

PARTICLE SWARM OPTIMIZATION BASED INTERVAL INVERSION OF DIRECT PUSH LOGGING DATA

Armand Abordán^{1,2}, Norbert Péter Szabó^{1,2**}*

**PhD student, **Associate professor*

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

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

ABSTRACT

An interval inversion approach based on global optimization is suggested for the interpretation of direct-push logging data. To further increase the overdetermination ratio of the inverse problem, the result of factor analysis is incorporated into the inversion procedure. The direct-push logging dataset consists of natural gamma-ray intensity, electrical resistivity, bulk density, neutron-porosity and cone resistance logs. First, factor analysis is carried out to estimate the water content of unsaturated sediments distributed along the borehole. Then, interval inversion is done by utilizing the information of factor analysis on water content to estimate the remaining model parameters such as clay and sand content. Gas content of the studied formation is derived from the inversion results using the material balance equation. It is shown that the factor analysis assisted interval inversion procedure gives highly accurate estimation to the model parameters. As an added advantage of the hybrid method, the starting model dependence of the inversion procedure can be greatly reduced owing to the Particle Swarm Optimization (PSO) technique applied to solve the inverse problem.

1. INTRODUCTION

Determination of geotechnical properties of shallow structures are routinely made by the so-called cone penetration test (CPT). This method was further improved by incorporating geophysical probes into the rod that is pushed into the ground [1]. The developed engineering geophysical sounding (EGS) method is capable of measuring not just cone resistance and sleeve friction but natural gamma-ray intensity, bulk density, neutron porosity and resistivity as well. By processing EGS data, quantitative information can be gathered about the investigated structures, including sand and clay contents, porosity, and water content. By means of factor analysis, water content can be estimated from EGS data [2]. Drahos suggested a local inversion approach [3] for the comprehensive evaluation of all soil parameters mentioned above. For increasing the estimation accuracy and decreasing the noise sensitivity of the estimated model parameters, Balogh [4] suggested an interval inversion method for inverting EGS data, which was originally developed by Dobróka [5] for wireline logging data.

In this paper, PSO is incorporated into the interval inversion procedure in order to reduce its starting model dependence. Once the solution is adequately close to the

optimum, the inversion process is switched to a linearized phase [6] that finds the solution fast and enables the calculation of estimation errors of model parameters. To further increase the overdetermination ratio of the inversion procedure, factor analysis is carried out to estimate water content, which can be then incorporated into the inversion procedure as a known parameter. This way the unknown model parameters to be estimated are only the clay and sand contents, since gas content is derived from the material balance equation.

2. ESTIMATION OF WATER CONTENT BY FACTOR ANALYSIS USING PSO

Factor analysis is a popular statistical tool [7] that is used to reduce a great number of variables into fewer factors. In case of EGS data, the extracted first factor can be related to water content [8] by regression analysis.

In the course of factor analysis, first the direct push logging data is standardized and put into an N -by- K matrix \mathbf{D} , where K denotes the number of direct-push probes and N is the number of measured points along the borehole. Then matrix \mathbf{D} is decomposed as

$$\mathbf{D} = \mathbf{F}\mathbf{L}^T + \mathbf{E}, \quad (1)$$

where \mathbf{F} is an N -by- M matrix of factor scores, M is the number of extracted factors. Matrix \mathbf{L} contains the K -by- M matrix of factor loadings, which describe the correlation between the measured logs and the calculated factors. Quantity \mathbf{E} denotes the N -by- K matrix of residuals. The first column of \mathbf{F} , which is referred to as the first factor log, represents the factor scores explaining the most of the data variance. Factor loadings can be calculated in a non-iterative way that was suggested by Jöreskog [9]

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

where $\mathbf{\Gamma}$ is the diagonal matrix of the first M number of eigenvalues of the sample covariance matrix \mathbf{S} and $\mathbf{\Omega}$ is the matrix of the first M number of eigenvectors, and \mathbf{U} is an arbitrarily chosen M -by- M orthogonal matrix. As it was suggested by Abordán and Szabó [10], it is favorable to estimate the factor scores by a non-linear global optimization scheme using PSO rather than the traditionally used approach suggested by Bartlett [11] that relies on a hypothesis of linearity and Gaussian distributed data.

PSO was developed by Eberhart and Kennedy [12]. It is a metaheuristic optimization approach that was created based on the social behavior of bird flocking and fish schooling. The algorithm itself is very simple yet effective. It utilizes a swarm of particles to solve problems with large search domains. To find the global best solution, particles move around in the search space according to Eqs. (3)–(4) i.e. the position update equation and the velocity update equation. By iteratively updating these equations in each iteration step, PSO converges to the optimal solution. Having an n -dimensional search domain, the position of the i -th particle of the swarm can be denoted by an n -dimensional vector, $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{in})^T$. The velocity of the particle can also be denoted by an n -dimensional vector, $\mathbf{v}_i = (v_{i1}, v_{i2}, \dots, v_{in})^T$. The formerly found best position of the i -th particle is written as $\mathbf{p}_i = (p_{i1}, p_{i2}, \dots, p_{in})^T$. The velocity of the i -th particle is refined using the position and velocity update equations

$$\mathbf{x}_i(t+1) = \mathbf{x}_i(t) + \mathbf{v}_i(t+1), \quad (3)$$

$$\mathbf{v}_i(t+1) = w\mathbf{v}_i(t) + r_1c_1(\mathbf{p}_i(t) - \mathbf{x}_i(t)) + r_2c_2(\mathbf{g}(t) - \mathbf{x}_i(t)), \quad (4)$$

where $t=1, \dots, T$ denotes the iteration steps, and $i=1, \dots, E$ represents the particle index (E is the size of the swarm). C_1 is the cognitive scaling parameter and c_2 is the social scaling parameter, both are set constants with a value of 2. In Eq. (4), parameters r_1 and r_2 are uniformly distributed random numbers from the range of 0 and 1, w represents the inertia weight that was suggested by Shi and Eberhart [13] to better control the optimization procedure. The best position found by the swarm until a given iteration step is stored in vector \mathbf{g} . Fig. 1 illustrates the movement of particle $\mathbf{x}_i(t)$ to its next position $\mathbf{x}_i(t+1)$ by using its own personal best position in the search space $\mathbf{p}_i(t)$, the best position of the whole swarm $\mathbf{g}(t)$ and its velocity $\mathbf{v}_i(t)$.

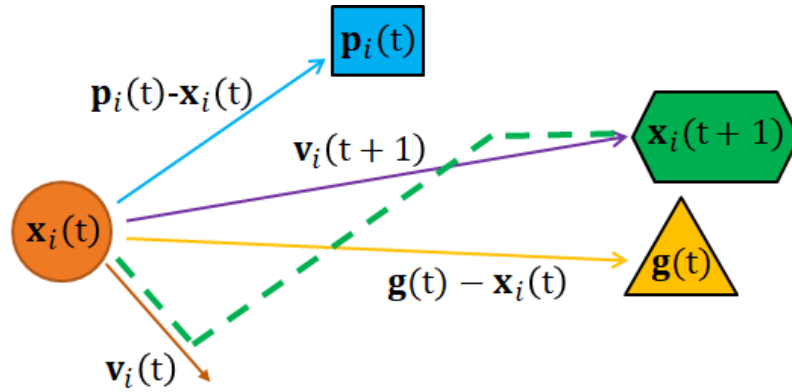


Fig 1.
Searching mechanism of the PSO algorithm

To estimate the factor scores with PSO, first the model of factor analysis in Eq. (1) has to be rearranged to

$$\mathbf{d} = \tilde{\mathbf{L}}\mathbf{f} + \mathbf{e}, \quad (5)$$

where \mathbf{d} denotes the KN length vector of standardized observed data, $\tilde{\mathbf{L}}$ denotes the NK -by- NM matrix of factor loadings, \mathbf{f} is the MN length vector of factor scores and \mathbf{e} is the KN length vector of residuals. Factor loads $\tilde{\mathbf{L}}$ are estimated by Eq. (2) and rotated with the varimax algorithm suggested by Kaiser [14] for having more meaningful factors.

Then the factor scores \mathbf{f} are estimated by the PSO algorithm using Eqs. (3)–(4) and the following objective function that is based on the L_2 norm

$$E = \frac{1}{NK} \sum_{i=1}^{NK} (d_i^{(m)} - d_i^{(c)})^2 = \min, \quad (6)$$

where $\mathbf{d}^{(m)}$ and $\mathbf{d}^{(c)}$ are the measured and calculated (standardized) direct push logging data vectors, respectively. In the course of the global optimization phase, $\tilde{\mathbf{L}}\mathbf{f}$ represents the calculated data and \mathbf{d} denotes the measured data. In the first step of finding the optimal values of the factor scores \mathbf{f} , a random population of particles is generated within the search space with uniform distribution. We set the inertia weight w in each iteration step according to Feng et al. [15]. The suggested chaotic descending inertia weight is as follows: a random number z is generated in the range from 0 to 1. Then by logistic mapping, z is set according to $z=4z(I-z)$ and then the inertia weight can be calculated as

$$w = (w_1 - w_2)(q_{\max} - q) / q_{\max} + w_2 z, \quad (7)$$

where w_1 and w_2 are the initial and the final value of inertia weight, respectively. The maximal number of iteration steps is denoted by q_{\max} and q is the current iteration step. Then in each iteration step, the positions of the particles are updated according to Eqs. (3)–(4), and Eq. (6) is recalculated with the new values of factor scores \mathbf{f} . After the last iteration step, \mathbf{f} is accepted as the optimal solution of factor scores. The first factor extracted by the above detailed method can be directly connected to the water content of the investigated structures. By regression analysis between the first factor (F_1) and the water content, an exponential relationship was earlier found by Szabó et al. [2]

$$V_w = ae^{(bF_1)} + c. \quad (8)$$

2. LOCAL INVERSION OF EGS DATA BY PSO

First, the vector of model parameters needs to be defined, which is estimated depth-by-depth in different inversion runs. For this study, the model includes clay (V_{cl}), sand (V_s), and water (V_w) both for local and interval inversion

$$\vec{m} = [V_{cl}, V_s, V_w]^T, \quad (9)$$

where T stand for transpose. For limiting the number of unknowns, gas content (V_g) is determined from the material balance equation as

$$V_g = 1 - (V_{cl} + V_s + V_w). \quad (10)$$

The model parameters defined in Eq. (9) are the unknowns to be estimated during the inversion procedure. For calculation of theoretical logs depth-by-depth, the following response equations are used. GR for gamma-ray intensity (kcpm), DEN for bulk density (g/cm^3) Φ_N for neutron-porosity (v/v) and RES for resistivity (ohmm)

$$GR = V_{cl}GR_{cl} + V_sGR_s, \quad (11)$$

$$DEN = V_w\rho_w + V_{cl}\rho_{cl} + V_s\rho_s, \quad (12)$$

$$\Phi_N = V_w \Phi_{N,w} + V_{cl} \Phi_{N,cl} + V_s \Phi_{N,s}, \quad (13)$$

$$RES = a(V_w + V_g + V_{cl})^{-m} \left(\frac{V_{cl}/(V_w + V_{cl})}{R_{cl}} + \frac{1 - [V_{cl}/(V_w + V_{cl})]}{R_w} \right)^{-l} \left(\frac{V_w + V_{cl}}{V_w + V_g + V_{cl}} \right)^{-n}, \quad (14)$$

where rock constituents and pore fluids are *cl* (clay), *s* (sand), *w* (water) and *g* (gas). The zone parameters a , m and n are the tortuosity factor, cementation exponent and saturation exponent, respectively and are taken as fixed constants during the inversion procedures. The theoretical logs calculated by the above equations are collected into a column vector

$$\mathbf{d}^{(c)} = [GR^{(c)}, DEN^{(c)}, \Phi_N^{(c)}, RES^{(c)}]^T, \quad (15)$$

and the observed data is also collected into a column vector

$$\mathbf{d}^{(m)} = [GR^{(m)}, DEN^{(m)}, \Phi_N^{(m)}, RES^{(m)}]^T. \quad (16)$$

Then the deviation of vectors (15) and (16) is minimized to find the optimal values of the model parameter vector. This is done by minimizing Eq. (17) by the PSO algorithm. In case of linearized inversion, this is usually done by the damped least squares (DLSQ) method [16].

$$E = \sqrt{\frac{1}{N} \sum_{i=1}^N \left(\frac{\mathbf{d}_i^{(m)} - \mathbf{d}_i^{(c)}}{\mathbf{d}_i^{(m)}} \right)^2} \cdot 100[\%]. \quad (17)$$

3. INTERVAL INVERSION OF EGS DATA BY PSO

In the course of interval inversion, the petrophysical parameters are assumed to be the functions of depth, which are estimated in one inversion procedure. The model parameters given in Eq. (9) are discretized by series expansion

$$m_i(z) = \sum_{q=1}^{Q_i} B_q^{(i)} \psi_q(z), \quad (18)$$

where $B_q^{(i)}$ is the q -th series expansion coefficient describing the i -th model parameter ($i=1,2,\dots,M$), where M is the number of model parameters. Q_i is the number of expansion coefficients and ψ_q is the q -th depth-dependent basis function. As basis functions, Legendre polynomials are used because of their orthogonality. They can be written for the zero-th, first and n -th degree as

$$P_0(z) = 1, \quad (19)$$

$$P_1(z) = z, \quad (20)$$

$$P_n(z) = \frac{1}{n} \{(2n-1)zP_{n-1}(z) - (n-1)P_{n-2}(z)\}, \quad (21)$$

Thus Eq. (18) can be written as

$$m_i(z) = \sum_{q=1}^{Q_i} B_q^{(i)} P_{q-1}(z). \quad (22)$$

Now the unknowns of the inverse problem are the series expansion coefficients, which are a lot lower in number than the number of observed data, therefore the overdetermination ratio is increased compared to that of local inversion.

To find the optimal solutions of the series expansion coefficients, an objective function needs to be defined that measures the distance between the measured and calculated data

$$D = \frac{1}{S} \sum_{s=1}^S \left(\sqrt{\frac{1}{N} \sum_{i=1}^N \left(\frac{d_i^{(m)} - d_i^{(c)}}{d_i^{(m)}} \right)^2} \right) \cdot 100(\%), \quad (23)$$

where S denotes the number of measured points and N is the number of logging tools. To reduce the starting model dependence of the procedure, first Eq. (23) is minimized by PSO until an adequately good solution is found and then the search is switched to a fast, linearized inversion phase, so that the estimation error of model parameters can be calculated. In the linearized phase, the so called damped least squares method (DLSQ) is used. Once the series expansion coefficients are estimated, the model parameters can be calculated using Eq. (22). In order to determine the estimation error of model parameters, first we have to derive the covariance matrix of the model parameters from the error of series expansion coefficients

$$[\text{cov}\mathbf{m}(z)]_{ij} = \sum_{n=1}^{Q^{(i)}} \sum_{m=1}^{Q^{(j)}} P_{n-1}(z) (\text{cov}\mathbf{B})_{hh'} P_{m-1}(z), \quad (24)$$

where \mathbf{B} denotes the vector of series expansion coefficients, indices are $(i=1,2,\dots,M; j=1,2,\dots,M; h=n+Q_1+Q_2+\dots+Q_{i-1}; h'=m+Q_1+Q_2+\dots+Q_{j-1})$. Using Eq. (24) one can derive the correlation coefficient of model parameters

$$\text{corr}(\mathbf{m}(z))_{ij} = \frac{\text{cov}(\mathbf{m}(z))_{ij}}{[\text{var}(\mathbf{m}(z))_i \text{var}(\mathbf{m}(z))_j]^{1/2}}, \quad (25)$$

which informs about the strength of correlation of estimated petrophysical parameters and their reliability.

4. FIELD RESULTS

The dataset consists of natural gamma-ray intensity, resistivity, bulk density, neutron-porosity and cone resistance logs. First, factor analysis was carried out using all available logs with the method detailed in section 1. Factor loadings were estimated by Eq. (2) for 2 factors. The resultant loadings for the first factor are $L^{RCPT} = -0.2403$, $L^{GR} = 0.1231$, $L^{DEN} = 0.7730$, $L^{NPHI} = 0.8878$, $L^{RES} = -0.8969$, which analogously to the correlation coefficient shows the degree of dependence between the factors and input data. It can be seen, that the first factor is mainly influenced by the bulk density, neutron-porosity and the resistivity log. Then, the factor scores are estimated by the minimization of Eq. (6) by the algorithm of PSO. The number of unknown variables is 472. In this case, a random population of 150 particles is generated within the search space with uniform distribution. For Eq. (7) that controls the search mechanism, w_1 is set to 0.6 and w_2 is 0.4. Then the algorithm minimizes the objective function defined in Eq. (6) by using Eqs. (3)–(4). After 15,000 iteration steps Eq. (6) reached a minimum at 19.78 [-].

Water content was estimated by the local inversion method detailed in section 2. Then, by regression analysis between the first standardized factor and the estimated water content, the following local relationship was found

$$V_w = 0.25e^{(0.46F_1)} - 0.12. \quad (26)$$

After completing the phase of factor analysis, interval inversion of the dataset was performed as detailed in section 3. All model parameters, clay (V_{cl}), sand (V_s), and water content (V_w) were discretized by Legendre polynomials of 40 degree. Thus, the number of unknowns is 123 and the number of measured data is 944, therefore the overdetermination ratio is 7.67 (in contrast to local inversion, where the data-to-unknowns ratio is just 4/3). First, Eq. (23) was minimized in 150 iteration steps by PSO, then the inversion process was switched to a linearized phase using the DLSQ method. The convergence to the optimal solution can be seen in Fig. 2 in red.

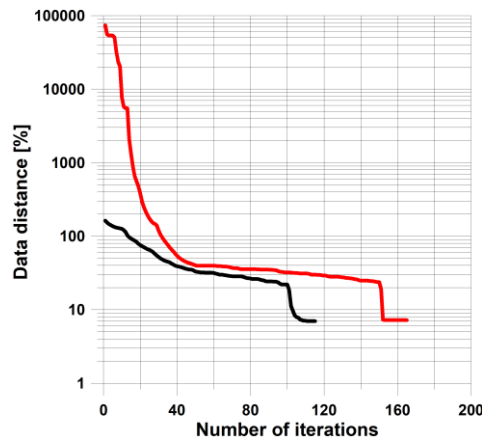


Fig. 2

Convergence of the interval inversion procedure made for the three model parameters (red), and for only two model parameters (black)

After the last iteration step, the data distance was 6.97%. Fig. 3 shows the raw logs and their assumed uncertainty intervals and the estimated model parameters. The average estimation error of each model parameter is $V_{cl}=0.031$ v/v, $V_s=0.052$ v/v, $V_w=0.024$ v/v for the measured interval.

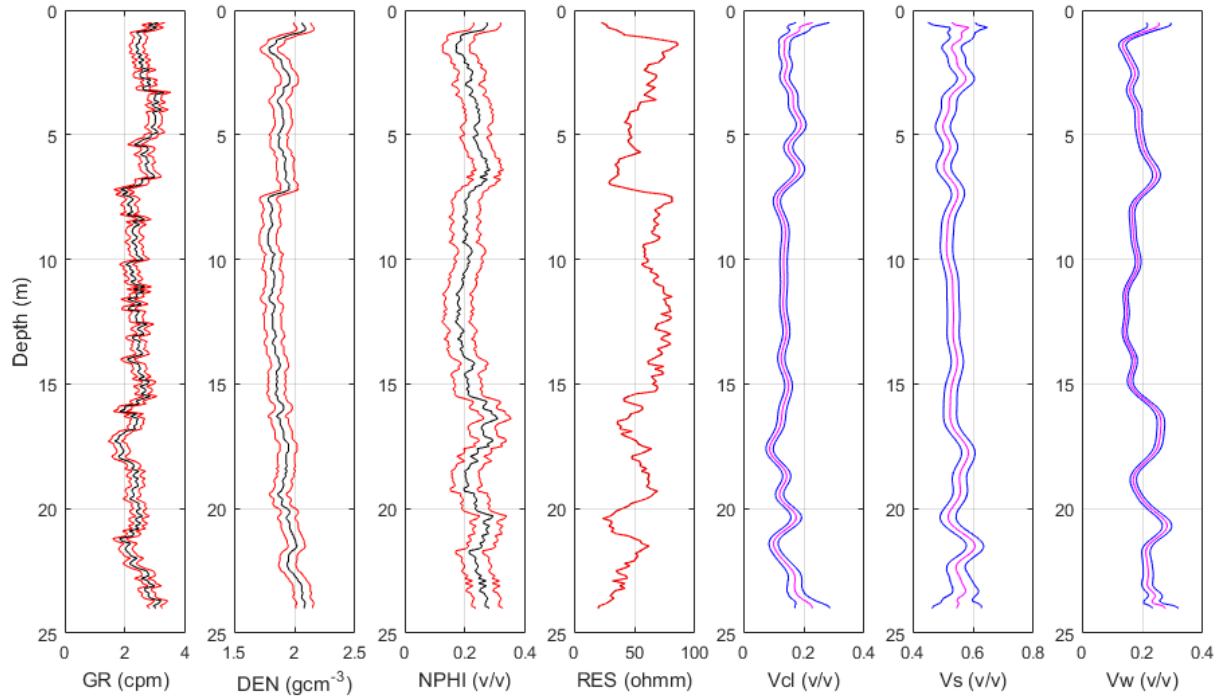


Fig. 3

Input direct push logs (black) and their measurement accuracies (red) in first 4 track, and resultant model parameters (magenta) and their estimation errors (blue) in the last 3 track

In the next step, the same interval inversion procedure was done, but only seeking two model parameters. This time the results of factor analysis was also utilized. The estimated V_w by factor analysis was considered as known values during the inversion procedure beside the measured logs. First, PSO ran for a hundred iterations and then DLSQ found the optimum solution as it can be seen in Fig. 2 in black.

As only two model parameters were estimated (V_s and V_{cl}), the number of unknowns were just 82, thus the overdetermination ratio increased to 14.39 and therefore the initial data distance was also lower as it can be seen if Fig. 2. The results of interval inversion can be seen in Fig. 4 for V_s and V_{cl} and V_w estimated by factor analysis using Eq. (26).

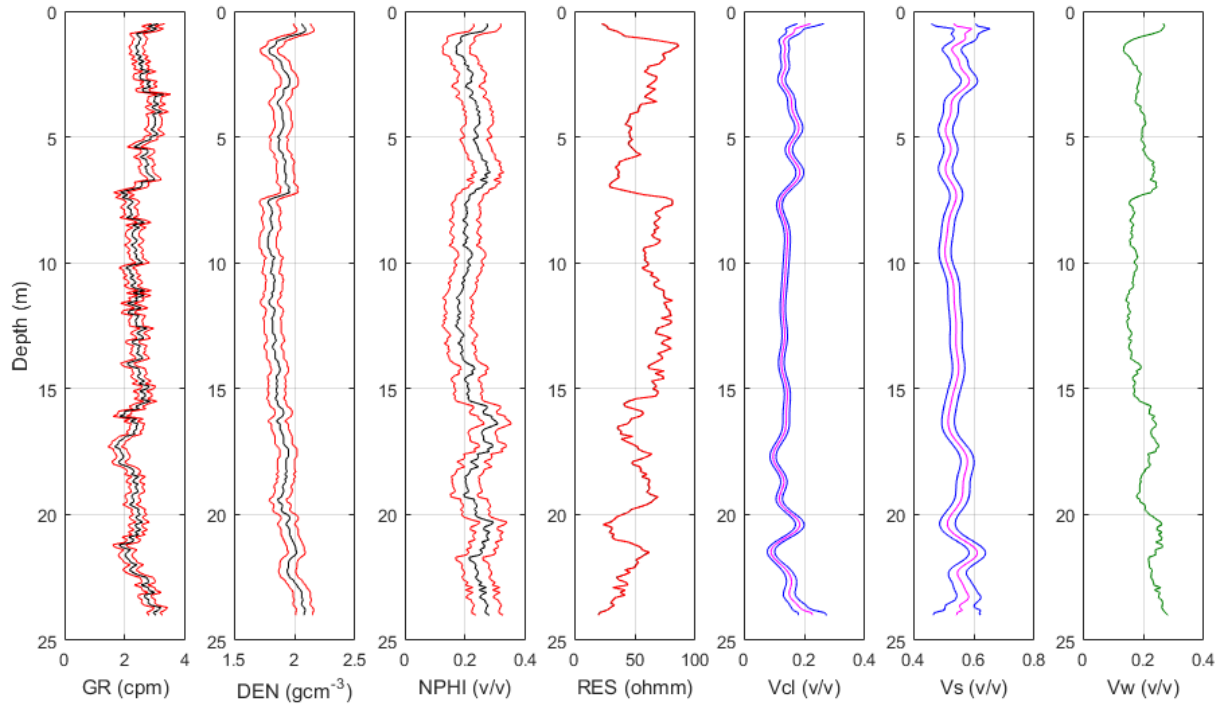


Fig. 4

Input direct push logs (black) and their measurement accuracies (red) in first 4 track, resultant model parameters (magenta) and their estimation errors (blue) in tracks 5-6 and water content estimated by factor analysis in the last track (green)

The average estimation error of the model parameters are $V_{cl}=0.025$ v/v and $V_s=0.050$ v/v for the whole interval. This indicates 20% improvement in the estimation error of clay content and 3% for sand content.

CONCLUSIONS

The presented study shows the results of interval inversion of direct push logging data aided by factor analysis. By initializing the starting model of the inversion by PSO, getting stuck in local minima can be avoided. In addition to the initial model independence of the inversion procedure, there is a possibility to improve the estimation accuracy of inversion results. By incorporating the factor analysis derived water content log into interval inversion as unvarying quantity in the iteration process, the overdetermination ratio can be further increased. In this study, by an increase of 6.72. The estimation error of model parameters decreased, clay content improved 20% and sand content improved 3%. A more accurate estimation to clay content gives a more reliable result for effective porosity and hydraulic conductivity underlying a better characterization of near-surface formations.

ACKNOWLEDGMENT

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.

REFERENCES

- [1] Fejes I. and Jóna E. 1990. **The engineering geophysical sounding method. Principles, instrumentation, and computerised interpretation.** In: Geotechnical and environmental geophysics, Environmental and groundwater, Vol. 2 (ed. S.H. Ward), pp. 321–331, SEG, ISBN 978-0-931830-99-0.
- [2] Szabó N.P., Balogh G.P., Stickle J 2018. **Most frequent value-based factor analysis of direct-push logging data.** Geophys Prospect 66:530–548
- [3] Drahos D. 2005. **Inversion of engineering geophysical penetration sounding logs measured along a profile.** Acta Geodaetica et Geophysica 40, 193–202.
- [4] Balogh G.P. 2016. **Interval inversion of engineering geophysical sounding logs.** Geosciences and Engineering, Vol. 5, No. 8, pp. 22-31.
- [5] Dobróka M., Szabó N.P., Tóth J. and Vass P. 2016. **Interval inversion approach for an improved interpretation of well logs.** Geophysics 81, (2), D155–D167.
- [6] Menke, W. 1984. **Geophysical Data Analysis: Discrete Inverse Theory.** Elsevier.
- [7] Lawley D.N. and Maxwell A.E. 1962. **Factor analysis as a statistical method.** The Statistician 12, 209–229.
- [8] Szabó N.P., Dobróka M. and Drahos D. 2012. **Factor analysis of engineering geophysical sounding data for water saturation estimation in shallow formations.** Geophysics 77, (3), WA35–WA44.
- [9] 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, 2007, pp 47–77
- [10] Abordán A., Szabó N.P. 2018. **Particle swarm optimization assisted factor analysis for shale volume estimation in groundwater formations,** Geosciences and Engineering, Vol. 6, No. 9, pp. 87-97.
- [11] Bartlett M.S. 1937. The statistical conception of mental factors. Br. J. Psychol., 28, 97-104.
- [12] Kennedy J, Eberhart R.C. 1995. **Particle swarm optimization,** Proceedings of IEEE international conference on neural networks, 4:1942–1948.
- [13] Shi Y., Eberhart R. 1998. **Modified particle swarm optimizer, Evolutionary Computation Proceedings.** IEEE World Congress on Computational Intelligence., The 1998 IEEE International Conference on, 1998, 69–73.
- [14] Kaiser H.F. 1958. **The varimax criterion for analytical rotation in factor analysis,** Psychometrika, 23, 187–200.
- [15] Feng Y, Teng G.F, Wang A.X, Yao Y.M. 2007. **“Chaotic Inertia Weight in Particle Swarm Optimization”**, In Innovative Computing, Information and Control.
- [16] Marquardt D.W. 1959. **Solution of non-linear chemical engineering models.** Chemical Engineering Progress 55, 65–70.