JrnlID 10040_ArtID 1749_Proof# 1 - 23/02/2018

Hydrogeology Journal
https://doi.org/10.1007/s10040-018-1749-7

PAPER

A genetic meta-algorithm-assisted inversion approach: hydrogeological study for the determination of volumetric rock properties and matrix and fluid parameters in unsaturated formations

Norbert Péter Szabó [1,2]

Received: 6 May 2017 / Accepted: 16 February 2018

Abstract
An evolutionary inversion approach is suggested for the interpretation of nuclear and resistivity logs measured by direct-push tools in shallow unsaturated sediments. The efficiency of formation evaluation is improved by estimating simultaneously (1) the petrophysical properties that vary rapidly along a drill hole with depth and (2) the zone parameters that can be treated as constant, in one inversion procedure. In the workflow, the fractional volumes of water, air, matrix and clay are estimated in adjacent depths by linearized inversion, whereas the clay and matrix properties are updated using a float-encoded genetic meta-algorithm. The proposed inversion method provides an objective estimate of the zone parameters that appear in the tool response equations applied to solve the forward problem, which can significantly increase the reliability of the petrophysical model as opposed to setting these parameters arbitrarily. The global optimization meta-algorithm not only assures the best fit between the measured and calculated data but also gives a reliable solution, practically independent of the initial model, as laboratory data are unnecessary in the inversion procedure. The feasibility test uses engineering geophysical sounding logs observed in an unsaturated loessy-sandy formation in Hungary. The multi-borehole extension of the inversion technique is developed to determine the petrophysical properties and their estimation errors along a profile of drill holes. The genetic meta-algorithmic inversion method is recommended for hydrogeophysical logging applications of various kinds to automatically extract the volumetric ratios of rock and fluid constituents as well as the most important zone parameters in a reliable inversion procedure.

Keywords Geophysical methods · Genetic algorithm · Zone parameter · Unsaturated zone · Hungary

Introduction

Direct-push measurements are widely used for the in-situ characterization of shallow water-bearing formations. Geophysical logging tools are usually incorporated in the cone-penetration equipment for a more detailed and reliable site investigation. Nuclear measurements are normally applied to estimate the shale volume, porosity and water saturation in unconsolidated formations. The electrical conductivity log is applicable for a general hydrostratigraphic analysis by giving information on the vertical variations in hydraulic conductivity (Butler 2005). By using the surface geoelectric method using special electrode arrays, the drillhole information can be spatially extended and the results of hydrogeophysical interpretation can be effectively improved (Szalai et al. 2015). Detailed facies mapping can also be done by the lateral correlation of electric logs measured in neighboring drill holes (Schulmeister et al. 2003). Recently an inexpensive nuclear magnetic resonance probe was developed to separate the mobile and bound water fractions and provide information on the pore-size distribution and hydraulic conductivity in direct-push drill holes (Walsh et al. 2013). Additional data acquisition techniques such as

✉ Norbert Péter Szabó
norbert.szabo.phd@gmail.com

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

[2] Geoengineering Research Group, |MTA-ME, Miskolc-Egyetemváros 3515, Hungary

hydrogeochemical, seismic and tomographic measurements, and groundwater and soil gas sampling possibilities, are detailed in Kirsch (2006).

A direct-push technology called engineering geophysical sounding (EGS) was developed to solve environmental and groundwater problems in Hungary (Fejes and Jósa 1990). In addition to conventional geotechnical parameters such as cone resistance and sleeve friction, the EGS instruments observe the natural gamma radioactivity, neutron-porosity, bulk density and electrical resistivity in a drill hole. During the measurement, a steel tube is pushed into the ground (Fig. 1a), which is the only material that separates the sensor from the soil. The observed data are transferred to the surface unit through the rods (Fig. 1b). Since there is no invasion of drilling fluid into the formation, the measured data should not be corrected for borehole effects. The EGS data collected by different probes (Fig. 1c) are recorded in digital format and represented by drill hole logs (Fig. 1d).

The geophysical model used in this study for describing the petrophysical characteristics of the unsaturated zone contains two types of parameters. In addition to the fractional volumes of rock constituents and pore-filling fluids, one has to determine a certain number of zone parameters including the physical properties of clay, fluids and minerals, and the textural features of rocks. The parameters of the first group change rapidly with the depth, while the zone parameters are almost constant in the entire domain of the geophysical sounding (the maximum penetration depth is 20–30 m in loose sediments). The model parameters included in the probe response equations are normally estimated by solving a local inverse problem at a given depth. Since having a smaller number of data types than unknowns at the depth point, the zone parameters are normally treated as constant, the initial values of which are a priori given using laboratory data or by literature. If the zone parameters were estimated by local inversion, an underdetermined inverse problem should be solved for an

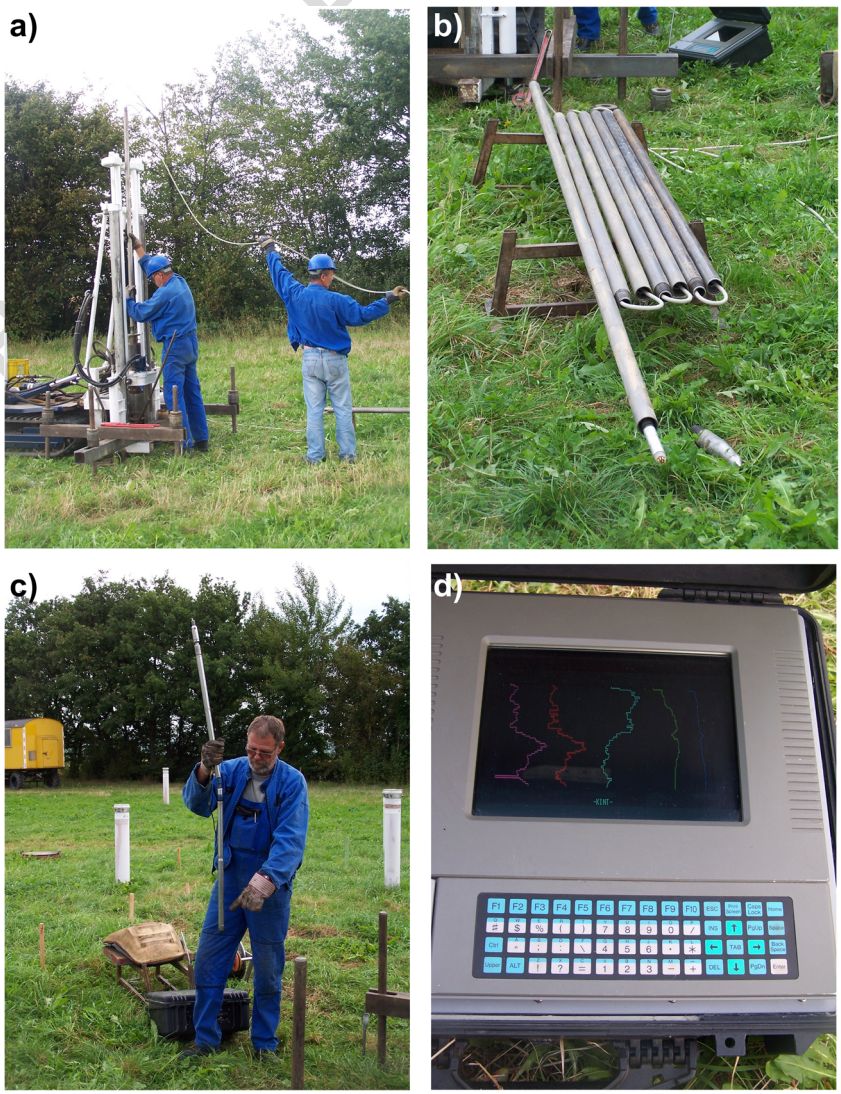

Fig. 1 Engineering geophysical sounding operation in Jülich, Germany: **a** the drill rod is being connected to the others that have been pushed into the ground, **b** wired rods prepared for measurement with a connector and a cone head, **c** combined gamma-ray probe used for the measurement of natural gamma-ray intensity and formation density, **d** real-time visual display of observed parameters on the screen of E1009-type surface unit

infinite number of solutions. In that case, the problem of ambiguity and numerical instability of the inversion procedure could not be avoided. Instead of this, Drahos (2005) suggested solving an overdetermined (local) inverse problem to give an estimate to the volumetric rock properties by fixed zone parameters. In order to compensate for different measurement accuracies and scales, the weighted least squares method is applied (Menke 1984). The preceding EGS inversion approach is very quick, but rather noise sensitive, as there is barely more data than unknowns at a given depth. Moreover, it requires the choice of appropriate zone parameters, which may pose a risk, because the inversion procedure is exposed to uncontrolled modeling error.

So far, there has not been significant progress in the automated determination of zone parameters even in oilfield-well-log analysis. Narayan and Yadav (2006) give a least squares estimate to a limited number of matrix parameters, where only the mean values of observed well logs are optimized in a particular water-bearing horizon. The zone parameters of quartz and water are fixed during the optimization procedure, while those of other minerals (if present) are changed to a maximum of 20% of their initial values originally given by cross-plot techniques. Petrophysical software used by leading oil companies also treat the zone parameters as constant, which have to be set before the processing of well logs. One possible solution is the use of the interval inversion technique, which allows the estimation of suitably chosen zone parameters and layer-thicknesses within the inversion procedure (Dobróka and Szabó 2012). The interval inversion method inverts the data set of a longer depth-interval to predict the vertical distributions of petrophysical parameters in a joint inversion procedure. In the framework of interval inversion, the model parameters including the zone parameters are expanded into a series by using orthogonal polynomials as basis functions, and the expansion coefficients are estimated by inversion. This approach is more stable and gives more accurate results than local inversion (Dobróka et al. 2016). For a high overdetermination (data-to-unknowns) ratio, which is responsible for the significant improvement in estimation accuracy, far fewer expansion coefficients are applied as unknown parameters than inverted data. In that case, however, the vertical resolution of the inversion method is limited. A better resolution does not allow the incorporation of too many unknowns. Dobróka and Szabó (2011) give an estimate to three textural parameters and four volumetric ratios in a Hungarian gas-bearing formation by using a simulated annealing based interval inversion method. Other alternative methods such as factor analysis, can help reduce the number of unknowns as they allow the independent estimation of some petrophysical parameters such as shale volume (Szabó et al. 2014) and hydraulic conductivity (Szabó 2015).

The zone parameters often correlate strongly to each other and volumetric parameters, which may lead to ambiguity in the inverse problem. Balázs (2015) theoretically shows that the amount of correlation between the zone parameters and volumetric properties increases with the number of inversion unknowns in linear inversion. On the other hand, parameter sensitivity experiments show that some zone parameters, like the saturation exponent, hardly bear influence on the observed well logs in fully saturated formations (Dobróka and Szabó 2011). Low or zero sensitivity to the saturation exponent and other zone properties may cause the failure of gradient-searching inversion methods. In addition to this problem, the linearized inversion procedures tend to find a solution at a local minimum of the objective function and are generally badly conditioned from the numerical point of view. In such cases, when a linear inversion approach is inapplicable, one can successfully apply a global optimization method, e.g. simulated annealing or evolutionary algorithms (Sen and Stoffa 2013). Global optimization methods give both a derivative-free and a practically initial-model independent solution, while they seek the absolute extremum of the objective function. The class of genetic algorithms has been used as an effective global optimization tool for biological and engineering problems and artificial intelligence approaches (Cranganu et al. 2015). Similar to natural selection, the genetic search iteratively improves a population of artificial individuals. The population consists of a number of individuals each representing one feasible solution in the model space. In geophysical inversion, each model parameter is coded into a chromosome, the basic elements of which, called genes, are randomly exchanged and modified during the searching process. In advanced genetic algorithms, the model parameters are encoded as real numbers and the genes are randomly modified by real operations. The float-encoded genetic algorithm (FGA) assures a high resolution of the model space and a relatively fast solution (Michalewicz 1992).

An FGA meta-algorithm-assisted EGS inversion method, named as genetic meta-algorithmic inversion (GMI), is suggested for the simultaneous estimation of volumetric parameters and physical properties of clay and matrix components (i.e. zone parameters) using engineering geophysical sounding data. The originally underdetermined EGS inverse problem is solved in two overdetermined inversion phases. In the inner loop of GMI, the volumetric parameters are locally estimated by a linearized inversion algorithm, whereas the zone parameters are determined by the FGA meta-algorithm in the outer loop. The GMI assures the best fit between the measured and calculated data and gives a reliable solution practically independent

of the initial model. Therefore, laboratory information on zone parameters is unnecessary in this stage of interpretation. The GMI is numerically tested using engineering geophysical sounding logs observed in a Hungarian unsaturated loessy-sandy formation to confirm the feasibility of the proposed inversion workflow.

Materials and methods

Forward problem

In the petrophysical characterization of unsaturated (clastic) sediments, the rock matrix is assumed to be composed of coarse and fine grain components and the pore-space is occupied by freshwater and some amount of air. In the GMI procedure, the fractional volumes of water (V_w), air (V_a), clay (V_{cl}) and sand (V_{sd}) are calculated locally at a given depth, while the matrix and clay properties are estimated as constant for the entire length of a drill hole. From the inversion results, one can derive the hydraulic conductivity (Nyári et al. 2010) and other important geotechnical parameters, e.g. undrained shear strength (Shin and Kim 2011) and dry density (Szabó 2012).

In the forward modeling process, theoretical EGS logs are calculated from the known values of model parameters, which are adjusted to observed data by inversion. The input of the inversion procedure is composed of natural gamma-ray intensity (GR), bulk density (ρ_b), neutron-porosity (Φ_N), and resistivity (R) logs. In shallow sediments, the nuclear parameters are calculated by the response functions suggested by Drahos (2005) and Szabó et al. (2012), while the resistivity is calculated by the model of De Witte (1955)

$$\mathrm{GR} = \mathrm{V_{cl}GR_{cl}} + \mathrm{V_{sd}GR_{sd}} \tag{1}$$

$$\rho_b = \mathrm{V_w}\rho_w + \mathrm{V_{cl}}\,\rho_{cl} + \mathrm{V_{sd}}\rho_{sd} \tag{2}$$

$$\Phi_N = \mathrm{V_w}\Phi_{N,w} + \mathrm{V_{cl}}\Phi_{N,cl} + \mathrm{V_{sd}}\Phi_{N,sd} \tag{3}$$

$$\mathrm{R} = (\mathrm{V_w} + \mathrm{V_a} + \mathrm{V_{cl}})^{-m}\left(\frac{\mathrm{V_{cl}}/(\mathrm{V_w}+\mathrm{V_{cl}})}{\mathrm{R_{cl}}} + \frac{1-[\mathrm{V_{cl}}/(\mathrm{V_w}+\mathrm{V_{cl}})]}{\mathrm{R_w}}\right)^{-1}\left(\frac{\mathrm{V_w}+\mathrm{V_{cl}}}{\mathrm{V_w}+\mathrm{V_a}+\mathrm{V_{cl}}}\right)^{-2} \tag{4}$$

$$\mathrm{V_w} + \mathrm{V_a} + \mathrm{V_{cl}} + \mathrm{V_{sd}} = 1 \tag{5}$$

where the zone parameters are gamma-ray intensity in sand (GR_{sd}) and clay (GR_{cl}), density of water (ρ_w), sand (ρ_{sd}) and clay (ρ_{cl}), neutron-porosity of water ($\Phi_{N,w}$), sand ($\Phi_{N,sd}$) and clay ($\Phi_{N,cl}$), resistivity of water (R_w) and clay (R_{cl}), and the cementation exponent (m). Equation (5) represents the material balance in the soil, which derives one of the volumetric parameters from the others.

Estimation of volumetric rock properties

In the local inversion phase (inner loop) of the GMI, the fractional volume of sand, water and clay is estimated in each depth separately. The air volume is estimated outside the inversion procedure using Eq. (5). The column vector of EGS data observed in a given depth is

$$\mathbf{d}^{(obs)} = \left[\mathrm{GR}^{(obs)}, \rho_b^{(obs)}, \Phi_N^{(obs)}, R^{(obs)}\right]^T \tag{6}$$

where T denotes the symbol of matrix transpose. The model vector of the inverse problem defined at the same depth is

$$\mathbf{m} = [V_w, V_{cl}, V_{sd}]^T \tag{7}$$

The theoretical value of the EGS data "measured" by the k-th probe is given by the forward modeling procedure using the relevant choice of Eqs. (1)–(4)

$$\mathrm{d}_k^{(cal)} = f_k(V_w, V_{cl}, V_{sd}, \mathbf{p}) \tag{8}$$

where $\mathbf{p}$ is the vector of zone parameters ($k = 1,\ldots,4$). In the inner loop of the GMI, the zone parameters are assumed to be fixed as constant and a set of marginally overdetermined inverse problems (as having four data against three unknowns) is solved along the drill hole, which has an unique solution. Because of Eq. (4), the set of probe response functions $\mathbf{f}$ is nonlinear; thus, the nonlinear inverse problem is usually linearized (Menke 1984). The linear approximation of data vs. model relation is $\mathbf{d}^{(cal)} = \mathbf{Jm}$, where $\mathbf{J}$ is the 4-by-3 Jacobi's matrix including the partial derivatives of EGS data with respect to the model parameters given in Eq. (7). The distance between the measured and calculated data normalized to the data variances as a measure of instrument uncertainty is given in the 4-by-1 deviation vector $\mathbf{e}$. The objective function of the inverse problem is

$$E = \|\mathbf{e}\|_2^2 + \varepsilon^2\|\mathbf{m}\|_2^2 = \min \tag{9}$$

where ε^2 is a positive regularization parameter used for stabilizing the solution of the ill-posed inverse problem. If one has no prior information on the data variances, the deviation error $\mathbf{e}$ can be normalized by the observed data to take the relative importance of each data type into consideration in the inversion procedure. The actual model is refined until the best fit between the measured and calculated data is achieved. The vector of model parameters is continuously improved by $\mathbf{m} = \mathbf{m}_0 + \delta\mathbf{m}$ in an iterative procedure, where $\mathbf{m}_0$ is the initial model. The model correction vector $\delta\mathbf{m}$ is estimated by the damped least squares method (DLSQ) suggested by Marquardt (1959)

$$\delta\mathbf{m} = \left(\mathbf{J}^T\mathbf{J} + \varepsilon^2\mathbf{I}\right)^{-1}\mathbf{J}^T\delta\mathbf{d} \tag{10}$$

where $\delta\mathbf{d}$ is the difference between the measured and actually computed EGS data vector and $\mathbf{I}$ is the unity matrix.

The EGS inversion method allows the checking of the quality of the estimated model. The covariance matrix of the model parameters estimated by a linearized inversion method can be related to the data covariance matrix including the variances of observed data

$$\mathrm{cov}\mathbf{m} = \mathbf{F}^{-1}\mathrm{cov}\mathbf{d}^{(\mathrm{obs})}\left(\mathbf{F}^{-1}\right)^{\mathrm{T}} \tag{11}$$

where $\mathbf{F}^{-1}$ is the generalized inverse matrix of the DLSQ method (Menke 1984). The estimation error of volumetric parameters is calculated as the square root of variances obtained in the main diagonal of cov $\mathbf{m}$. The reliability of inversion results can be quantified via the Pearson's correlation matrix (corr $\mathbf{m}$) derived from Eq. (11). One can use the following scalar for the measure of average correlation

$$\mathrm{S}(\mathbf{m}) = \left\{\frac{1}{\mathrm{M}(\mathrm{M}-1)}\sum_{\mathrm{u}=1}^{\mathrm{M}}\sum_{\mathrm{v}=1}^{\mathrm{M}}\left[(\mathrm{corr}\mathbf{m})_{uv}-\delta_{\mathrm{uv}}\right]^2\right\}^{1/2} \tag{12}$$

where δ denotes the Kronecker delta, and M is the number of model parameters. The solution is considered to be reliable, when the estimated three fractional volumes correlate poorly. In opposite cases, high correlation causes ambiguity, which prevents the reliable determination of the individual parameters by an inversion procedure.

Estimation of zone parameters

The global optimization phase (outer loop) of GMI provides a heuristic search for the unknown zone parameters. In the FGA inversion meta-algorithm, only the zone parameters are updated, whereas the volumetric properties preliminary estimated in the inner loop of iterations are treated as fixed parameters. The model vector is composed of the matrix and clay properties being the unknowns of the inverse problem

$$\tilde{\mathbf{p}} = \left[\mathrm{GR}_{\mathrm{cl}}, \mathrm{GR}_{\mathrm{sd}}, R_{\mathrm{cl}}, \rho_{\mathrm{cl}}, \rho_{\mathrm{sd}}, \Phi_{\mathrm{N,cl}}\right]^{\mathrm{T}} \tag{13}$$

the constant values of which are estimated for the entire processing interval. To achieve this, all data collected from the drill hole are integrated in one data vector

$$\tilde{\mathbf{d}}^{(\mathrm{obs})} = \left[\mathbf{d}_1^{(\mathrm{obs})}, \mathbf{d}_2^{(\mathrm{obs})}, \ldots, \mathbf{d}_i^{(\mathrm{obs})}, \ldots, \mathbf{d}_N^{(\mathrm{obs})}\right]^{\mathrm{T}} \tag{14}$$

where N is the total number of processed depths. In the genetic search, the lower and upper bounds of zone parameters are first initialized. A number of randomly generated models of zone properties as individuals is simultaneously tested during the optimization process, in which those having parameters out of the physical range are effectively rejected. In the first phase of the GMI, a set of DLSQ inversion runs are executed by fixed values of zone parameters for the determination of the depth distribution of volumetric parameters (inner loop of GMI). Then, the updated values of fractional volumes are fixed, and the zone parameters are re-calculated by fitting the observed and predicted data by the FGA process (outer loop of GMI). The two phases of the GMI are repeated until a stop criterion is met.

In the outer loop of GMI, a population of model vectors $\tilde{\mathbf{p}}_j$ ($j = 1,2,\ldots,P$, where P is the population size) is gradually improved to minimize the misfit between the observed data given in Eq. (14) and those calculated using Eq. (8). For data prediction, the same probe response functions given in Eqs. (1)–(4) are used as in local inversion. The chromosome of an individual is built up from the model parameters of Eq. (13). According to the Darwinian theory of evolution, the fittest individuals survive and reproduce, while others disappear from the population over time. In this study, the fitness of the j-th model is calculated from the normalized data deviation along the entire length of the drill hole

$$\mathrm{F}\left(\tilde{\mathbf{p}}_j\right) = -\left[\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=1}^{K}\left(\frac{\tilde{d}_{ik}^{(\mathrm{obs})} - f_k\left(\mathbf{m}_i, \tilde{\mathbf{p}}_j\right)}{\tilde{d}_{ik}^{(\mathrm{obs})}}\right)^2\right]^{1/2} \tag{15}$$

where $\mathbf{m}_i$ is the vector of volumetric parameters fixed at the i-th depth ($i = 1,2,\ldots,N$) and K is the number of applied probes. The solution of the optimization problem is to be found at $\mathrm{F}(\tilde{\mathbf{p}}) = \max$. The value of fitness multiplied by −100 characterizes the relative distance between the measured and calculated EGS logs in percent. Next, a suitable combination of real genetic operations is used to increase the average fitness of subsequent generations. In order to get a quick solution and best resolution, floating-point genetic operations such as selection, crossover and mutation are implemented (Houck et al. 1995).

At first the fittest individuals are selected for reproduction. In this study, the mating of individuals is performed by tournament selection. In the selection process, a number of individuals are randomly chosen from the population, with replacement, and the fittest ones are copied into the new population. This procedure is repeated until P number of individuals have been selected. This approach is probability-free unlike other selection methods, e.g. roulette wheel or ranking operations (Razali and Geraghty 2011). The number of tournaments as a control parameter is set to ensure the development of convergence of the genetic search. Less adaptable individuals have a smaller chance to be selected for the next generation if this tournament size is large. In the next step, a pair of individuals ($\tilde{\mathbf{p}}_1$ and $\tilde{\mathbf{p}}_2$) is chosen from the actual population that had undergone the selection process and some

amount of genetic information is exchanged between them. By the use of simple or multi-point crossover, the model vectors are divided into two or more segments, the genes of which are exchanged between the crossover points to get new individuals. The heuristic crossover extrapolates two individuals as follows

$$\tilde{\mathbf{p}}_1 = \tilde{\mathbf{p}}_1 + \alpha\left(\tilde{\mathbf{p}}_1 - \tilde{\mathbf{p}}_2\right)$$
$$\tilde{\mathbf{p}}_2 = \tilde{\mathbf{p}}_1 \quad (16)$$

where α is a random number generated from U(0,1). It is assumed that the fitness value of $\tilde{\mathbf{p}}_1$ is higher than that of $\tilde{\mathbf{p}}_2$. If any value of $\tilde{\mathbf{p}}_1$ is out of bounds, a new random number is generated and Eq. (16) is recalculated. After a certain number of failures, the new values of model parameters are set to equal the old ones. The third genetic operation is a uniform mutation. For the mutation process, the model vector $\tilde{\mathbf{p}}_1$ is selected and its l_0-th parameter is substituted with a random floating-point number (β) generated from its pre-defined range

$$\tilde{\mathbf{p}}_1' = \begin{cases} \beta, & \text{if } l = l_0 \\ \tilde{\mathbf{p}}_{1,l}, & \text{otherwise} \end{cases} \quad (17)$$

where $\tilde{\mathbf{p}}_1'$ is the mutated individual. The selection and the preceding genetic operations (Eqs. 16–17) are applied in successive generations until a termination criterion is met. An elitism-based reproduction can be optionally used, which preserves the fittest individual(s) of the previous generation and replaces the weakest with them in the next generation (Bijani et al. 2012). The stop criterion is defined either as the maximum number of generations or a specified threshold in the distance between the measured and calculated data. In the last generation, the model vector with maximum fitness is regarded as the solution of the global optimization problem. The estimation error of a zone parameter is calculated as the standard deviation of its values obtained for different individuals in the last generation.

Feasibility results

Geological setting

The GMI is tested in the Bátaapáti (Üveghuta) site, South-West Hungary (Fig. 2a). Ground geophysical surveys were previously carried out for establishing a nuclear waste repository in the igneous bedrock (Vértesy et al. 2004). The aim of geophysical measurements in the shallow unconsolidated sediments was to detect neotectonic events by tracing horizons connected with soils and explore local hydrogeological conditions. The loessy-sandy formation is deposited on a partially weathered granite basement, where the water table is situated approximately at the top of the granite. The EGS logs were measured in the upper 20–30 m of the loose unsaturated formation in seven drill holes (H4–H10) located 50 m apart from each other along a 300-m-long profile (Fig. 2b). The zone

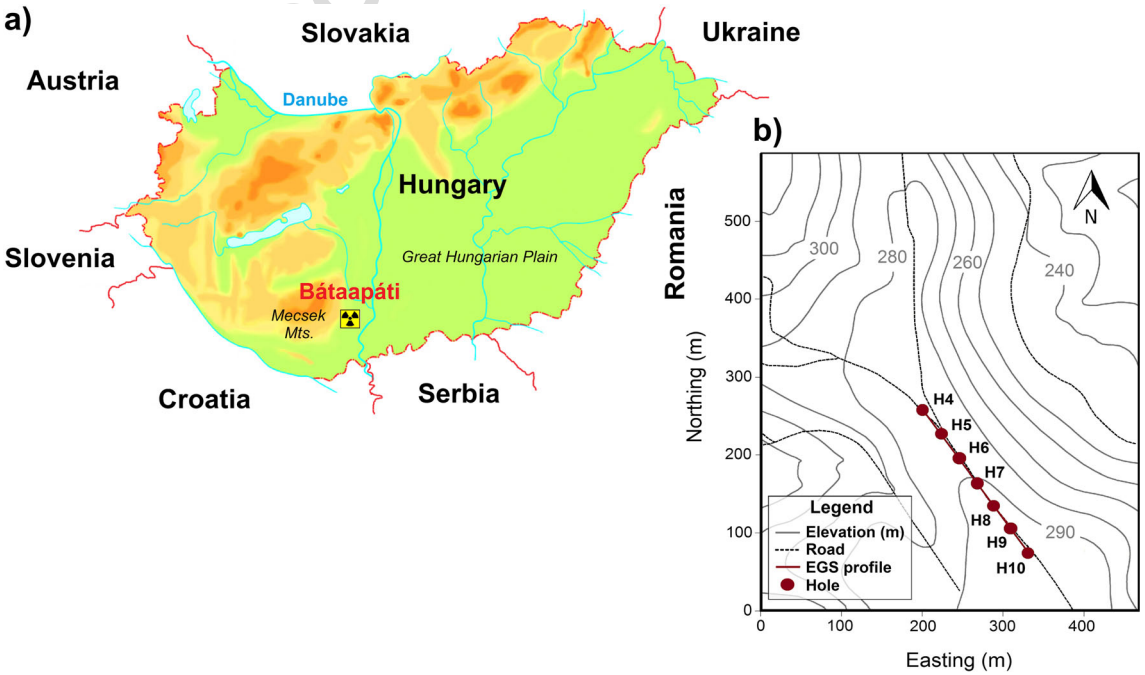

Fig. 2 Location map of the studied area: **a** Bátaapáti well site in South West Hungary, **b** topographic map and the profile of the investigated drill holes

parameters in Eqs. (1)–(4) other than clay and matrix properties are given from the literature (Drahos 2005), i.e. $R_w = 12$ ohmm, $m = 1.7$, $\rho_w = 1.0$ g/cm^3, $\Phi_{N,w} = 1.0$ v/v, $\Phi_{N,sd} = 0$ v/v. In Eqs. (2)–(3), the air volume is not included, because the mass density and neutron-porosity of air are set to zero.

Single borehole application

In a one-dimensional (1D) application of the GMI, the vertical distribution of water, clay and quartz volumes and the constant values of matrix and clay properties are simultaneously estimated in drill hole H4 (Fig. 2b). The input of the inversion procedure is composed of the GR, ρ_b, Φ_N, R logs measured with a sampling interval of 0.1 m. Table 1 shows a moderate correlation between the EGS measurement types on average. The mean of Pearson's correlation coefficients calculated by Eq. (12) is $S = 0.55$. The strongest correlation is indicated between the resistivity and porosity logs. The uncertainty of EGS data represented by the data variances in Eq. (11) are chosen from Drahos (2005) such as $\sigma^2_{GR} = 0.05$ kcpm^2, $\sigma^2_{\rho_b} = 0.01$ g^2/cm^6, $\sigma^2_{\Phi_N} = 2.5 \cdot 10^{-3}$ v^2/v^2, $\sigma^2_R = 0.05$ ohm^{2}m^2. The forward problem is solved using Eqs. (1)–(4) in both phases of the 1D GMI.

In the inner loop of iterations, the volumetric properties are updated by Eq. (10), the initial values of which for each depth are chosen as $V_{w,0} = 0.3$ v/v, $V_{sd,0} = 0.4$ v/v, $V_{cl,0} = 0.15$ v/v. The air volume is directly derived from the EGS inversion results using Eq. (5). The optimal solution is found after 10 iteration steps in each depth. The regularization factor in Eq. (10) is chosen as $\varepsilon^2 = 0$, which shows a stable linearized inversion process. At the end of the local inversion phase, the 1D GMI is switched to the FGA searching (outer) loop. The fractional volumes estimated in the last iteration of local inversion are fixed, and the start population with 30 individuals is initialized (Fig. 3a). The search domain of the matrix and clay properties is listed in Table 2. The fitness of individuals is calculated by Eq. (15), which has local maxima at certain selected values of zone parameters (Fig. 3c–d). The genetic operators including tournament selection, heuristic crossover and uniform mutation are used to find the optimal values of zone parameters. The control parameters of FGA are chosen as tournament size (200), crossover retry (50) and mutation probability (0.05). An elitism-based reproduction is performed, as the vector of zone parameters with the maximum fitness is automatically copied into the next generation. After each local inversion loop including 10 iterations, the FGA procedure runs over 1,000 generations. The total number of iterations of the 1D GMI is set to 10. As a result of the combined use of linear and global inversion methods, the relative distance between the measured and calculated data decreases continuously (Fig. 3b). In the first step of 1D GMI, the relative data distance is 39%, which reduces to 4% at the end of the inversion procedure. The measured and calculated EGS logs as well as the optimal values of volumetric parameters can be found in Fig. 4. The estimation errors of volumetric parameters calculated by Eq. (11) are obtained as $\sigma_{Vw} = 0.6$ v/v, $\sigma_{Vsd} = 5.4$ v/v, $\sigma_{Vcl} = 0.93$ v/v (Fig. 5). The mean correlation between the volumetric parameters calculated by Eq. (12) is $S = 0.51$. As part of the inversion result, the estimated values of zone parameters including their estimation errors are listed in Table 2. The inversion results show stable and reliable inversion procedure. The required CPU time of the 1D GMI using a quad-core processor workstation is 4 min 43 s.

t1.1 **Table 1** Pearson's correlation matrix of engineering geophysical sounding logs measured in drill hole H4

t1.2		GR	ρ_b	Φ_N	R
t1.3	GR	1	0.23	0.11	−0.29
t1.4	ρ_b	0.23	1	0.65	−0.73
t1.5	Φ_N	0.11	0.65	1	−0.83
t1.6	R	−0.29	−0.73	−0.83	1

GR gamma-ray intensity; ρ_b bulk density; Φ_N neutron-porosity; R resistivity logs

Multi-borehole application

The two-dimensional (2D) version of the GMI allows the estimation of the volumetric properties and zone parameters not only along the depth coordinate but also as a horizontal profile of drill holes. It is assumed that the zone parameters do not change significantly in lateral direction. Therefore, the volumetric parameters are determined in each depth by local inversion, whereas the zone parameters are treated as constants in the whole shallow region and estimated by the FGA procedure. The only difference between the 1D and 2D GMIs is that the fitness of the j-th model, in the 2D case, is calculated using the data from all drill holes

$$F^*\left(\tilde{\mathbf{p}}_j\right) = -\left[\frac{1}{HKN^*}\sum_{h=1}^{H}\sum_{k=1}^{K}\sum_{i=1}^{N_h}\left(\frac{\hat{d}^{(obs)}_{hki} - f_k\left(\mathbf{m}_{hi}, \tilde{\mathbf{p}}_j\right)}{d^{(obs)}_{hki}}\right)^2\right]^{1/2} \quad (18)$$

where $\mathbf{m}_{hi}$ is the vector of volumetric parameters fixed at the i-th depth in the h-th hole, N_h is the number of processed depths in the h-th hole and H is the total number of penetration holes. The solution of the inverse problem is to be found at $F^*(\tilde{\mathbf{p}}) = \max$. In the 2D GMI, the GR, ρ_b, Φ_N, R logs measured in drill holes H4–H10 are simultaneously inverted (Fig. 2b). The sampling interval is 0.1 m and the

JnlID 10040_ArtID 1749_Proof# 1 - 23/02/2018

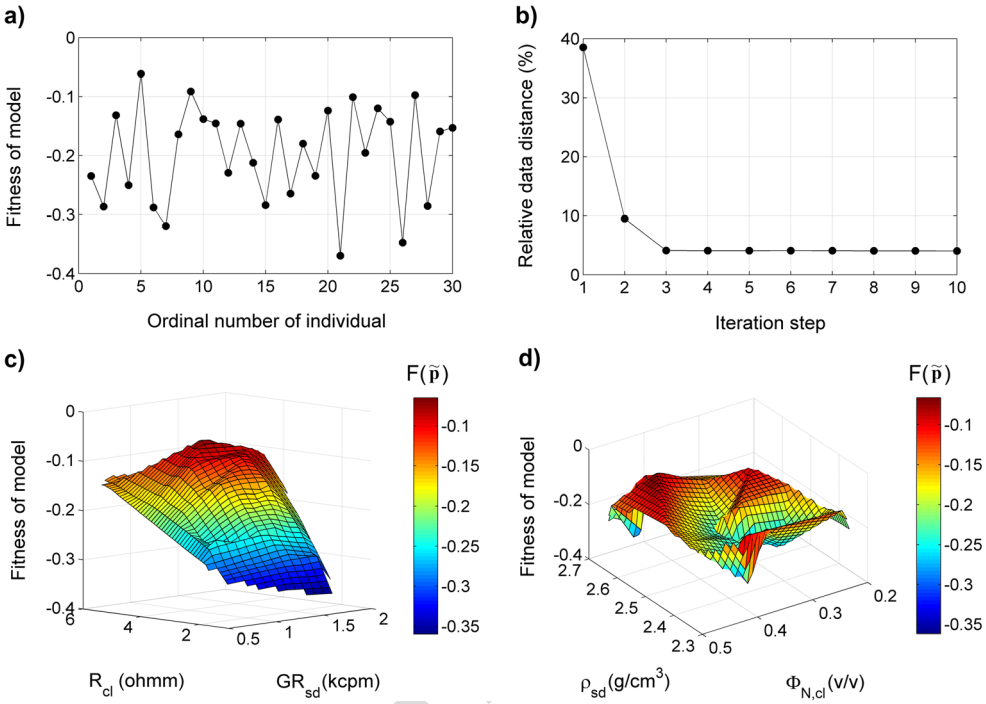

Fig. 3 Genetic meta-algorithmic inversion process in drill hole H4 **a** fitness values of individuals including the zone parameters generated in the initial population, **b** data misfit vs. iteration step during the GMI process, **c** fitness vs. clay resistivity (R_{cl}) and gamma-ray intensity in sand (GR_{sd}), **d** fitness vs. sand density (ρ_{sd}) and neutron-porosity of clay ($\Phi_{N,cl}$)

lengths of drill holes are not exactly the same (the bottom is between 19.9–27.7 m). The total number of EGS data is $N^* = N_1 + N_2 + \ldots + N_H = 6{,}972$. The mean spread calculated by Eq. (12) is $S = 0.27$, which shows weaker correlation between the observed physical quantities than in hole H4. This may be due to the lateral variation of lithology and data noise. The strongest correlation is indicated between the resistivity and neutron-porosity logs (Table 3). The theoretical values of EGS data are calculated using Eqs. (1)–(4), in both phases of the 2D GMI, and the same data variances are assumed as in section 'Single borehole application'.

In the inner loop of iterations, the DLSQ method is used to produce the 2D interpolated sections of volumetric properties (Fig. 6); note that the same control parameters are used as in section 'Single borehole application'. In the FGA search phase, a population of model vectors (13) is improved in 250 generations using the same genetic operations as in section 'Single borehole application'. The fitness of individuals is evaluated using Eq. (18). The estimated values of the matrix and clay properties are found in Table 4, the errors of which are calculated as the standard deviations of model parameters given in the last generation. At the end of the 2D GMI, the average of relative data distance decreases from 42.4 to 4.9%. The 2D distribution of relative data distance is plotted in Fig. 7a. The largest deviations are associated primarily to interpolation errors because of the different drilling depths. The quadratic mean of the estimation errors of volumetric parameters calculated by Eq. (11) is below 0.046 v/v (Fig. 7b), its average is 0.038 v/v, and the mean spread of the same quantities is $S = 0.49$. This result also shows stable and reliable inversion procedure. The CPU time of the 2D GMI using a quad-core processor workstation is 8 min 41 s.

Table 2 Zone parameters estimated by 1D genetic meta-algorithmic inversion in drill hole H4

Zone parameter	Search domain	Estimated value	Estimation error	Unit
GR_{cl}	8.0–12.0	8.86	0.01	kcpm
GR_{sd}	0–2.0	1.83	0.01	kcpm
ρ_{cl}	1.9–2.3	2.07	0.02	g/cm^3
ρ_{sd}	2.3–2.7	2.41	0.01	g/cm^3
$\Phi_{N,cl}$	0.2–0.5	0.43	0.05	v/v
R_{cl}	1.0–6.0	4.56	0.61	ohmm

GR_{cl} and GR_{sd} gamma-ray intensity of clay and sand, respectively; ρ_{cl} and ρ_{sd} density of clay and sand, respectively; $\Phi_{N,cl}$ neutron-porosity of water; R_{cl} resistivity of clay

JrnlID 10040_ArtID 1749_Proof# 1 - 23/02/2018

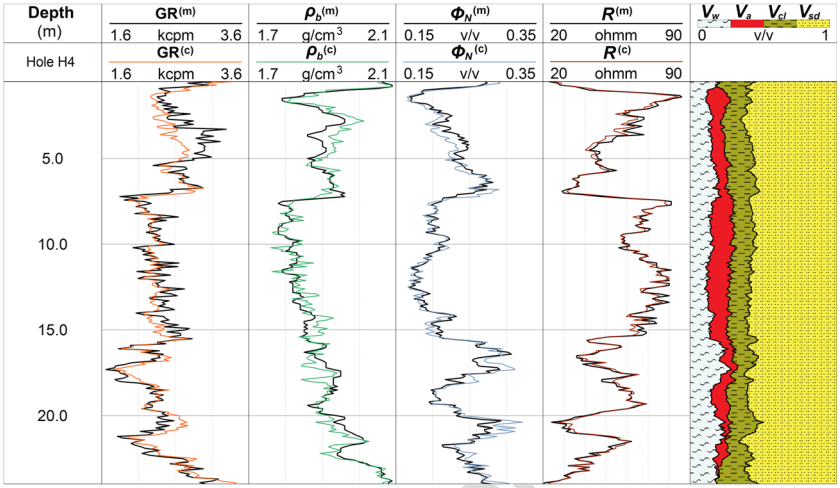

Fig. 4 Engineering geophysical sounding logs measured (m) and calculated (c) in drill hole H4 (tracks 1–4). GR is natural gamma-ray intensity, ρ_b is bulk density, Φ_N is neutron-porosity, R is resistivity. Petrophysical parameters estimated by genetic meta-algorithmic inversion are fractional volumes of water (V_w), air (V_a), clay (V_{cl}) and sand (V_{sd}) (track 5)

Discussion

The estimation of zone parameters has been one of the most challenging task in formation evaluation. It was previously studied that well-logging data are often insensitive to the variation of zone parameters such as tortuosity coefficient and saturation exponent, which also show high correlation to each other. During the inversion tests, high correlation was observed between the resistivity of water and density parameters. Because of the problem of ambiguity, the number of zone parameters estimated by inversion is restricted; however, these numerical tests show that the matrix and clay properties can be simultaneously estimated with the volumetric parameters in a stable meta-algorithmic inversion procedure. The use of global optimization is justified since there are several local maxima of the fitness function at selected values of zone parameters. The adaptability, numerical stability and initial-model independence of the FGA are proven to be higher than those of linearized inversion methods which usually become unstable in ambiguous situations and thus give a divergent solution. On the other hand, in the traditional local inversion approach, it is inevitably impossible to find a reliable solution for the volumetric and zone parameters together in one underdetermined inversion procedure (i.e. the estimation of six zone parameters and three volumetric properties— nine unknowns in total—from four data types).

The inversion results of the 1D and 2D GMIs show a close agreement. The estimation of volumetric parameters is made without regularization in both cases. The uncertainty of the fractional volumes is proportional to that of the input data (Fig. 5). In case of extremely noisy data sets, the application of robust weighting procedures is advisable, which effectively suppress the instrument noise, especially the influence of outliers (Steiner 1991). The most accurate parameters are the water volume and clay content, the standard deviation of which is obtained around 1 v/v. For a more reliable estimation of the matrix volume, it is

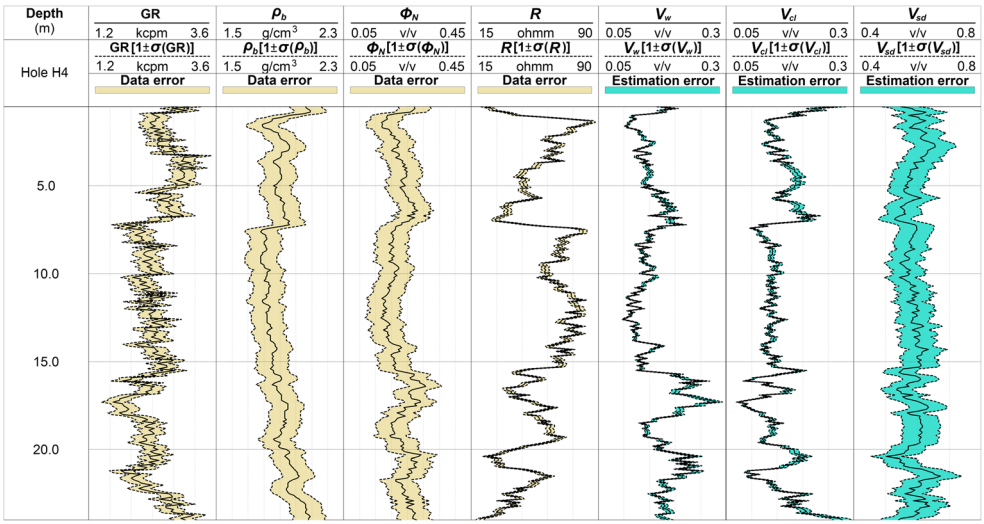

Fig. 5 Quality results of 1D genetic meta-algorithmic inversion in drill hole H4. Observed engineering geophysical sounding logs and their confidence intervals (tracks 1–4). Estimated petrophysical parameters and their estimation errors (tracks 5–7) (σ is standard deviation)

t3.1 **Table 3** Pearson's correlation matrix of engineering geophysical sounding logs measured in drill holes H4–H10

t3.2		GR	ρ_b	Φ_N	R
t3.3	GR	1	0.03	0.13	0.07
t3.4	ρ_b	0.03	1	−0.01	0.08
t3.5	Φ_N	0.13	−0.01	1	0.63
t3.6	R	0.07	0.08	0.63	1

GR gamma-ray intensity; ρ_b bulk density, Φ_N neutron-porosity; R resistivity logs

necessary to include new measurement types, the number of which is relatively limited in this stage of the EGS technology. New directions of tool development aim to measure such parameters that give information on the pore-space distribution and the type and level of soil contamination. The obtained values and estimation errors of zone parameters also confirm the feasibility of both the 1D and 2D GMI. In comparing the estimation results, very similar results are given for the resistivity of clay, which is a basic parameter of the dual water model. Much larger differences are obtained for the density of quartz and the neutron-porosity of clay, which will be studied in the future using different real data sets. The estimated values of zone parameters show a close agreement also with the results of Drahos (2005). The CPU time of the GMI is acceptable, which can be further improved by using more powerful workstations or by using the very fast simulated reannealing, as a quick global optimization method proposed by Ingber (1989) in the outer loop of the GMI.

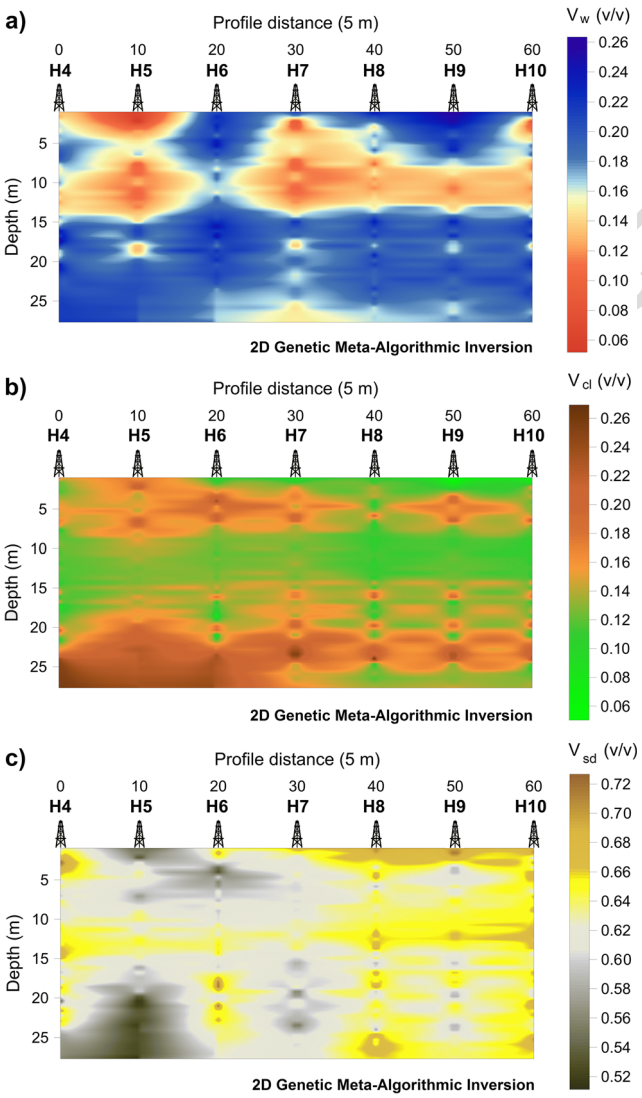

Fig. 6 Results of 2D genetic meta-algorithmic inversion in drill holes H4–H10: **a** water volume (V_w) estimated by 2D GMI, (**b** clay volume (V_{cl}) estimated by 2D GMI, **c** sand volume (V_{sd}) estimated by 2D GMI

Conclusions

A new inversion strategy assisted by evolutionary computation is proposed for evaluating the petrophysical properties of the shallow unsaturated zone. The novel approach named GMI combines the estimation of volumetric rock properties and zone parameters for improving the reliability of hydrogeological interpretation. By the inversion procedure, the physical properties of clay and matrix are estimated from an objective source. By using an adaptive genetic algorithm, the inversion process is practically initial-model independent, which does not require the initial value of the zone parameters. There is no need for laboratory data for the inversion process; they are only necessary for the initialization of FGA to set a possible range of zone parameters. Another advantage of the GMI is that with more accurate definition of zone parameters, the response functions of the forward problem are also automatically calibrated. The local inversion phase of the GMI provides both a fast solution and a high vertical resolution for the volumetric rock model, while the global optimization process provides a derivative-free solution for matrix and fluid parameters by effectively avoiding the ambiguity domains of correlated zone parameters. The latter computes the estimation error of the zone parameters from the entire population. By incorporating the data measured on core samples, the domains of zone parameters can be further constrained or the inversion results can be confirmed with the laboratory measurements.

The GMI hopefully becomes a useful tool not only in hydrogeophysical but also in oilfield well-logging applications where a large number of zone parameters affect the outcome of the evaluation of hydrocarbon reservoirs. Future research will extend the GMI to 3D applications, which will further improve the hydrogeological model of near-surface sediments. The optimization tool used in the inversion procedure will be made robust by using

t4.1 **Table 4** Zone parameters
t4.2 estimated by 2D genetic meta-
algorithmic inversion in drill
t4.3 holes H4–H10

Zone parameter	Search domain	Estimated value	Estimation error	Unit
GR_{cl}	8.0–12.0	8.02	0.17	kcpm
GR_{sd}	0–2.0	1.98	0.01	kcpm
ρ_{cl}	1.9–2.3	1.97	0.03	g/cm^3
ρ_{sd}	2.3–2.7	2.31	0.01	g/cm^3
$\Phi_{N,cl}$	0.2–0.5	0.33	0.01	v/v
R_{cl}	1.0–6.0	4.19	0.49	ohmm

GR_{cl} and GR_{sd} gamma-ray intensity of clay and sand, respectively; ρ_{cl} and ρ_{sd} density of clay and sand, respectively; $\Phi_{N,cl}$ neutron-porosity of water; R_{cl} resistivity of clay

statistically efficient weighting procedures, e.g. Cauchy or Steiner weights, to exclude the harmful effect of outliers and enable a more robust processing of non-Gaussian data. The output of inverse modeling can be used fruitfully in the factor analysis of the same EGS logs. The statistical factors can be applied for a more reliable estimation of water saturation and the derivation of hydraulic conductivity. In a reverse process, one may increase the overdetermination of the inverse problem by using the factor analysis-derived water saturation as a fixed parameter during the inversion process. The combination of the GMI with the aforementioned interval inversion method would make it possible to further improve the estimation accuracy of volumetric parameters and to determine the layer-thicknesses in an automated inversion procedure. The latter can also be extended to multi-borehole applications to locate the soil boundaries and to detect the lateral variation of the water table.

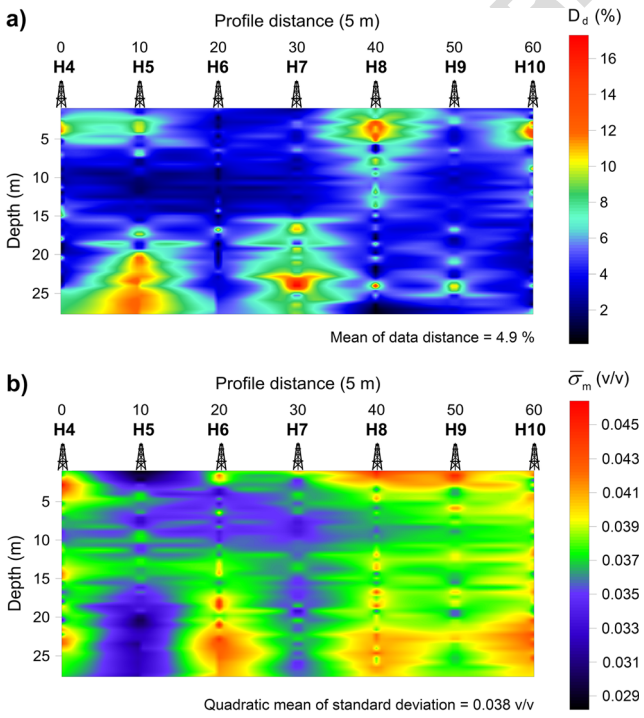

Fig. 7 Quality results of 2D genetic meta-algorithmic inversion in drill holes H4–H10: **a** 2D section of relative data distance (D_d) between observed and calculated engineering geophysical sounding logs, **b** 2D section of average estimation error of fractional volumes ($\bar{\sigma}_m$)

Acknowledgements 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. The author thanks Professor Mihály Dobróka, University of Miskolc, Associate Professor Dezső Drahos, Roland Eötvös University of Budapest, and János Stickel from Elgoscar-2000 Ltd. for the long-term scientific cooperation.

Funding Information The research was funded by the European Union and co-financed by the European Structural and Investment Funds.

References

Balázs L (2015) Inversion of well logging measurements with a constant interval parameter. Geosci Eng 4(6):93–104

Bijani R, Ponte Neto CF, Martins SS, Travassos JM (2012) 2-D tomography of first-arrivals using the genetic algorithm with elitism. SEG Technical Program Expanded Abstracts 2012:1–6. https://doi.org/10.1190/segam2012-0489.1

Butler JJJR (2005) Hydrogeological methods for estimation of spatial variations in hydraulic conductivity. In: Rubin Y, Hubbard SS (eds) Hydrogeophysics. Springer, Dordrecht, The Netherlands, pp 23–58

Cranganu C, Luchian H, Breaban ME (2015) Artificial intelligent approaches in petroleum geosciences. Springer, Heidelberg, Germany

De Witte L (1955) A study of electric log interpretation methods in shaly formations. Petroleum Trans AIME 204:103–110

Dobróka M, Szabó NP (2011) Interval inversion of well-logging data for objective determination of textural parameters. Acta Geophys 59(5): 907–934. https://doi.org/10.2478/s11600-011-0027-z

Dobróka M, Szabó NP (2012) Interval inversion of well-logging data for automatic determination of formation boundaries by using a float-encoded genetic algorithm. J Pet Sci Eng 86–87:144–152. https://doi.org/10.1016/j.petrol.2012.03.028

Dobróka M, Szabó NP, Tóth J, Vass P (2016) Interval inversion approach for an improved interpretation of well logs. Geophysics 81(2): D163–D175. https://doi.org/10.1190/GEO2015-0422.1

Drahos D (2005) Inversion of engineering geophysical penetration sounding logs measured along a profile. Acta Geodaet Geophys 40(2):193–202. https://doi.org/10.1556/AGeod.40.2005.2.6

Fejes I, Jósa E (1990) The engineering geophysical sounding method: principles, instrumentation, and computerised interpretation. In: Ward SH (ed) Geotechnical and environmental geophysics, environmental and groundwater, vol 2. SEG, Tulsa, OK, pp 321–331

Houck CR, Joines J, Kay M (1995) A genetic algorithm for function optimization: a MATLAB implementation. NCSU-IE technical report 95–09. North Carolina State University, Raleigh, NC, pp 1–14

Ingber AL (1989) Very fast simulated re-annealing. Math Comput Model 12(8):967–973. https://doi.org/10.1016/0895-7177(89)90202-1

Kirsch R (2006) Groundwater geophysics: a tool for hydrogeology. Springer, Heidelberg, Germany

Marquardt DW (1959) Solution of non-linear chemical engineering models. Chem Eng Prog 55(6):65–70

Menke W (1984) Geophysical data analysis: discrete inverse theory. Academic, San Diego

Michalewicz Z (1992) Genetic algorithms + data structures = evolution programs. Springer, Heidelberg, Germany

Narayan JP, Yadav L (2006) Application of adaptive processing technique for the inversion of open hole logs recorded in oil fields of Indian basins. In: 6th International Conference & Exposition on petroleum geophysics "Kolkata 2006". Society of Petroleum Geophysicists, Uttarakhand, India, pp 505–512

Razali NM, Geraghty J (2011) Genetic algorithm performance with different selection strategies in solving TSP. In: Proceedings of the World Congress on Engineering 2011, vol II, London, July 2011, pp 1–6

Schulmeister MK, Butler JJ, Healey JM, Zheng L, Wysocki DA, McCall GW (2003) Direct-push electrical conductivity logging for high-resolution hydrostratigraphic characterization. Ground Water Monit Remediation 23(2):52–62. https://doi.org/10.1111/j.1745-6592.2003.tb00683.x

Sen MK, Stoffa PL (2013) Global optimization methods in geophysical inversion. Cambridge University Press, Cambridge, UK

Shin YJ, Kim D (2011) Assessment of undrained shear strength based on cone penetration test (CPT) for clayey soils. KSCE J Civ Eng 15(7):1161–1166. https://doi.org/10.1007/s12205-011-0808-6

Steiner F (1991) The most frequent value: introduction to a modern conception of statistics. Academic, Budapest

Szabó NP (2012) Dry density derived by factor analysis of engineering geophysical sounding measurements. Acta Geodaet Geophys 47(2):161–171. https://doi.org/10.1556/AGeod.47.2012.2.5

Szabó NP (2015) Hydraulic conductivity explored by factor analysis of borehole geophysical data. Hydrogeol J 23(5):869–882. https://doi.org/10.1007/s10040-015-1235-4

Szabó NP, Dobróka M, Drahos D (2012) Factor analysis of engineering geophysical sounding data for water saturation estimation in shallow formations. Geophysics 77(3):WA35–WA44. https://doi.org/10.1190/geo2011-0265.1

Szabó NP, Dobróka M, Turai E, Szűcs P (2014) Factor analysis of borehole logs for evaluating formation shaliness: a hydrogeophysical application for groundwater studies. Hydrogeol J 22(3):511–526. https://doi.org/10.1007/s10040-013-1067-z

Szalai S, Lemperger I, Metwaly M, Kis A, Wesztergom V, Szokoli K, Novák A (2015) Increasing the effectiveness of electrical resistivity tomography using γ11n configurations. Geophys Prospect 63(2):508–524. https://doi.org/10.1111/1365-2478.12215

Vértesy L, Fancsik T, Fejes I, Gulyás Á, Hegedűs E, Kovács Acs, Kovács P, Kiss J, Madarasi A, Sőrés L, Szabó Z, Tóth Z (2004) Ground-based geophysical surveys at the Bátaapáti (Üveghuta) Site and in its vicinity. In: Annual Report of the Geological Institute of Hungary 2003, Budapest, pp 239–256

Walsh D, Turner P, Grunewald E, Zhang H, Butler JJ, Reboulet E, Knobbe S, Christy T, Lane JW, Johnson CD, Munday T, Fitzpatrick A (2013) A small-diameter NMR logging tool for groundwater investigations. Groundwater 51(6):914–926. https://doi.org/10.1111/gwat.12024

Zs N, Kanli AI, Stickel J, Tillmann A (2010) The use of non-conventional CPTe data in determination of 3-D electrical resistivity distribution. J Appl Geophys 70(3):255–265. https://doi.org/10.1016/j.jappgeo.2010.01.008