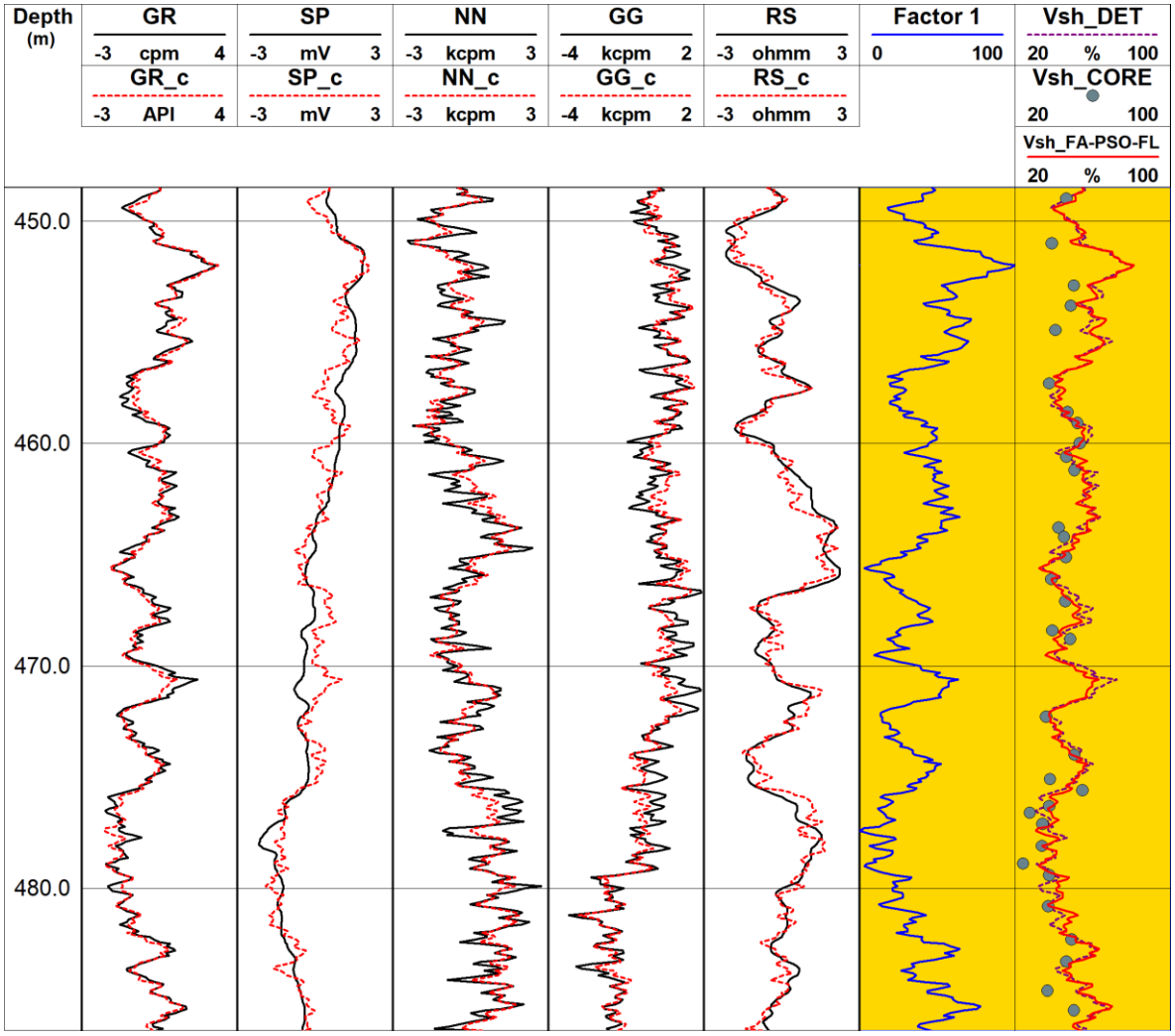

Fig. 39 Resultant logs of the fully PSO based factor analysis. Measured (black) and calculated logs (red dashed line) in tracks 1-5, scaled first factor in track 6 and the resultant shale volumes in track 7 in Well-A

The fit between the measured and calculated logs is fairly good. In the sixth track, the scaled first factor log is shown in blue and the last track contains the estimated shale volumes. The deterministically calculated shale volume (Larionov 1969) is plotted in purple dashed line and the one estimated by factor analysis is shown in red. The fit between the two is quite good, which verifies the applicability of the method. Shale volumes are also confirmed by laboratory measurements (dots).

DOI: 10.14750/ME.2020.007

5.2 Summary

In this chapter, I have further improved the particle swarm optimization based factor analysis (FA-PSO) suggested in Chapter 1.2. To give a fully optimized solution for the model of factor analysis, I have implemented an inner loop into the FA-PSO method, in which the factor loadings are optimized based on the misfit between measured well logs and calculated data ($\tilde{\mathbf{L}}\mathbf{f}$). The tests on an actual wireline logging dataset indicate that the data distance in the data space can be slightly decreased compared to the normal FA-PSO solution. This also results in a moderate improvement for the correlation coefficient between the first extracted factor and shale volume.

Thesis 5.

I have further developed the particle swarm optimization-based factor analysis algorithm (FA-PSO) given in Thesis 1 by iteratively re-calculating the factor loadings. The factor scores and factor loadings are optimized simultaneously in a stable inversion procedure. The newly developed algorithm of factor analysis - named FA-PSO-FL - optimizes the factor loadings as well by PSO based on the misfit between the observed and calculated well logging data. By processing in situ data, I proved the feasibility of the suggested method in Hungarian groundwater formations, and estimated the shale volume directly from the global optimization-derived factor scores. The estimation result was also confirmed by core measurements.

DOI: 10.14750/ME.2020.007

6. SUMMARY

In my PhD thesis, I have summarized my method developments for the advanced processing of wireline logging data that I have carried out at the Department of Geophysics, University of Miskolc. My aim was to develop advanced data processing methods that can be applied on actual well logging datasets. For testing purposes, I gathered datasets ranging from groundwater formations to hydrocarbon bearing formations. I have carried out all method developments introduced in this thesis in MATLAB environment.

As a first step, I have modified the mathematical algorithm of factor analysis. I have developed methods where the factor scores are estimated by Metropolis algorithm-based simulated annealing and the particle swarm optimization technique based on the misfit of measured and calculated data. This approach gives an alternative way to extract the factors from well logging datasets in a reliable way. Based on previous studies, I have shown that the first factor extracted by the developed FA-SA and FA-PSO methods strongly correlates to shale volume of different geological formations. The relationship between the first factor and the shale volume of the investigated formation can be readily established by regression analysis.

I have determined the correlation relationship between the first factor estimated by the particle swarm optimization based factor analysis (FA-PSO) and the decimal logarithm of permeability of hydrocarbon bearing formations. The method has been tested on Hungarian hydrocarbon bearing formations using both core plug and deterministic modeling derived permeability data. It is proved that the developed method can provide an independent estimate for permeability in oilfield practice.

I have developed a particle swarm optimization based interval inversion method for estimating the petrophysical parameters of inhomogeneous layers from wireline logging data. The implementation of the metaheuristic approach almost completely eliminates the starting model dependence of the inverse problem. Thus the starting model does not need to be very close to the optimal solution as in case of linearized inversions. I have also increased the overdetermination ratio of the interval inversion method by incorporating the particle swarm optimization assisted factor analysis (FA-PSO) derived shale volume log into the inversion procedure, thus improving the estimation accuracy of the model parameters.

I have further developed the suggested FA-PSO method by using a hyperparameter estimation approach in which the optimal values of control parameters c_1 and c_2 are set in an automated way by simulated annealing in an outer iteration loop (FA-PSO-CC). These

parameters control the movement of the solution candidates in the search space. Thus I generalized the FA-PSO method in regard of these two control parameters, so there is no need to set them manually in the initialization phase.

To offer an optimal solution for both the factor scores and factor loadings by the PSO based factor analysis, I have built in an inner loop where the factor loadings are recalculated along with the factor scores. The newly developed algorithm (FA-PSO-FL) optimizes the factor loadings as well by PSO based on the misfit between the observed and calculated data. Thus both the factor scores and factor loadings are optimized simultaneously in an iterative inversion procedure to further improve the results of factor analysis.

In the framework of my PhD thesis I carried out inversion based method developments in MATLAB environment for the advanced processing of wireline logging data. The developed methods are capable to effectively process wireline logging datasets and to estimate the petrophysical parameters of formations which are necessary for the quantitative assessment of hydrocarbon and mineral resources. The improved factor analysis based data processing methods (FA-SA, FA-PSO, FA-PSO-CC, FA-PSO-FL) can provide a reliable and independent estimate to shale volume and permeability of sedimentary formations through regression analysis. Once the regression relationships are found between the first factor extracted from a well logging dataset and the petrophysical parameters (i.e., shale volume, permeability) of a given formation in a specific area, the relationships might be used in neighboring wells as well as an independent estimate to these parameters. This can reduce the operating costs, such as the need for taking more core samples to determine the before mentioned petrophysical parameters. The suggested globally optimized interval inversion method aided by factor analysis can estimate the petrophysical parameters from well logging data in a highly overdetermined procedure, which results in more accurate estimations of the desired petrophysical parameters. Since these parameters are the basis for calculating movable hydrocarbons 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. In the future I intend to use this procedure 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.

DOI: 10.14750/ME.2020.007

7. ÖSSZEFOGLALÁS

A PhD dolgozatomban a Miskolci Egyetem Geofizikai Tanszékén elvégzett fúrólyuk-geofizikai módszerfejlesztéseimet foglaltam össze. A kutatómunkám célja olyan korszerű kiértékelési eljárások fejlesztése volt, melyek a terepi adatrendszerek feldolgozásánál is hatékonyan tárják fel az adatrendszerben rejlő földtani információt. Az új lyukgeofizikai módszerek alkalmazhatóságát vízkutatófúrás és szénhidrogén-kutatófúrások adatrendszerén tesztelem. Az értekezésben bemutatott módszerfejlesztések mindegyikét MATLAB környezetben fejlesztettem.

A dolgozatban elsőként a faktoranalízis matematikai algoritmusát fejlesztettem tovább. A faktorérékek meghatározására két új módszert vezettem be, melyekben a faktorértékeket simulated annealing és particle swarm optimization eljárásokkal határozom meg a mért és számított adatok távolsága minimalizálása révén. A kifejlesztett módszerek megbízhatóan alkalmazhatóak a faktorértékek fúrólyuk-geofizikai adatrendszerekből történő meghatározására. Korábbi tanulmányok eredményeire támaszkodva bizonyítottam, hogy a kifejlesztett FA-SA és FA-PSO módszerekkel származtatott első faktor erősen korrelál különböző földtani szerkezetek agyagtartalmával. A vizsgált szelvényezési szakaszon az első faktor és az agyagtartalom közötti korrelációs kapcsolat regresszióelemzéssel könnyen meghatározható.

A kifejlesztett globális optimalizáción alapuló faktoranalízis (FA-PSO) alkalmazásával meghatároztam a korrelációs kapcsolatot fúrólyuk-geofizikai adatrendszerekből származtatott első faktorszelvény és szénhidrogén-tároló formációk áteresztőképességének tízes alapú logaritmusa között. A módszer alkalmazhatóságát hazai szénhidrogén-kutatófúrásokban mért mélyfúrási geofizikai adatrendszereken és magadatokon teszteltem, mellyel bizonyítottam, hogy a kifejlesztett eljárással a gyakorlatban független becslés adható az áteresztőképességre.

Particle swarm optimization alapú intervallum inverziós módszert fejlesztettem, amellyel inhomogén rétegek petrofizikai paraméterei határozhatók meg fúrólyuk-geofizikai adatokból. A kifejlesztett meta-heurisztikus módszer nagymértékben csökkenti az inverz probléma startmodell-függőségét. Megnöveltem az intervallum inverzió túlhatározottságát a faktoranalízisből (FA-PSO) származtatott agyagtartalom szelvény inverzióba való beintegrálásával, így növelve az inverzió által becsült modellparaméterek megbízhatóságát.

Tovább javítva a kifejlesztett FA-PSO módszer hatékonyságát, hiperparaméter becslésen alapuló eljárást fejlesztettem, melyben a c_1 és c_2 vezérlőparaméterek

DOI: 10.14750/ME.2020.007

megválasztása automatikusan történik egy külső ciklusban simulated annealing segítségével (FA-PSO-CC). Ezen paraméterek nagymértékben befolyásolják az egyedek mozgását a paramétertérben. A javasolt eljárással elkerülhető, hogy értéküket empirikus összefüggések alapján kelljen megválasztanunk az eljárás inicializálásakor.

Továbbfejlesztettem a PSO alapú faktoranalízist egy belső ciklus beépítésével a faktorsúlyok iteratív újraszámításhoz. Az újonnan fejlesztett eljárás (FA-PSO-FL) a faktorsúlyokat is a PSO eljárással optimalizálja a mért és számított adatok eltérésének minimalizálása alapján. Így a faktorértékek és a faktorsúlyok szimultán optimalizálásával összességében optimálisabb megoldást kínálva.

A PhD dolgozatomban inverziós módszerfejlesztést végeztem fúrólyuk-geofizikai adatok hatékony és korszerű feldolgozása céljából. Az inverziós algoritmusokat MATLAB fejlesztési környezetben implementáltam. A kifejlesztett módszerek alkalmasak a különböző földtani képződmények petrofizikai paramétereinek a meghatározására, melyek segítségével lehetőség nyílik az ásványi nyersanyagkészletek meghatározására. A továbbfejlesztett faktoranalízis alapú módszerekkel (FA-SA, FA-PSO, FA-PSO-CC, FA-PSO-FL) független, megbízható becslés adható az üledékes rétegek agyagtartalmára és permeabilitására regresszióanalízisen keresztül. Fúrólyuk-geofizikai szelvényekből kinyert első faktor és adott földtani formáció petrofizikai paraméterei (agyagtartalom, permeabilitás) között fennálló regressziós függvény ismeretében, a modell kiterjeszthető egy nagyobb mérési területre, amely által független becslés adható e paraméterekre. Így akár az ezen petrofizikai paraméterek meghatározására irányuló magmintavételek száma is csökkenthető. A kifejlesztett faktoranalízissel támogatott globális optimalizáción alapuló intervallum inverziós eljárás a petrofizikai paraméterek becslését egy nagymértékben túlhatározott feladat keretein belül végzi el, mellyel növelhető a modellparaméterek becslési pontossága és megbízhatósága. A kombinált inverziós eljárásban becsült kőzetfizikai paraméterek kiemelt fontossággal bírnak, mivel a gyakorlatban a kitermelhető szénhidrogén mennyiségét ezek alapján számítjuk. A megnövelt túlhatározottsággal lehetőség nyílik egyes zónaparaméterek inverzión belüli meghatározására is, miközben a becsült modellparaméterek felbontása nem romlik. Az eljárást a jövőben nem-konvencionális, több ásványtípusból felépülő szénhidrogén-tárolók kőzetfizikai modellezésére tervezem alkalmazni, ahol az ismeretlenek száma jóval magasabb, mint a hagyományos tárolók esetében.

DOI: 10.14750/ME.2020.007

ACKNOWLEDGEMENTS

Firstly, I would express my gratitude to Prof. Dr. Norbert Péter Szabó for his efforts as my academic supervisor. He taught me many valuable lessons using his insight and knowledge of geophysics. I am grateful for his guidance; my academic achievements would have not been possible without his advice.

I would also like to thank Prof. Dr. Mihály Dobróka for his advice and suggestions that I could build into my research.

As a member of the Department of Geophysics, I would also like to thank the help and advice of all my colleagues.

Lastly, I would like to thank my family for creating a peaceful environment, in which I could work easily and effortlessly.

For providing the data used in this PhD thesis, I am thankful to:
Labóczki Enid, Mol Hungarian Oil and Gas Plc., Petrophysical Department, Szolnok
Anita É. Csoma, Ph.D., E&P Innovation Lead & Head of Subsurface MOL
The Alaska Oil and Gas Conservation Commission (AOGCC). Accessed the digital well logs at http://doa.alaska.gov/ogc/DigLog/diglogindex.html.

The research was carried out within the GINOP-2.3.2-15-2016-00031 "Innovative solutions for sustainable groundwater resource management" project of the Faculty of Earth Science and Engineering of the University of Miskolc in the framework of the Széchenyi 2020 Plan, funded by the European Union, co-financed by the European Structural and Investment Funds.

The described study was carried out as part of the EFOP-3.6.1-16-2016-00011 "Younger and Renewing University – Innovative Knowledge City – institutional development of the University of Miskolc aiming at intelligent specialisation" project implemented in the framework of the Szechenyi 2020 program. The realization of this project is supported by the European Union, co-financed by the European Social Fund.

REFERENCES

Abordán A., Szabó N.P.: Metropolis algorithm driven factor analysis for lithological characterization of shallow marine sediments. Acta Geodaetica et Geophysica 53: 189. (2018a)

Abordán A., Szabó N.P.: Particle swarm optimization assisted factor analysis for shale volume estimation in groundwater formations. Geosciences and Engineering 6(9):85–95 (2018b)

Abordán A.: Invasive weed optimization driven factor analysis for lithological characterization of aquifers, MultiScience - XXXIII. microCAD International Multidisciplinary Scientific Conference University of Miskolc, 5-6 September, (2018)

Abordán A., Szabó N.P.: Particle swarm optimization based interval inversion of direct push logging data, 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), p. 134–140. (2019b)

Abordán A., Szabó N.P.: Uncertainty reduction of interval inversion estimation results using a factor analysis approach. GEM - International Journal on Geomathematics 11, 11. (2020)

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

Archie G. E.: The electrical resistivity logs as an aid in determining some reservoir characteristics: Transactions of the AIME, 146, 54-62. (1942)

Asquith G., Krygowski D.A.: Basic well log analysis. Second edition, AAPG Methods in Exploration Series (16), The American Association of Petroleum Geologists, Tulsa (2004)

Baker Atlas.: OPTIMA: eXpress reference manual: Baker Atlas, Western Atlas International, Inc. (1996)

Ball S.M., D.M. Chace and W.H. Fertl.: The Well Data System (WDS): An advanced formation evaluation concept in a microcomputer environment: Proceedings of the SPE Eastern Regional Meeting, paper 17034, 61–85. (1987)

Bartlett M.S.: The statistical conception of mental factors: British Journal of Psychology, 28, 97–104 (1937)

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

Dobróka M.: Együttes inverziós algoritmusok bevezetése a mélyfúrási geofizikai értelmezésbe. Kutatási zárójelentés, Miskolci Egyetem, Geofizikai Tanszék, 91 p. (1995)

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)

Douglas W. Oldenburg and Yaoguo Li: 5. Inversion for Applied Geophysics: A Tutorial, Investigations in Geophysics: 89-150. (2005)

Drahos D.: Inversion of engineering geophysical penetration sounding logs measured along a profile. Acta Geodaetica et Geophysica Hungarica 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 Geodaetica et Geophysica Hungarica, (46) 309-316. (2011)

Emily S. Finzel, Kenneth D. Ridgway, Rocky R. Reifenstuhl, Robert B. Blodgett, James M. White, Paul L. Decker: Stratigraphic framework and estuarine depositional environments of the Miocene Bear Lake Formation, Bristol Bay Basin, Alaska: Onshore equivalents to potential reservoir strata in a frontier gas-rich basin. AAPG Bulletin, V. 93, No. 3, P. 379-405. (2009)

Feng Y., Teng G.F., Wang A.X., and Yao Y.M.: Chaotic Inertia Weight in Particle Swarm Optimization. In: Second International Conference on Innovative Computing, Information and Control. p. 475. IEEE (2007)

Geman S., Geman D.: Stochastic relaxation, Gibbs distributions, and Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-6:721-741. (1984)

Holland J.H.: Adaptation in natural and artificial systems. University of Michigan Press, 232 p. (1975)

Ingber L.: Very fast simulated re-annealing. Mathematical and Computer Modelling, 12:8, pp. 967-973. (1989)

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 RC (eds) Factor analysis at 100: historical developments and future directions. Erlbaum, Mahwah, pp 47–77. (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, Proceedings of IEEE international conference on neural networks, 4:1942–1948. (1995)

Klinkenberg L.J.: The permeability of porous media to liquids and gases, Drilling and Production Practice, American Petroleum Inst., pp. 200–213. (1941)

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

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

Marquardt D.W.: Solution of non-linear chemical engineering models. Chemical Engineering Progress. 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)

Mehrabian A.R., Lucas C.: A novel numerical optimization algorithm inspired from weed colonization. Ecological Informatics, 1:4, pp. 355-366. (2006)

Menke W.: Geophysical data analysis: Discrete inverse theory: Academic Press Inc. (1984)

Metropolis N., Rosenbluth M. N., Teller A. H., Teller E.: Equation of state calculations by fast computing machines. Journal of Chemical Physics 21, 1087–1092. (1953)

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., A. Santilano, and A. Godio: Particle swarm optimization of 2D magnetotelluric data. Geophysics, 84 (3), E125-E141. (2019)

Revil A., Cathles L.M.: Permeability of shaly sands. Water Resources Research, 35(3), pp. 651-662. (1999)

Schlumberger: Log interpretation principles/applications. Seventh printing. Schlumberger, Houston (1989)

Schön J.H., Georgi D.T.: Dispersed shale, shaly-sand permeability - a hydraulic analog to the waxman-smits equation. SPWLA 44th Annual Loggin Symposium, June 22-25. (2003)

Sen M.K., and P.L. Stoffa: Global optimization methods in geophysical inversion: Cambridge University Press. (2013)

Serra O.: Fundamentals of well-log interpretation 1. The acquisition of logging data. Developments in Petroleum Science, Vol. 15. Elsevier. (1984)

Shaw R., and S. Srivastava: 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. IEEE World Congress on Computational Intelligence., The 1998 IEEE International Conference on, 69–73. (1998)

Spearman C.: The proof and measurement of association between two things. The American Journal of Psychology, Vol. 15, pp. 72-101. (1904)

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

Szabó N.P., Dobróka M.: Geostatistical Approach for Shale Volume Estimation in Water-bearing Formations, Near Surface 2011 – 17th European Meeting of Environmental and Engineering Geophysics, Leicester, UK, 12-14 September (2011)

Szabó N.P., Dobróka M., Drahos D.: Factor analysis of engineering geophysical sounding data for water-saturation estimation in shallow formations. Geophysics 77, WA35–WA44. (2012)

Szabó N.P., Dobróka M.: Extending the application of a shale volume estimation formula derived from factor analysis of wireline logging data. Mathematical Geosciences 45, 837–850. (2013)

Szabó N.P., Dobróka M., Turai E., Szűcs P.: Factor analysis of borehole logs for evaluating formation shaliness: a hydrogeophysical application for groundwater studies. Hydrogeology Journal 22, 511–526. (2014)

Szabó N.P.: Hydrocarbon formation evaluation using an efficient genetic algorithm-based factor analysis method. 15th European Conference on the Mathematics of Oil Recovery, Mo P071, 12 p. (2016)

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

Szabó N.P., Balogh G. P., Stickel J.: Most frequent value-based factor analysis of direct-push logging data. Geophysical Prospecting, 66(3), pp. 530-548. (2018)

Szabó N.P., Dobróka, M.: Series Expansion-Based Genetic Inversion of Wireline Logging Data. Mathematical Geosciences (2019) 51: 811. (2019)

Szabó N.P., Dobróka M.: Interval inversion as innovative well log interpretation tool for evaluating organic-rich shale formations. Journal of Petroleum Science and Engineering, 186:106696 (2020)

Tarantola A.: Inverse problem theory and methods for model parameter estimation. Society of Industrial and Applied Mathematics (SIAM), Philadelphia (2005)

Timur A.: An investigation of permeability, porosity, and residual water saturation relationships. SPWLA-1968-J, SPWLA 9th Annual Logging Symposium, New Orleans. (1968)

Wiley T.J.: Sedimentary basins of offshore Alaska and adjacent regions. U.S. Geological Survey (1986)

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