HYPERPARAMETER INVERSION OF WELL LOGS FOR SIMULTANEOUS ESTIMATION OF VOLUMETRIC AND ZONE PARAMETERS

Running head: Hyperparameter inversion of well logs

Norbert Péter Szabó, Institute of Exploration Geosciences, Department of Geophysics, University of Miskolc, H-3515, Miskolc-Egyetemváros, Hungary, norbert.szabo@uni-miskolc.hu

Original paper date of submission: 12 March 2024

ABSTRACT

We present a new alternative for the joint inversion of well logs to predict the volumetric and zone parameters in hydrocarbon reservoirs. Porosity, water saturation, shale content, kerogen and matrix volumes are simultaneously estimated with the tool response function constants with a hyperparameter estimation assisted inversion of the total and spectral natural gamma-ray intensity, neutron porosity and resistivity logs. We treat the zone parameters, i.e., the physical properties of rock matrix constituents, shale, kerogen, and pore-fluids, as well as some textural parameters, as hyperparameters and estimate them in a meta-heuristic inversion procedure for the entire processing interval. The selection of inversion unknowns is based on parameter sensitivity tests, which show the automated estimation of several zone parameters is favorable and their possible range can also be specified in advance. In the outer loop of the inversion procedure, we use a real-coded genetic algorithm for the prediction of zone parameters, while we update the volumetric parameters in the inner loop in addition to the fixed values of zone parameters estimated in the previous step. We apply a linearized inversion process in the inner loop, which allows for the quick prediction of volumetric parameters along with their estimation errors from point to point along a borehole. Derived parameters such as hydrocarbon saturation and total organic content show good

agreement with core laboratory data. The significance of the inversion method is in that zone parameters are extracted directly from wireline logs, which both improves the solution of the forward problem and reduces the cost of core sampling and laboratory measurements. In a field study, we demonstrate the feasibility of the inversion method using real well logs collected from a Miocene tight gas formation situated in the Derecske Trough, Pannonian Basin, East Hungary.

INTRODUCTION

Open-hole wireline logs provide high resolution quasi in-situ information about the lithological and petrophysical characteristics of hydrocarbon reservoirs (Serra, 1984; Asquith and Krygowski, 2006). For reliable formation evaluation and reserve calculation, we can estimate the volumetric quantities such as porosity, shale volume, water saturation, matrix volumes in a quality-checked inversion procedure (Schlumberger, 1989). In conventional (shaly sand) reservoirs, we normally invert a suite of well logs sensitive to lithology, porosity, and fluid saturation jointly in a linearized inversion procedure (Menke, 1984; Tarantola, 1987). In this approach, the weighted least squares method is used to minimize the misfit between the observed data and the theoretical

ones calculated by the estimated petrophysical model and properly chosen tool response functions (Ball et al., 1987; Alberty and Hashmy, 1984). Inversion procedures used generally in industrial practice invert the local dataset at a given depth to estimate the petrophysical parameters at the same depth (Baker Atlas, 1996). By solving a set of local inverse problems along a borehole one receives a quick solution for the petrophysical parameters and their uncertainty. As a drawback, however, since this inversion method calculates the model parameters only from a dataset measured at a given depth, it is therefore typically marginally overdetermined (i.e., we have barely more data than unknowns), and quite a noise-sensitive inversion procedure (Dobróka et al., 2016). A consequence of this is that the number of petrophysical parameters to be determined by local inversion is rather limited and the inversion procedure must use arbitrarily chosen (fixed value) response function constants, which may cause uncontrolled modeling error in the inversion results.

Well logging inverse problems usually have two types of petrophysical unknowns, the volumetric and the zone parameters. The volumetric quantities change rapidly with depth, while the zone parameters are unvarying (or just slowly varying) in a longer processing interval (e.g., in the reservoir zone). Zone parameters representing the physical properties of rock constituents and fluids (e.g., density of hydrocarbon, neutron porosity of shale, resistivity of pore-water or Archie's

exponents etc.) can be found in the tool response functions, but there are some influential quantities that are not included explicitly in them (e.g., temperature, pressure). If we treated all petrophysical variables included in the response functions as unknown, we would end up with a highly underdetermined inverse problem, the solution of which is not possible within the framework of local inversion. In this case, to avoid the problem of ambiguity, one treats the zone parameters as constant, the values of which are fixed using the available well site information, laboratory data or literature resources. In unconventional reservoirs, where the number of matrix and fluid components and zone parameters included in response functions is higher than in shaly sand formations, the application of local inversion methods apart from the determination of some basic parameters is not suitable for estimating additional parameters. Since the number of local volume characteristics and zone parameters is much larger than that of the observed data, one would have to solve an underdetermined inverse problem at each depth, which may lead to an unstable inversion procedure and ambiguous solution.

In practice, no significant progress has been made in the automated estimation of zone parameters. The basic reason for this is not just the marginal overdetermination of the popular local inversion procedure, but also the fact that the parameter sensitivity and correlation relations of the

zone properties and their consequences were not really investigated either. Narayan and Yadav (2006) apply the least squares method to give an estimate to a limited number of matrix parameters, where they compare only the mean values of the measured wireline logs to the theoretical logs calculated in a particular depth interval. In their approach, they keep the zone parameters of sand and pore-water fixed during the inversion process, while they update those of other minerals (if present) with a maximum 20% of their initial value preliminary set by cross-plot techniques. Petrophysical software systems used by leading oil companies normally apply local inversion workflows that also treat the zone parameters as constant, which must be estimated before inversion. Other approaches also exist to determine a limited number of zone parameters, e.g., Sousa et al. (2020) applied a regression technique to estimate the interval mineral and fluid densities using well logs.

Heidari et al. (2012) suggested a nonlinear inversion approach to invert well logging data collected from more depths simultaneously to improve the efficiency of well log interpretation in thinly bedded formations. The same inversion methodology was successfully applied to carbonate formations (Heidari et al., 2013). The interval inversion method suggested by Dobróka and Szabó (2011) allows for the automated estimation of suitably chosen zone parameters. A favorably high

overdetermination (data-to-unknowns) ratio can be achieved by discretizing the depth-dependent model parameters with a series expansion technique. We directly estimate the expansion coefficients as unknowns of the inverse problem, which replaces the depth variation of petrophysical properties. The interval inversion approach greatly improves the accuracy of estimation, which is directly proportional to the degree of overdetermination (Dobróka et al., 2016). In practice, the number of series expansion coefficients can be selected significantly smaller than that of the well logging data measured on the processing interval, thus additional parameters can be involved in the inversion procedure such as the zone parameters or the layer thicknesses (Dobróka and Szabó, 2012). However, to achieve an appropriate vertical resolution of model parameters, the number of expansion coefficients must not be too small, thus the number of involved zone parameters cannot be increased too much. Increasing the number of series expansion coefficients that are estimated by inversion can quickly reduce the accuracy. When estimating the zone parameters together with the volume characteristic parameters, one must maintain an optimal rate of overdetermination. Not to mention that increasing the number of unknowns within the same inversion procedure increases the risk of higher intercorrelations and the problem of ambiguity as it was shown for constant interval parameters by Balázs (2015). Szabó and Dobróka (2020)

extended the application of the interval inversion method to shale gas reservoirs, where a hybrid application of global and linear optimization methods provided reliable estimates of porosity, water saturation, mineral volumes, and total organic carbon content.

Several studies show the importance of evaluating multimineral formations, e.g., shale gas reservoirs (Liu et al., 2022). In this study, we estimate the volumetric and zone parameters of tight gas formations in a joint inversion process. Instead of using interval inversion, we apply a different approach based on hyperparameter estimation. Analogous to their machine learning use (Feurer and Hutter, 2019), we define the hyperparameters unknown variables with the automated adjustment of which we can control the inversion process and improve its result. We define the zone parameters of the theoretical tool response equations as hyperparameters which are estimated in the outer loop of a two-step embedded inversion algorithm. In the inner loop of the inversion procedure the volumetric parameters are estimated by traditional linear inversion. In both steps, the criterion for improving the inversion parameters is the appropriate matching of the measured and calculated wireline logs. Hyperparameter inversion has been previously applied to estimate the a priori data errors in a Bayesian inversion scheme in seismic tomography (Bodin et al., 2012). Xiao et al. (2021) used the Markov chains to estimate the mean, variance, and scale of the data

covariance function treated as hyperparameters in the processing of hydraulic head data. Tilahun and Korus (2023) tuned the hyperparameters to better estimate hydraulic conductivity in high resolution 3D groundwater modeling. Zhang et al. (2022) tested how the hyperparameters of deep learning neural networks affect the seismic impedance results, which draws attention to the advantages of optimizing the hyperparameters. Kuwatani et al. (2022) estimated the smoothness and precision hyperparameters to improve the seismic image reconstruction and deblurring. Previously, we estimated the matrix and clay parameters simultaneously with the water, shale, and sand content of shallow unsaturated sediments by the inversion of direct push logs (Szabó, 2018). We investigated the subsoils by the genetic meta-algorithmic inversion (GMI) method used for processing data acquired from single and multiple boreholes, respectively. The inversion methodology allows us to reliably determine the basic soil parameters and quantify their estimation accuracy. During the 2D inversion of engineering geophysical sounding data, we predicted the volumetric parameters point-by-point along the drill-holes, while we fitted thousands of observed and calculated data to estimate the zone parameters assumed constant in the investigation area. The hyperparameter estimation based GMI procedure proved to be stable and found a reliable solution after some iterations. We validated the inversion results by exploratory factor analysis,

which provides reliable information especially for water saturation estimation (Szabó et al., 2012). In this study, we assume the hyperparameter inversion procedure can also be applied to evaluate tight gas formations and thus its application can be extended to oilfield wireline logs. The petrophysical model of unconventional reservoirs has a more complex structure than that of near-surface structures, therefore, it may be a serious challenge for the inversion method to predict many more volume characteristics and zone parameters. Meeting this challenge, we demonstrate the feasibility of hyperparameter inversion using real data, which is also the first application to evaluate the Hungarian unconventional hydrocarbon formations by this inversion approach.

METHODS

Forward problem

For computing wireline logs in tight gas reservoirs, we establish a petrophysical model that can be connected to the data by proper tool response equations. We assume the unit volume of rock is composed of pore-space occupied by brine and gas, and solid particles forming the rock matrix including silica, shale, and kerogen. Accordingly, we select the porosity (Φ), shale content (V_{sh}), kerogen volume (V_k), sand volume (V_{sd}), water saturation defined in the invaded (S_{x0}) and

virgin zone (S_w) as volumetric parameters of the petrophysical model. The natural gamma-ray intensity (GR in API), potassium concentration (K in %), thorium concentration (Th in ppm), uranium concentration (U in ppm), neutron-porosity (Φ_N in p.u.), and deep resistivity (R_d in Ωm) are available for the evaluation of Hungarian tight gas formations.

We apply linear response functions to simulate the nuclear logs (Alberty and Hashmy, 1984). The theoretical data are calculated as the weighted average of shale content, sand and kerogen volume, and porosity (if included in the equation)

$$\mathrm{GR} = V_{\mathrm{sh}}\mathrm{GR}_{\mathrm{sh}} + V_{sd}\mathrm{GR}_{sd} + V_k\mathrm{GR}_k, \tag{1}$$

$$K = V_{\mathrm{sh}}K_{s\,h} + V_{sd}K_{sd} + V_kK_k, \tag{2}$$

$$U = V_{\mathrm{sh}}U_{s\,h} + V_{sd}U_{sd} + V_kU_k, \tag{3}$$

$$Th = V_{\mathrm{sh}}Th_{s\,h} + V_{sd}Th_{sd} + V_kTh_k, \tag{4}$$

$$\Phi_N = \Phi\left[S_{\mathrm{x0}}\,\Phi_{\mathrm{N,mf}} + (1 - S_{\mathrm{x0}})\,\Phi_{\mathrm{N,h}}\right] + V_{\mathrm{sh}}\,\Phi_{\mathrm{N,\,sh}} + V_{sd}\,\Phi_{\mathrm{N,\,sd}} + V_k\,\Phi_{\mathrm{N,\,k}}, \tag{5}$$

where the subindices *mf*, *h*, *sh*, *sd*, *k* denote the mud filtrate, hydrocarbon (gas), shale, sand, and kerogen, respectively. All functional constants indicated by these indices which do not represent

volume characteristics are considered as zone parameters. The full list of zone parameters is given in Table 1. Water saturation is usually calculated by non-linear tool response functions. In tight gas formations, Archie's equation must be corrected by the total organic content (*TOC*) related gas saturation (Xu et al., 2017). Thus, we apply a modified Archie's formula, which takes the kerogen content into account in calculating the true (or corrected deep) resistivity ($R_t \approx R_d$)

$$R_d = \frac{\mathrm{a}R_w}{\Phi^m S_w^n} - R_{sh}(V_{sh} - V_k)^2 + V_k^2 K_{rf}, \tag{6}$$

where K_{rf} is the kerogen (resistivity) factor that can be specified from knowledge of the local *TOC* vs. R_t relationship (Kadkhodaie and Rezaee, 2016), and the textural parameters *a*, *m*, *n* are the dimensionless Archie's constants (Archie, 1942). Equation 6 is valid in the domain of 0-2000 ohmm resistivity and the range of 0-20 wt% TOC ($1999 < K_{rf} < 33317$). By combining equations 5 and 6 shale-free porosity can be estimated by inversion in formations with high shale volume. We can gather the theoretical data into a column vector

$$\mathbf{d}^{(th)} = \left[GR, K, U, Th, \Phi_N, R_d\right]^{\mathrm{T}}. \tag{7}$$

In forward modeling, we calculate the data of the above well log types at successive depth points using equations 1-6, where the input values of volumetric and zone parameters are estimated with a preliminary inversion.

Two-step inversion procedure

Genetic inversion (Phase I)

In the global optimization step of the inversion procedure, the entire well logging dataset is jointly inverted for the purpose of determining the zone parameters considered constant at an arbitrarily chosen depth interval. The zone parameters can be freely selected as inversion unknowns. The parameter selection process can be based on synthetic modeling tests, quality check of inversion tests, and parameter sensitivity calculations. In this study, we predict the following vector of zone parameters

$$\mathbf{c} = \left[GR_{sh}, GR_k, K_{sh}, U_{sh}, U_k, Th_{sh}, \Phi_{N,mf}, \Phi_{N,sh}, R_w, m, K_{rf}\right]^{\mathrm{T}}, \tag{8}$$

where T is the symbol of transpose. We use the real-coded genetic algorithm for the estimation of zone parameters, which is a highly adaptive global optimization method giving a derivative-free

(Sen and Stoffa, 2013) and practically initial-model independent solution (Szabó and Dobróka, 2019).

Genetic algorithms as an evolutionary computation method use the analogy to natural selection of living organisms (Holland, 1975). According to Darwin's theory, the fittest individuals survive and reproduce during an evolution process, while others die off quickly in the population. In our genetic algorithm, we improve a set of zone parameter vectors considered as individuals of the artificial population by maximizing the fit between the theoretical and observed data. The fitness value of the p-th zone parameter vector $\mathbf{c}_p$ (p=1,2,…,P, where P is the population size) is calculated as

$$F(\mathbf{c}_p) = -\left[(NK)^{-1}\sum_{i=1}^{N}\sum_{j=1}^{K^*}\left(\frac{d_{ij}^{(m)} - g_j(\mathbf{m}_i,\mathbf{c}_p)}{\sigma_j^{(m)}}\right)^2\right]^{\frac{1}{2}}, \tag{9}$$

where $d_{ij}^{(m)}$ is the data measured in the i-th depth with the j-th tool, $\sigma_j^{(m)}$ is the standard deviation of observed data representing the accuracy of measurement, g_j is the response function of the j-th well log, $\mathbf{m}_i$ is the vector of volumetric parameters estimated in the i-th depth we do not let vary in the outer loop of the inversion algorithm (N is the number of depth points over the processing interval and K^* is the number of applied logging tools). (In the absence of information on data

accuracies one can substitute the variance with the measured data itself.) The goal of the evolutionary search process is to increase the average fitness of the population in successive generations and finally to find the global maximum of the fitness function $F(\mathbf{c})$. However, the fitness function may have many local maximum locations in the domain of the zone parameters. Let us generate 100 random populations of zone parameters to analyze the landscape of the fitness function. We change all zone parameters (23 parameters) in the range shown in the last column of Table 1. One can see in Figure 1, in some bivariate subdomains of the independent variables, numerous candidate (equivalent) solutions appear with maximum fitness that can be far from the absolute optimum. We evaluate the red maxima as equally good places, which include the best fit of the measured and calculated well logs. The areas characterized by the largest data distance (fitness minima) are forbidden ranges. The areas marked in dark red contain the most probable solutions, the global optimization must find the best solution (absolute maximum) in these ranges. The above shows that it is advisable to use the genetic algorithm during the hyperparameter inversion.

In the genetic inversion step (Phase I), we apply a differential evolution-based float-encoded genetic algorithm (Michalewicz, 1992). We simultaneously improve a set of zone

parameter vectors considering all of them as chromosome made up of floating-point value genes to avoid the local maximum locations in the search space. In the outer loop of the inversion algorithm shown in Figure 2, the volume characteristics are treated as fixed constants during the iterative process, in which the floating-point genetic operations, i.e., selection, crossover, and mutation are repeatedly applied on the individuals in consecutive generations (Houck et al., 1995). First, we pick the fittest individuals for mating using a tournament selection operator. We randomly choose a number of individuals from the population (allowing re-selection) and we copy the fittest into a temporary population. We repeat this process till P number of individuals have been selected. The number of tournaments as a tunable control parameter influences the development of convergence of the genetic search. If we set the tournament number to be large, individuals with relatively small fitness have a smaller chance to be selected for the next generation. Next, we exchange some of the zone parameters (as genes of the chromosomes) between a pair of individuals ($\mathbf{c}_1$and $\mathbf{c}_2$) that have undergone the selection process. In this study, we use the heuristic crossover that extrapolates two individuals as follows

$$\left.\begin{aligned}\hat{\mathbf{c}}_1 &= \mathbf{c}_1 + \alpha(\mathbf{c}_1 - \mathbf{c}_2)\\ \hat{\mathbf{c}}_2 &= \mathbf{c}_1\end{aligned}\right\}, \tag{10}$$

where $\hat{\mathbf{c}}_1$ and $\hat{\mathbf{c}}_2$ are the two offspring, α is a random number uniformly distributed in the interval of 0 and 1. The feasibility condition for using the crossover operator is if the fitness of $\mathbf{c}_1$ is higher than that of $\mathbf{c}_2$. If any parameter of vector $\hat{\mathbf{c}}_1$ is out of bounds, we generate a new value of α and recalculate equation 10. After a certain number of retries, we let the new values of zone parameters equal the old ones. We use the uniform mutation as the third genetic operation, in which we exchange the l_0-th gene with a real number (β) randomly generated from the possible range of the actual zone parameters

$$\tilde{\mathbf{c}}_1 = \begin{cases} \beta, & \text{if } l = l_0 \\ \hat{c}_{1,l}, & \text{if } l \neq l_0 \end{cases}, \tag{11}$$

where $\tilde{\mathbf{c}}_1$ is the mutated individual (l=1,2,…,L). After applying the three genetic operators, we use elitism-based reproduction, which preserves the fittest individual of the previous generation and replaces the weakest with it in the next generation (Bijani et al., 2012). We apply the three genetic operations in successive generations until a proper fit is achieved (Figure 2). We calculate the data distance as a measure of fit between the observed and calculated wireline logs in percent by D_d=−100F. After reaching the maximal number of generations, we stop the inversion process and regard the zone parameter vector with maximum fitness as the solution of the global optimization

problem. We compute the estimation error of the zone parameter as the standard deviation of the values of relevant genes of different individuals obtained in the last generation.

Linear inversion (Phase II)

We estimate the volumetric parameters in an inner iterative loop embedded inside the genetic inversion procedure. In this phase, the zone parameters for which we obtain preliminary estimates in the outer loop of the inversion process (Phase I) are fixed as constant (Figure 2). In the linear optimization step of the inversion procedure (Phase II), the fractional volumes of the rock constituents are estimated by local inversion where the theoretical data is fitted to the observed data in each depth, separately. Below we specify the model parameters to be estimated

$$\mathbf{m} = \left[\Phi, S_{x0}, S_w, V_{sh}, V_k, V_{sd}\right]^{\mathrm{T}}. \tag{12}$$

One can calculate the well log data included in equation 7 to the given depth using $\mathbf{d}^{(th)}=\mathbf{g}(\mathbf{m},\mathbf{c})$, where the vector $\mathbf{m}$ is given in equation 12 and the vector $\mathbf{c}$ is constant. We write the linear approximation of this relation as $\mathbf{d}^{(th)}=\mathbf{G}\mathbf{m}$, where $\mathbf{G}$ denotes the Jacobi matrix including the numerical derivatives of well log data with respect to the model parameters. We derive the deviation vector $\mathbf{e}$ by the difference between the observed data and the calculated ones, which is

normalized by the data variances that measure the instrument uncertainty. Based on this, we formulate the objective function to be minimized as

$$E = \|\mathbf{e}\|_2^2 + \varepsilon^2 \|\mathbf{m}\|_2^2, \tag{13}$$

where ε^2 is a positive number chosen as the ratio of data and model variances (Menke, 2022) or empirically set based on preliminary inversion runs. The quantity ε acts a regularization parameter used for numerically stabilizing the inversion procedure. We estimate the model correction vector of the volumetric parameters using the algorithm of Marquardt (1959)

$$\delta\mathbf{m} = \left(\mathbf{G}^{\mathrm{T}}\mathbf{G} + \varepsilon^2\mathbf{I}\right)^{-1} + \mathbf{G}^{\mathrm{T}}\delta\mathbf{d}, \tag{14}$$

by which we update the volumetric parameters estimated in the previous iteration ($\mathbf{m}_0$) using $\mathbf{m}=\mathbf{m}_0+\delta\mathbf{m}$ (Figure 2). In equation 14, vector $\delta\mathbf{d}$ denotes the difference between the measured ($\mathbf{d}^{(m)}$) and calculated data vector $\mathbf{d}_0^{(th)} = \mathbf{g}(\mathbf{m}_0,\mathbf{c})$ and $\mathbf{I}$ is the unity matrix.

Our workflow also allows us to quality check the inversion results. In the linearized inversion phase, we can quantify the uncertainty of the estimated volumetric parameters in each iteration using the Jacobi matrix. According to the inverse theory of Menke (1984), the covariance

matrix of the model parameters estimated by a linear inversion method can be related to the data covariance matrix including the data variances on its main diagonal

$$\mathrm{cov}(\mathbf{m}) = \mathbf{G}^{-g}\mathrm{cov}\left(\mathbf{d}^{(m)}\right)\left(\mathbf{G}^{-g}\right)^{\mathrm{T}}, \tag{15}$$

where $\mathbf{G}^{-g}$ is the generalized inverse of the damped least squares method. The variance of input data including the instrumental error and environmental effects can be given by tool calibration tests and user manuals. Horváth (1973) discusses the possible sources of interpretation errors and gives an estimate of the accuracy of data measured by different probe types. Balázs (2021a) studies the effect of correlation between the corrected data on the inversion results. The estimation accuracy of volumetric parameters is calculated as the square root of variances obtained in the main diagonal of the model covariance matrix derived from equation 15. The reliability of inversion results can be quantified via the Pearson's correlation matrix of estimated model parameters. As is well known in inversion practice, we consider the solution unique when the estimated model parameters correlate weakly. In case of high correlation, the model parameters cannot be estimated individually but just a combination of them. In these cases, we can encounter with the problem of ambiguity, which prevents the reliable estimation of the geophysical model.

At the end of the linear inversion procedure (Phase II), we fix the newly estimated volumetric parameters and return to the genetic inversion step (Phase I), where we recalculate the zone parameters. We repeat the two inversion phases until the maximum number of iterations is reached (Figure 2).

INVERSION RESULTS

Parameter sensitivity test

Parameter sensitivity informs about the extent of influence on data, which is exerted by a model parameter. Dobróka (1988) introduced parameter sensitivity functions for studying the absorption and dispersion characteristics of guided waves. Gyulai (1995) applied the same functions for geoelectric parameters of dipping bed structures in analytic forward modeling. Szalai et al. (2002) proved based on parameter sensitivity experiments that null array data are spatially more variable than traditional geoelectric arrays when investigating near-surface structures. Wang et al. (2022) studied the influence of neutron capture cross section of rock constituents to water saturation in cased hole well log analysis. In inverse modeling, we prefer to determine the parameters of relatively high sensitivity, which influence the objective function to a sufficient

extent during the search. Ingber (1989) improved the Fast Simulated Annealing method by modifying the cooling schedule of each model parameter according to their sensitivity values considering that how they affect the value of the energy function. Model parameters with low sensitivity can delay the development of convergence, and since they cannot be changed to an adequate extent, their estimation can lead to an ambiguous solution. The selection of the inversion unknowns based on their sensitivity is quite important, accordingly we select those with high sensitivity, while the rest are determined outside of the inversion procedure.

We apply the following parameter sensitivity function for measuring the variability of wireline logging data influenced by a zone parameter

$$\Psi_{jq}(c_q) = \frac{\partial d_j \quad c_q}{\partial c_q g_j(\mathbf{m},c_q)}, \tag{16}$$

where Ψ_{jq} represents the parameter sensitivity of the *j*-th data type as a function of the *q*-th zone parameter. When the model parameters are evenly spaced, we can approximate the partial derivative by the quotient of the finite differences (Menke, 2022). We compute the numerical derivatives of the theoretical data with respect to the zone parameters using the response functions 1-6. As a result, quantity Ψ is dimensionless and scaled between the interval of 0 and 1. We

generate the parameter sensitivity functions in tight gas formations to determine the individual extent of influence on well logging data, which is exerted by each zone parameter. When choosing the range of zone parameters, we take the recommendation of Baker Atlas (1996) into account and choose a wide interval according to the reservoir type.

We calculate the parameter sensitivity functions for the zone parameters listed in Table 1. We set the reference values of zone parameters such as the Archie's constants as default values (Archie 1942). All natural radioactivity levels of sand are equal to zero assuming it to be a non-radioactive rock. We give the neutron porosity of sand as matrix corrected value of zero porosity quartz sandstone. One can estimate the kerogen factor using the R_t vs. TOC chart proposed by Kadkhodaie and Rezaee (2016). We set the rest of zone parameters close to their values used in the field study of Szabó et al. (2022). In equation 16, we always change only one zone parameter, while the others are considered constant. The basic volumetric parameters are fixed as follows Φ=0.08 (v/v), V_{sh}=0.60 (v/v), V_k=0.01 (v/v), V_{sd}=0.30 (v/v), S_w=0.60 (v/v), S_{x0}=0.88 (v/v). In Figures 3 and 4, semi-log plots show the sensitivity functions of the total and spectral gamma-ray intensity, neutron porosity and deep resistivity logs. Each plot includes four curves, which represent how the sensitivity changes at four different values of the relevant volumetric quantity.

By analyzing these plots, one can see very sensitive parameters like the shale properties (e.g., GR_{sh}, Th_{sh}), and parameters with low sensitivity, which are not favorable to estimate by an inversion procedure (e.g., sand and kerogen properties). Szabó and Dobróka (2020) showed the kerogen volume can be estimated with great accuracy by inversion, however, this is not valid for all kerogen related zone parameters (e.g., Th_k, $\Phi_{N,k}$). In water bearing formations, the sensitivity function for $\Phi_{N,h}$ and n are both zero, which is the main reason for not choosing them as unknowns. The latter is easy to see by analyzing the modified Archie's model (equation 6), where the sensitivity remains zero for any n, because S_w equals to 1. For the sake of simplicity, we illustrate the absolute value of the sensitivity for parameters $\Phi_{N,sd}$ and R_{sh}. In this case, only the amount of change matters, which is not very big, the negative sign shows inverse relation between the zone parameter and the data. We derive the sensitivities from the curves by substituting the reference values of zone parameters indicated in the third column of Table 1. We select a zone parameter as unknown if its sensitivity reaches 20 % of the maximal sensitivity ($\Psi \geq 0.2$) in the near vicinity of the reference value of the zone parameter. The last column of Table 1 shows whether the zone parameter is selected or fixed during the inversion. Based on the parameter sensitivity experiments, we collect

the most sensitive parameters into vector **c** (equation 15), which represent the unknows in the genetic inversion phase of our workflow.

Inversion of field data

We apply the hyperparameter estimation-based inversion method to the interpretation of well logs observed in Berettyóújfalu-4 (Be-4) well drilled in the Derecske basin in East Hungary. The Neogene basin and the surrounding area is an important gas accumulation region in Hungary (Haás, 2013). The investigated tight formations of Mid-Miocene age settled under a thick sedimentary sequence at a depth greater than 3000 m in a complex geological environment (Holditch, 2013). The Berettyóújfalu tight gas occurrence is in the southwestern part of the NE-SW-striking Derecske trench, the depth of which is highly variable (approx. 3800-4600 m). The deep basin is filled by a Miocene clastic and volcano-sediment complex, which settles directly on the Paleozoic bedrock, with a minimum thickness of 300–700 m. The Miocene interval includes the reservoir and source rock layered into each other between 3471 and 3616 m in the Be-4 well. The organic matter content is adequate everywhere based on the measured TOC values, in the lower part it is classified as good or excellent, also suitable for oil and gas generation III type

kerogen. Due to the rapid subsidence and thick sedimentary complex covering the hydrocarbon play in the basin, the young source rock is in matured position which is indicated by the values of vitrinite reflectance of 0.91-0.96 in the zone of interest. The tight gas reservoirs are mainly composed of sandstones and siltstones with low porosity and permeability. Szabó et al. (2022) published a detailed geological description of the hydrocarbon field, and the exploration activity including seismic, well logging and laboratory measurements. The same study includes the estimation results of interval inversion for the basic volumetric parameters, which we use for constructing the initial model.

We process the nuclear and resistivity logs given in equation 7. In the depth interval of 3550 and 3617 m, we have 5292 data (N=882, K^*=6). We aim to estimate the vector of the zone parameters (L=11) given in equation 8 in Phase I. The rest of the zone parameters, which are fixed as constant, are given in Table 1 (see the reference values). In Phase II, we predict the volumetric parameters included in equation 12. It is possible to increase the overdetermination of the local inverse problem using the material balance equation, since we can compute the sand volume also deterministically using $V_{sd}=1-\Phi-V_{sh}-V_k$. The water saturation of the flushed and virgin zone is generally highly correlated to each other or other volumetric and zone parameters burdening the

result of the inversion with a relatively large error (Balázs, 2021b; Dobróka et al., 2016). The estimation error of water saturation cannot be significantly reduced even by increasing the overdetermination of the inverse problem (Abordán and Szabó, 2020), therefore, it is reasonable to estimate one of the water saturations outside of inversion. We can further increase the data-to-unknowns ratio by calculating the flushed zone saturation from the empirical relation $S_{x0} = S_w^{\kappa}$ ($0.2 \leq \kappa \leq 0.5$) suggested by the Hungarian oil industry experience. We set the exponent κ to 0.25 based on preliminary inversion runs.

In the genetic inversion step (Phase I), we select a wide search domain of the zone parameters (Table 2) and we generate an initial population of 200 individuals (2200 genes). The fitness values of these individuals are plotted in Figure 5a, where the data distance of the individuals (D_d) vary between 24% and 547%. This wide search domain assures to find the global optimum, at the same time, is not too large to achieve fast convergence. We control the genetic search by setting the number of individuals participating in each tournament to 50, the number of crossover retries to 100 and the mutation rate to 0.05. We apply the elitism-based reproduction mechanism, which replaces the worst individual of the current generation with the fittest one of the previous generation. We refine the zone parameters over 20000 generations beside fixed values

of volumetric quantities to find the global optimum. In our inversion workflow, after completing the genetic inversion step (Phase I), the local inversion process begins (Phase II) that runs over 100 iterations in each depth, separately, then the genetic inversion process (Phase I) follows again that runs over 20000 generations and so on. In Phase II, we estimate the volumetric properties while the zone parameters obtained as a result of Phase I are kept constant. The starting values of volumetric parameters are Φ_0=0.08 (v/v), $V_{sh,0}$=0.60 (v/v), $V_{k,0}$=0.02 (v/v), $S_{w,0}$=0.60 (v/v). We set the initial value of the damping factor ε to 1000, and gradually decrease it by 90% of its actual value in each iteration to maintain it near zero till the end of the local inversion process. We repeat the two inversion steps (Phases I and II) 10 times (Maximum iteration step in Figure 2). One can see the gradual increase of the maximal fitness of individuals in the first genetic inversion step in Figure 5b. Both the maximal and minimal fitness of the optimal population found in Phase I improve continuously over the 10 iterations of the two-step inversion procedure (Figure 5c). In the zeroth iteration, we see the fitness value of the initial models, which is calculated by substituting $\mathbf{m}_i^{(0)}$ and $\mathbf{c}_p^{(0)}$ into equation 9. We can see rapid convergence toward the optimum in Figure 5d, which proves the stability of the two-step inversion procedure. One can see from the shape of the curve that the inversion method already gives a suitable approximation in the first iteration, after

that it only slightly refines the zone parameters. The GMI inversion method applied in shallow unconsolidated sediments is characterized by similarly fast convergence (Szabó, 2018). We show the change of zone parameters in the last genetic inversion phase (10^{th} iteration of the two-step inversion procedure) in Figure 6a. The accepted values of the zone parameters during the inversion procedure are shown in Figures 6b-d. At the end of the last iteration, we accept the newly estimated volumetric and the zone parameters as solution of the inverse problem. The relative data distance is D_d=4.39 %, which indicates a proper fit between the measured and calculated data.

We show the inversion result in Figure 7, where one can find the measured (black solid line) and calculated (dashed line with different colors) well logs in the first six tracks. The last three tracks illustrate the estimated volumetric parameters. The movable gas saturation is calculated as the difference between the S_{x0} and S_w curves, while the irreducible gas saturation is computed as $1-S_{x0}$ (track 7). The average porosity of the gas bearing formation is 8.2 %, while the mean of the shale content is 45 % (track 8). The TOC given in weight percent is derived from the kerogen volume using the formula $TOC=(V_k\rho_k)/(K_c\rho_b)$, where Kc is the kerogen conversion factor chosen as 0.95, ρ_k is the density of kerogen selected as 1.8 gcm^{-3}, and ρ_b is the gamma-gamma log derived bulk density (Tissot and Welte, 1978). The $TOC^{(inv)}$ log estimated by inversion and

$TOC^{(core)}$ data measured in the laboratory are in good agreement (track 9). We give the estimated zone parameters in the fourth column of Table 2.

We check the quality of inversion results using the prior knowledge of the data uncertainties. We select the standard deviation of the observed wireline logging data as σ_{GR}=0.12, σ_K=σ_K=σ_{Th}=0.10, $\sigma_{\Phi N}$=0.13, σ_{Rd}=0.14 (Dobróka et al., 2016). The error range of the observed logs are represented with shaded area on tracks 1-6 in Figure 8. Substituting the data variances into equation 15, we derive the variances of volumetric parameters estimated at the end of the last linear inversion step. The average estimation error of volumetric quantities computed along the entire processing interval is: $\overline{\sigma_{\Phi}}$= 2.9·10^{-3}, $\overline{\sigma_{S_w}}$=5.6·10^{-2}, $\overline{\sigma_{V_{sh}}}$=1.6·10^{-2}, $\overline{\sigma_{V_k}}$=3.6·10^{-4}. We show the estimated volumetric parameters and their 95 % confidence intervals on tracks 7-10 in Figure 8. The highest uncertainty is marked by the estimation error of water saturation. We can estimate the kerogen volume with high accuracy, which has great practical importance in identifying the sweet spots. The average correlation between the volumetric parameters is moderate (0.47), which shows the reliability of the inversion result, too. We characterize the estimation accuracy of zone parameters with the standard deviation of their estimated values for 200 individuals. Table 2 includes the standard deviation of zone parameters both at the beginning and at the end of the

inversion procedure, as well as the relative estimation error. One can conclude the most sensitive parameters can be estimated with the highest accuracy such as the gamma ray intensity, potassium, and thorium concentration of shale, while the uranium concentration of shale, neutron porosity of mud-filtrate and the resistivity parameters such as pore water resistivity, cementation exponent and kerogen factor can be predicted with the lowest accuracy. The latter three are related to the estimation of water saturation being the least accurately estimated volumetric quantity.

DISCUSSION

In the practice of formation evaluation, it is common experience that the inversion of wireline logs is highly sensitive to the initial setting of zone parameters. If we have poor information about the studied formations and these quantities, the incorrectly set zone parameters can obviously prevent the accurate estimation of the volumetric parameters and reduce the efficiency of reservoir modeling. The current study reveals the importance of estimating the zone parameters automatically. Now we examine the difference between optimizing the zone parameters by an inversion method and arbitrarily fixing them before the inversion procedure. We make a comparative study between inversion beside fixed and varying hyperparameters using three petrophysical models given in Table 3. Model I assumes the zone parameters as unknown that are estimated by hyperparameter inversion. Model II and III include fixed zone parameters of tight gas formations, where just the volumetric parameters are estimated by local inversion (only inversion phase II is performed). The initial values of volumetric parameters are the same for all three

inversion runs: Φ_0=0.08 (v/v), $V_{sh,0}$=0.60 (v/v), $V_{k,0}$=0.02 (v/v), $S_{w,0}$=0.60 (v/v). The last column of Table 3 shows the relative data distance (D_d) between the observed and calculated data upon completion of the inversion process, which suggests that inappropriate choice (or even by slightly changing) of zone parameters can cause a 10- to 20-fold increase of the data misfit. One can also study the effect of the setting of zone parameters on the estimation of the volumetric parameters in Figure 9. The first six tracks show the fit between the observed and theoretical logs, which is the best for Model I, while for Models II and III the nuclear and resistivity logs show larger deviations. The volumetric parameters estimated by local inversion also appear to change significantly. In the case of Model II, whose zone parameters are relatively close to those of Model I, the inversion is unable to find the gas at some depth intervals (e.g., 3560-3565 m) and instead identifies a water bearing formations. The parameter setting of Model III forces the inversion process identify all the rocks as fully water saturated (S_w=1) ones over the entire depth interval. Regarding Model I and III, we obtain the following average relative change between the volumetric parameters: $\overline{\Delta\Phi/\Phi}$=13 %, $\overline{\Delta V_{sh}/V_{sh}}$=18 %, $\overline{\Delta V_k/V_k}$=27, $\overline{\Delta S_w/S_w}$=92 %. This shows that the saturation estimation has the biggest impact on the identification of the tight gas reservoirs. The derived *TOC* data shows also a 0.5–1.0 wt% difference. Model II basically overestimates (e.g., >3650 m), while Model III underestimates (3580-3590 m) the values of *TOC*. We conclude that we obtain the best inversion results when the zone parameters are estimated by hyperparameter inversion, which assists to identify the sweet spots and gives reliable volumetric quantities for the estimation of hydrocarbon reserves.

We compare the hyperparameter inversion approach with an advanced probabilistic inversion technique called interval inversion described in the “Introduction” section. In contrast to

depth-by-depth inversion, in the framework of interval inversion, we invert the dataset of a larger interval jointly and solve a highly overdetermined inverse problem to increase the accuracy of estimation and efficiency of inversion. In the interval of 3550-3617 m in Be-4 well, we invert the GR, K, Th, U, Φ_N, R_d logs while treating the zone parameters as fixed constant (see the reference values in Table 1). We discretize the volumetric parameters (Φ, V_{sh}, V_k, S_w) using Legendre polynomials of 24th degree serving as basis functions of the series expansion, then after estimating the expansion coefficients, we derive the depth distribution of the model parameters. In contrast, in hyperparameter inversion we treat the zone parameters as unknown quantities as detailed in the "Inversion of field data" section. The results obtained by the two inversion methods show good agreement, primarily regarding porosity and shale content (Fig. 10). The average of the correlation coefficients for the linear regression results is $\overline{R}$=0.84. Larger differences are basically based on the estimation (adjustment) of the zone parameters. For instance, the interval inversion reveals a hydrocarbon zone almost across the entire interval. However, the hyperparameter inversion method also identifies some water bearing layers (S_w=1), also where the interval inversion method estimates lower water saturation (gas indication). Both inversion methods are suitable for determining the estimation error of the model parameters. In both cases, the accuracy of the input data is assumed based on the values specified in the "Inversion of field data" section. In the case of the interval inversion approach, we apply equation 15 to calculate the covariance matrix of the series expansion coefficients, and then we derive the uncertainty of the volumetric parameters using the law of error propagation (Dobróka et al. 2016). Both inversion methods allow the calculation of the model covariance as a function of depth, i.e., the estimation error of the volumetric parameters is available for each depth. By plotting the occurrence of the standard deviation of the inverted parameters, we obtain a spatial distribution of the error values that is not

uniform (Fig. 11). Hyperparameter inversion is more accurate when estimating porosity (Fig. 11a) and kerogen volume (Fig. 11d), while interval inversion gives more reliable results for the water saturation (Fig. 11b) and shale content (Fig. 11c). We emphasize that the kerogen volume can be estimated most accurately with both inversion methods, for which the hyperparameter inversion achieves an order of magnitude smaller error, which is favorable from the point of view of the evaluation of unconventional reservoirs.

CONCLUSION

We extract the zone parameters appearing as delicate quantities of the tool response functions jointly with volumetric parameters by hyperparameter inversion of well logs. Avoiding the solution of the inherently underdetermined inverse problem, we divide the inversion procedure into two overdetermined inversion phases to get unique solutions for both type of unknowns. We select the unknowns by parameter sensitivity tests to improve the numerical stability of the inversion procedure. After this, we predict 11 zone parameters in a Hungarian tight gas reservoir. In other types of unconventional formations (e.g., in shale gas, volcanic reservoirs, other multimineral rocks), the number of estimated zone parameters is even higher, which requires a robust inversion strategy. The automated estimation of zone parameters not only improves the solution of the forward problem (i.e., reducing the modeling error) but may reduce the cost of

coring and laboratory measurement as one can determine the zone parameters solely (in situ and continuously) from the well logs. We also show a way to estimate the shale parameters and the kerogen volume with great accuracy as new alternative to *TOC* estimation based on wireline logs. The proposed inversion methodology not only allows the quality check of inversion results but also allows us to integrate new well log types into the inversion procedure. The rate of overdetermination can be further increased by replacing the local inversion phase with the interval inversion method, which adds the potential to estimate additional unknowns such as the fractional volumes of more matrix components. By interval inversion, the depth of zone boundaries can also be automatically determined, which designates the optimal interval for the estimation of the zone parameters. Discretizing the depth function of zone parameters by a series expansion technique, we can determine the slow (and faster) variation of the zone parameters along a borehole. By involving the interval inversion technique, the hyperparameter estimation method can be extended to an automated multi-well correlation of petrophysical parameters.

In the literature, the geophysical applications of hyperparameter estimation mostly include the determination of the not-exactly known data variances. At this point, our inversion method can be further developed, for example, to determine the control parameters of the genetic algorithm,

or to estimate the regularization parameter, the Lagrange multipliers of penalty term or other weighting factors included in the objective function of the Marquard algorithm. The significance of the proposed inversion method primarily lies in that different from the previous approaches we try to specify not the control parameters but certain geophysical quantities as hyperparameters and estimate them by inversion. In addition to the advantages presented in this study, the limitation of the inversion method is the increasing number of inverted zone parameters in complex formations, small parameter sensitivities, the strong correlation between the zone parameters and the volumetric properties, and CPU time requirement of genetic inversion. By carefully studying and improving these factors, one can apply the inversion method as innovative (gap filling) tool in formation evaluation.

ACKNOWLEDGMENTS

The research was carried out in the Project No. K-135323 supported by the National Research, Development and Innovation Office (NKFIH), Hungary. The research was partly funded by the Sustainable Development and Technologies National Programme of the Hungarian Academy of Sciences (FFT NP FTA). Thank is given to the Research Institute of Applied Earth

Sciences of the University of Miskolc for the permission of using the real data. The author expresses his gratitude to Prof. Dr. Mihály Dobróka for decades of academic mentoring work, and his valuable scientific advice and guidance.

NOMENCLATURE

a	=	tortuosity factor in Archie's equation
$\mathbf{c}$	=	vector of zone parameters
cov	=	covariance operator
$\mathbf{d}^{(th)}$	=	column vector of calculated (theoretical) data
D_d	=	relative data distance (%)
E	=	objective function of local (linearized) inversion
$\mathbf{e}$	=	deviation between the measured and calculated data vectors
F	=	fitness function to be maximized
$\mathbf{g}$	=	function relation between the data vector and model vector
$\mathbf{G}$	=	Jacobi matrix
$\mathbf{G}^{-g}$	=	symbol of generalized inverse matrix

GR = natural gamma-ray intensity (API)

GR_k = natural gamma-ray intensity of kerogen (API)

GR_{sd} = natural gamma-ray intensity of sand (API)

GR_{sh} = natural gamma-ray intensity of shale (API)

$\mathbf{I}$ = identity matrix

K = potassium concentration (%)

K^* = number of well logs simultaneously inverted in each depth

K_c = kerogen conversion factor

K_k = potassium concentration of kerogen (%)

K_{rf} = kerogen (resistivity) factor

K_{sd} = potassium concentration of sand (%)

K_{sh} = potassium concentration of shale (%)

L = number of zone parameters in genetic search

$\mathbf{m}$ = vector of model (volumetric) parameters

m = cementation exponent in Archie's equation

n = saturation exponent in Archie's equation

N = number of depth points over the processing interval

P = number of individuals in the given population

R = Pearson's correlation coefficient

R_d = deep resistivity (Ωm)

R_{sh} = shale resistivity (Ωm)

R_t = true resistivity (Ωm)

R_w = pore-water resistivity (Ωm)

T = symbol of matrix transpose

Th = thorium concentration (ppm)

Th_k = thorium concentration of kerogen (ppm)

Th_{sd} = thorium concentration of sand (ppm)

Th_{sh} = thorium concentration of shale (ppm)

TOC = total organic matter content (wt%)

U = uranium concentration (ppm)

U_k = uranium concentration of kerogen (ppm)

U_{sd} = uranium concentration of sand (ppm)

U_{sh} = uranium concentration of shale (ppm)

S_{x0} = water saturation in invaded zone (v/v)

S_w = water saturation in uninvaded zone (v/v)

V_k = kerogen volume (v/v)

V_{sh} = shale volume (v/v)

V_{sd} = sand volume (v/v)

α, β = random numbers generated in genetic search

ε = regularization term used in local (linearized) inversion

κ = empirical exponent used in saturation estimation

ρ_b = bulk density (gcm^{-3})

ρ_k = kerogen density (gcm^{-3})

σ_d = accuracy (standard deviation) of observed data

σ_m = accuracy (standard deviation) of model parameters

Φ = porosity (v/v)

Φ_N = neutron porosity (p.u.)

$\Phi_{N,h}$ = neutron porosity of hydrocarbon (p.u.)

$\Phi_{N,k}$ = neutron porosity of kerogen (p.u.)

$\Phi_{N,mf}$ = neutron porosity of mud-filtrate (p.u.)

$\Phi_{N,sd}$ = neutron porosity of sand (p.u.)

$\Phi_{N,sh}$ = neutron porosity of shale (p.u.)

Ψ = parameter sensitivity function

REFERENCES

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

Alberty, M., and K. Hashmy, 1984, Application of ULTRA to log analysis: Presented at the SPWLA Symposium Transactions, 1–17.

Archie, G. E., 1942, The electrical resistivity log as an aid in determining some reservoir characteristics: Transactions of the AIME, **146**, 54–62.

Asquith, G., and D. Krygowski, 2006, Basic well log analysis, second edition: AAPG.

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

Balázs, L., 2015, Inversion of well logging measurements with a constant interval parameter: Geosciences and Engineering, **4**, 93–104.

Balázs, L., 2021a, Correlation effect of transformed or corrected data inversion. Acta Geodaetica et Geophysica, **56**, 263–270.

Balázs, L., 2021b, Interval maximum likelihood inversion for well logging data and statistical properties of estimated parameters. Acta Geodaetica et Geophysica, **56**, 697–709.

Ball, S. M., D. M. Chace, and W. H. Fertl, 1987, 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.

Bijani, R., C. F. Ponte Neto, S. S. Martins, and J. M. Travassos, 2012, 2-D tomography of first-arrivals using the genetic algorithm with elitism: SEG Technical Program Expanded Abstracts, 1–6.

Bodin, T., M. Sambridge, N. Rawlinson, and P. Arroucau, 2012, Transdimensional tomography with unknown data noise: Geophysical Journal International, **189**, 1536–1556.

Dobróka, M., 1988, On the absorption-dispersion characteristics of channel waves propagating in coal seams of varying thickness: Geophysical Prospecting, **36**, 318–331.

Dobróka, M., and N. P. Szabó, 2011, Interval inversion of well-logging data for objective determination of textural parameters: Acta Geophysica, **59**, 907–934.

Dobróka, M., and N. P. Szabó, 2012, Interval inversion of well-logging data for automatic determination of formation boundaries by using a float-encoded genetic algorithm: Journal of Petroleum Science and Engineering, **86–87**, 144–152.

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

Feurer, M., and F. Hutter, 2019, Hyperparameter optimization, *in* F. Hutter, L. Kotthoff, and J. Vanschoren, eds., Automated machine learning. The Springer series on challenges in Machine Learning: Springer, 3–33.

Gyulai, Á., 1995, Investigating dipping bed structures by means of analytic forward modeling: Magyar Geofizika, **36**, 40–67 (in Hungarian).

Haás, J., 2013, Geology of Hungary: Springer-Verlag.

Heidari, Z., C. Torres-Verdín, and W. E. Preeg, 2012, Improved estimation of mineral and fluid volumetric concentrations from well logs in thinly bedded and invaded formations: Geophysics, **77**, WA79–WA98.

Heidari, Z., C. Torres-Verdín, and W. E. Preeg, 2013, Improved estimation of mineral and fluid volumetric concentrations in thinly bedded carbonate formations: Geophysics, **78**, D261–D269.

Holditch, S.A., 2013, Unconventional oil and gas resource development – let's do it right: Journal of Unconventional Oil and Gas Resources, **1**, 2–8.

Holland, J. H., 1975, Adaptation in natural and artificial systems: Univ. Michigan Press.

Horváth, Sz. B., 1973, The accuracy of petrophysical parameters as derived by computer processing: The Log Analyst, **14**, 16–33.

Houck, C. R., J. Joines, and M. Kay, 1995, A Genetic Algorithm for Function Optimization: A

Matlab Implementation: NCSU-IE Technical Report 95-09, North Carolina State University, 1–14.

Ingber, L., 1989, Very fast simulated re-annealing: Mathematical and Computer Modelling, **12**, 967–973.

Kadkhodaie, A., and R. Rezaee, 2016, A new correlation for water saturation calculation in gas shale reservoirs based on compensation of kerogen-clay conductivity: Journal of Petroleum Science and Engineering, **146**, 932–939.

Kuwatani, T., H. Hino, K. Nagata, T. Kawashima, M. Toriumi, and M. Okada, 2022, Hyperparameter estimation using a resolution matrix for Bayesian sensing: Inverse Problems, **38**, 124004.

Liu, W., C. Zhang, Z. Zhang, J. Li, L. Zhu, S. Hu, and X. Zhou, 2022, A new method of mineral inversion based on error analysis and static response equation error: A case study of a shale gas reservoir in the Wufeng-Longmaxi Formation: Interpretation, **10**, T117-T126.

Marquardt, D. W., 1959, Solution of non-linear chemical engineering models: Chemical Engineering Progress, **55**, 65–70.

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

Menke, W., 2022, Environmental data analysis with MATLAB or Python: Elsevier.

Michalewicz, Z., 1992, Genetic Algorithms + Data Structures = Evolution Programs: Springer-Verlag.

Narayan, J. P., and L. Yadav, 2006, Application of adaptive processing technique for the inversion of open hole logs recorded in oil fields of Indian basins: 6th International Conference & Exposition on petroleum geophysics "Kolkata 2006", 505–512.

Schlumberger, 1989, Log interpretation principles/applications: Seventh printing: Schlumberger Co.

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

Serra, O., 1984, Fundamentals of well-log interpretation: Elsevier.

Sousa, M. C., de Figueiredo, J. J. S., Régis, C. R. T. Régis, and C. B. da Silva, 2020, Interval mineral and fluid densities estimation from well-logs: Application to the Norne Field dataset: Journal of Applied Geophysics, **183**, 104199.

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

Szabó, N. P., 2018, A genetic meta-algorithm-assisted inversion approach: hydrogeological study for the determination of volumetric rock properties and matrix and fluid parameters in unsaturated formations: Hydrogeology Journal, **26**, 1935–1946.

Szabó, N.P., and M. Dobróka, 2019, Series expansion-based genetic inversion of wireline logging data: Mathematical Geosciences, **51**, 811–835.

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

Szabó, N. P., F. Remeczki, A. Jobbik, K. Kiss, and M. Dobróka, 2022, Interval inversion based well log analysis assisted by petrophysical laboratory measurements for evaluating tight gas formations in Derecske through, Pannonian basin, east Hungary: Journal of Petroleum Science and Engineering, **208**, 109607.

Szalai, S., L. Szarka, E. Prácser, F. Bosch, I. Müller, and P. Turberg, 2002, Geoelectric mapping of near - surface karstic fractures by using null arrays: Geophysics, **67**, 1769-1778.

Tarantola, A., 1987, Inverse problem theory: Methods for data fitting and model parameter estimation: Elsevier.

Tilahun, T., and J. Korus, 2023, 3D hydrostratigraphic and hydraulic conductivity modelling using supervised machine learning: Applied Computing and Geosciences, **19**, 100122.

Tissot, B. P., and D. H. Welte, 1978, Petroleum Formation and Occurrence: Springer.

Xiao, S., T. Xu, S. Reuschen, W. Nowak, and H.-J. H. Franssen, 2021, Bayesian inversion of multi-Gaussian log-conductivity fields with uncertain hyperparameters: an extension of

preconditioned Crank-Nicolson Markov chain Monte Carlo with parallel tempering: Water Resources Research, **57**, e2021WR030313.

Xu, J., L. Xu, and Y. Qin, 2017, Two effective methods for calculating water saturations in shale gas reservoirs: Geophysics, **82**, D187–D197.

Wang, Q., X. Zhengyao, H. Wang, X. Li, L. Tian, S. Yuan, R. Deng, 2022, Sensitivity analysis of log interpretation parameters of through casing reservoir saturation logging: Energy Reports, **8**, 392–400.

Zhang, S.-B., H.-J. Si, X.-M. Wu, and S.-S. Yan, 2022, A comparison of deep learning methods for seismic impedance inversion: Petroleum Science, **19**, 1019-1030.

LIST OF FIGURES

Figure 1. Fitness function of genetic algorithm expressing the probability of the zone parameters being selected for reproduction and how close they are to the solution of the inverse problem. The pair of independent variables: (a) gamma-ray intensity of shale (GR_{sh}) and uranium concentration of kerogen (U_k), (b) gamma-ray intensity of sand (GR_{sd}) and uranium concentration of shale (U_{sh}), (c) resistivity of shale (R_{sh}) and saturation exponent (n), (d) neutron porosity of kerogen (Φ_k) and neutron porosity of hydrocarbon (Φ_h).

Figure 2. Flowchart of the two-step hyperparameter estimation-based inversion algorithm used for simultaneously estimate the zone and volumetric parameters.

Figure 3. Parameter sensitivity functions (Ψ) measuring the extent of influence of zone parameters on natural gamma-ray logging data. The independent variables: the natural gamma-ray intensity of kerogen (GR_k), shale (GR_{sh}), sand (GR_{sd}); the potassium concentration of kerogen (K_k), shale (K_{sh}), sand (K_{sd}); the uranium concentration of kerogen (U_k), shale (U_{sh}), sand (U_{sd}); the thorium concentration of kerogen (Th_k), shale (Th_{sh}), sand (Th_{sd}).

Figure 4. Parameter sensitivity functions (Ψ) measuring the extent of influence of zone parameters on neutron porosity and deep resistivity data. The independent variables: the neutron porosity of kerogen ($\Phi_{n,k}$), shale ($\Phi_{n,sh}$), sand ($\Phi_{n,sd}$), hydrocarbon ($\Phi_{n,h}$), mud-filtrate ($\Phi_{n,sh}$); the resistivity of pore-water (R_w), shale (R_{sh}); tortuosity factor (a), cementation exponent (m), saturation exponent (n), kerogen resistivity factor (K_{rf}).

Figure 5. Convergence curves for the hyperparameter inversion procedure. Phase I represents the genetic inversion part of the two-step inversion procedure, Phase II is the linear optimization step. (a) Initial population of 200 individuals randomly generated at the beginning of the inversion procedure, (b) maximal fitness of individuals during the genetic search performed in the first loop of the inversion procedure, (c) fitness of the best zone parameter vector at each iteration when completing the two inversion phases, (d) relative data distance between the measured and calculated logs at each iteration when completing the two inversion phases.

Figure 6. Change of zone parameters during the hyperparameter inversion procedure. (a) The values of zone parameters during the genetic search (phase I) in the last loop (10^{th} iteration)

of the inversion procedure. (b)-(d) The values of zone parameters estimated at each iteration when completing the two inversion phases (phase I+II).

Figure 7. Hyperparameter inversion results obtained in Be-4 well, Derecske Trough, East Hungary. Measured (solid line) and calculated (dashed line) well logs: *GR* is natural gamma-ray intensity, *K, U, Th* are potassium, uranium, thorium concentration, respectively, Φ_N is neutron porosity, R_d is deep resistivity. Volumetric parameters estimated (or derived) by inversion are: Φ is porosity, S_{x0} is water saturation (invaded zone), S_w is water saturation (uninvaded zone), V_{sh} is shale volume, V_k is kerogen volume, V_{sd} is sand volume, *TOC* is total organic carbon content.

Figure 8. Uncertainty analysis of hyperparameter inversion results. Confidence intervals of measured data (data accuracy) and petrophysical properties (estimation error) are indicated with shaded regions around the well logs, σ is the standard deviation. The observed logs: *GR* is natural gamma-ray intensity, *K, U, Th* are potassium, uranium, thorium concentration, respectively, Φ_N is neutron porosity, R_d is deep resistivity. Volumetric parameters estimated by inversion: Φ is porosity, S_w is water saturation in the uninvaded zone, V_{sh} is shale volume, V_k is kerogen volume.

Figure 9. Hyperparameter inversion results obtained for three different petrophysical models. Zone parameters of Model I-III are given in Table 3. In tracks 1-6, measured logs are denoted by upper index (m), calculated logs are indicated by upper index (c). The input logs: *GR* is natural gamma-ray intensity, *K, U, Th* are potassium, uranium, thorium concentration, respectively, Φ_N is neutron porosity, R_d is deep resistivity. Volumetric parameters estimated (or derived) by inversion: Φ is porosity, S_w is water saturation in the uninvaded zone, V_{sh} is shale volume, V_k is kerogen volume, *TOC* is total organic content.

Figure 10. Linear regression relations between the same volumetric parameters estimated by hyperparameter inversion and interval inversion, respectively. The studied parameters are (a) porosity (Φ), (b) water saturation in the uninvaded zone (S_w), (c) shale volume (V_{sh}), (d) kerogen volume (V_k). Quantity R is the Pearson's correlation coefficient.

Figure 11. Histograms of the estimation error of volumetric parameters extracted by hyperparameter inversion and interval inversion, respectively. From the model covariance matrix, we derive (a) the standard deviation of porosity (σ_Φ), (b) the standard deviation of water saturation in the uninvaded zone (σ_{Sw}), (c) the standard deviation of shale volume (σ_{Vsh}), (d) the standard deviation of kerogen volume (σ_{Vk}).

LIST OF TABLES

Table 1. Zone parameters of tool response functions used for the evaluation of tight gas formations and the parameter selection results based on sensitivity analysis (the threshold value of sensitivity is chosen as 0.2).

Table 2. Zone parameter estimation results given by hyperparameter inversion of well logs in Be-4 well, Derecske Trough, East Hungary.

Table 3. Well logging inversion results obtained by different settings of zone parameters.

Table 1. Zone parameters of tool response functions used for the evaluation of tight gas formations and the parameter selection results based on sensitivity analysis (the threshold value of sensitivity is chosen as 0.2).

Zone parameter	Symbol (unit)	Reference value	Sensitivity value	Inversion parameter
Gamma ray intensity of shale	GR_{sh} (API)	140	0.60	unknown
Gamma ray intensity of kerogen	GR_k (API)	3000	0.40	unknown
Gamma ray intensity of sand	GR_{sd} (API)	0	$<10^{-3}$	fixed
Potassium concentration of shale	K_{sh} (%)	5	0.78	unknown
Potassium concentration of kerogen	K_k (%)	12	0.07	fixed
Potassium concentration of sand	K_{sd} (%)	0	$4 \cdot 10^{-3}$	fixed
Uranium concentration of shale	U_{sh} (ppm)	3.5	0.34	unknown
Uranium concentration of kerogen	U_k (ppm)	160	0.64	unknown
Uranium concentration of sand	U_{sd} (ppm)	0	$5 \cdot 10^{-3}$	fixed

Thorium concentration of shale	TH_{sh} (ppm)	30	1.0	unknown
Thorium concentration of kerogen	TH_{k} (ppm)	15	0.02	fixed
Thorium concentration of sand	TH_{sd} (ppm)	0	0.002	fixed
Neutron porosity of mud-filtrate	$\Phi_{N,mf}$ (p.u.)	98	0.32	unknown
Neutron porosity of shale	$\Phi_{N,sh}$ (p.u.)	23	0.72	unknown
Neutron porosity of kerogen	$\Phi_{N,k}$ (p.u.)	64	0.06	fixed
Neutron porosity of sand	$\Phi_{N,sd}$ (p.u.)	-4	0.06	fixed
Neutron porosity of hydrocarbon (gas)	$\Phi_{N,h}$ (p.u.)	10	$7 \cdot 10^{-3}$	fixed
Pore-water resistivity	R_w (Ωm)	0.01	0.66	unknown
Shale resistivity	R_{sh} (Ωm)	2	0.08	fixed
Saturation exponent	n	2	0.19	fixed
Cementation exponent	m	2	0.95	unknown
Tortuosity factor	a	1	0.19	fixed

Kerogen resistivity factor	K_{rf}	6800	0.32	unknown

Table 2. Zone parameter estimation results given by hyperparameter inversion of well logs in Be-4 well, Derecske Trough, East Hungary.

Zone parameter	Symbol (unit)	Search domain	Estimated value	Standard deviation (initial population)	Standard deviation (last generation)	Relative error (%)
Gamma ray intensity of shale	GR_{sh} (API)	100-200	187	29.4	3.4	12
Gamma ray intensity of kerogen	GR_k (API)	0-4000	1645	1130	247	22
Potassium concentration of shale	K_{sh} (%)	0-10	4.65	2.6	0.18	7
Uranium concentration of shale	U_{sh} (ppm)	0-10	2.28	3.0	1.85	62
Uranium concentration of kerogen	U_k (ppm)	100-300	198	55.1	13.9	25
Thorium concentration of shale	TH_{sh} (ppm)	0-50	28.7	14.4	0.78	5
Neutron porosity of mud-filtrate	$\Phi_{N,mf}$ (p.u.)	90-100	99.8	2.8	1.49	53
Neutron porosity of shale	$\Phi_{N,sh}$ (p.u.)	0-40	22.43	11.0	0.73	7
Pore-water resistivity	R_w (Ωm)	0.001-0.1	0.023	0.029	0.012	41

Cementation exponent	m	1.0-2.2	1.51	0.20	0.06	30
Kerogen resistivity factor	K_{rf}	4000-10000	6162	1729	796	46

Table 3. Well logging inversion results obtained by different settings of zone parameters.

	Zone parameters	GR_{sh} (API)	GR_k (API)	K_{sh} (%)	U_{sh} (ppm)	U_k (ppm)	TH_{sh} (ppm)	$\Phi_{N,mf}$ (p.u.)	$\Phi_{N,sh}$ (p.u.)	R_w (Ωm)	m	K_{rf}	$D_{d,0}$ (%)	D_d (%)
Model I	unknown	187	1645	4.6	2.3	198	28.7	99.9	22.4	0.02	1.51	6162	47.5	4.4
Model II	fixed	120	500	15	2	200	25	95	28	0.01	2	7000	137	49
Model III	fixed	110	1000	10	5	250	22	97	28	0.08	2	9000	253	107

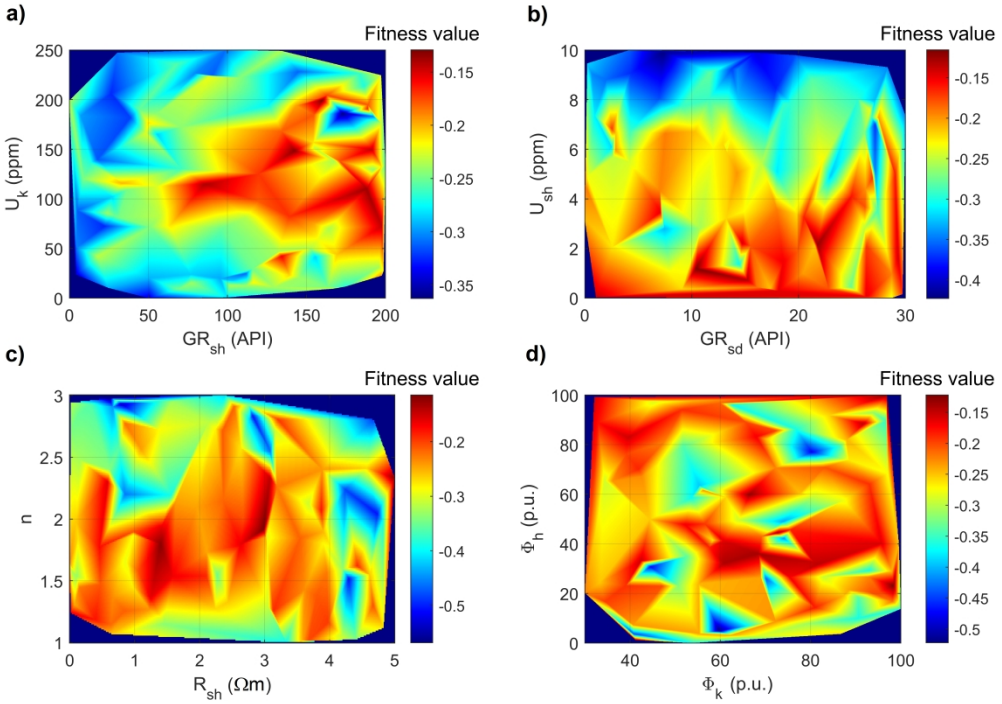

Figure 1. Fitness function of genetic algorithm expressing the probability of the zone parameters being selected for reproduction and how close they are to the solution of the inverse problem. The pair of independent variables: (a) gamma-ray intensity of shale (GRsh) and uranium concentration of kerogen (Uk), (b) gamma-ray intensity of sand (GRsd) and uranium concentration of shale (Ush), (c) resistivity of shale (Rsh) and saturation exponent (n), (d) neutron porosity of kerogen (Φk) and neutron porosity of hydrocarbon (Φh).

1263x890mm (120 x 120 DPI)

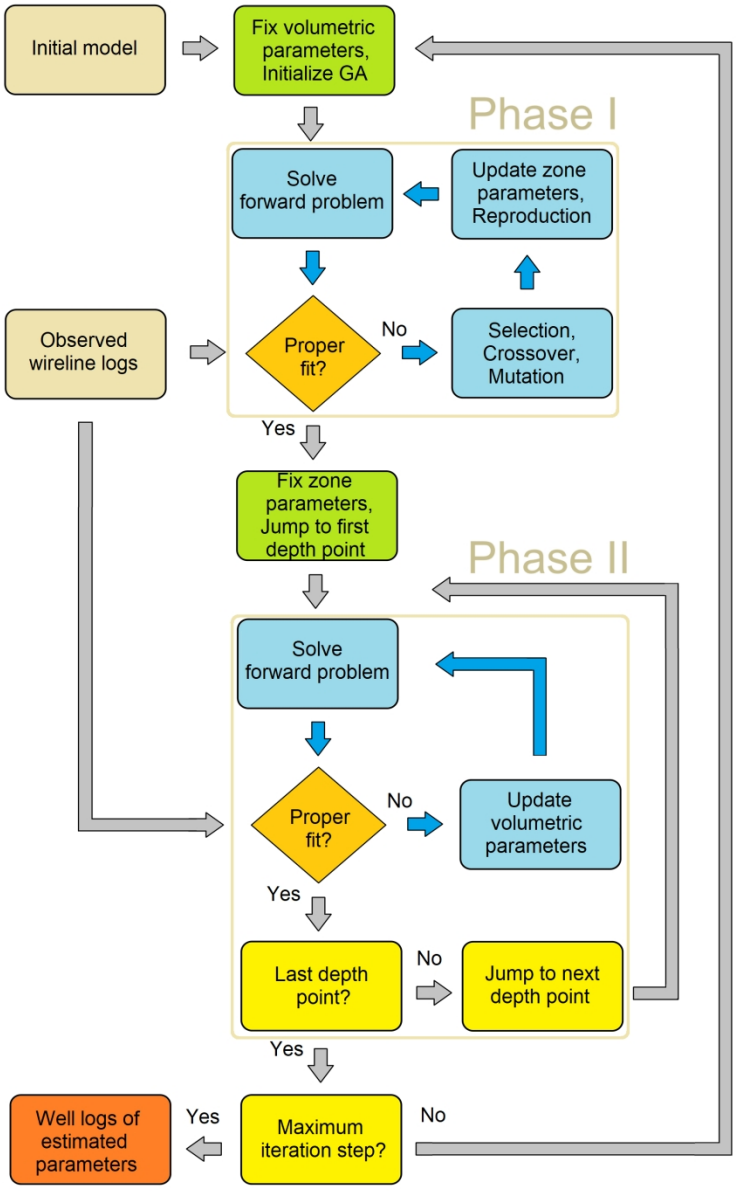

Figure 2. Flowchart of the two-step hyperparameter estimation-based inversion algorithm used for simultaneously estimate the zone and volumetric parameters.

396x608mm (120 x 120 DPI)

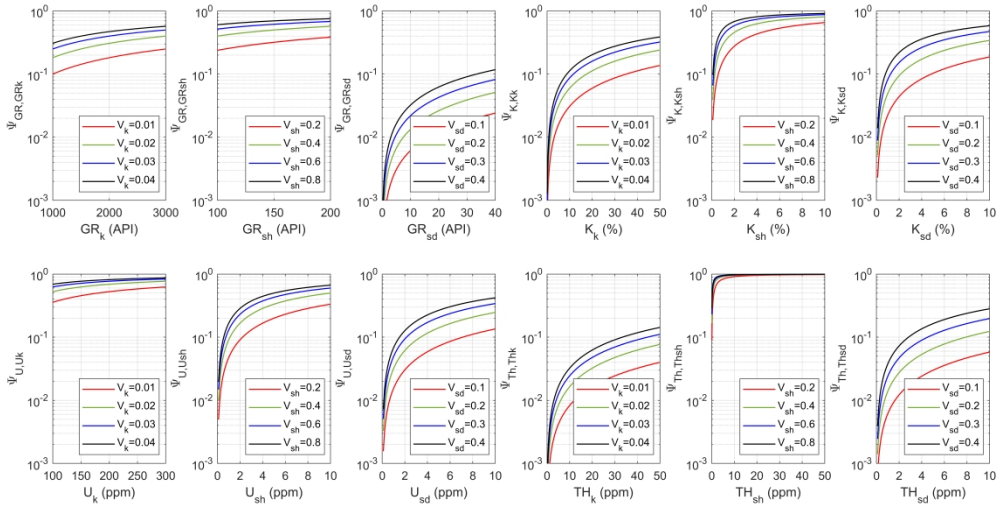

Figure 3. Parameter sensitivity functions (Ψ) measuring the extent of influence of zone parameters on natural gamma-ray logging data. The independent variables: the natural gamma-ray intensity of kerogen (GRk), shale (GRsh), sand (GRsd); the potassium concentration of kerogen (Kk), shale (Ksh), sand (Ksd); the uranium concentration of kerogen (Uk), shale (Ush), sand (Usd); the thorium concentration of kerogen (Thk), shale (Thsh), sand (Thsd).

331x167mm (600 x 600 DPI)

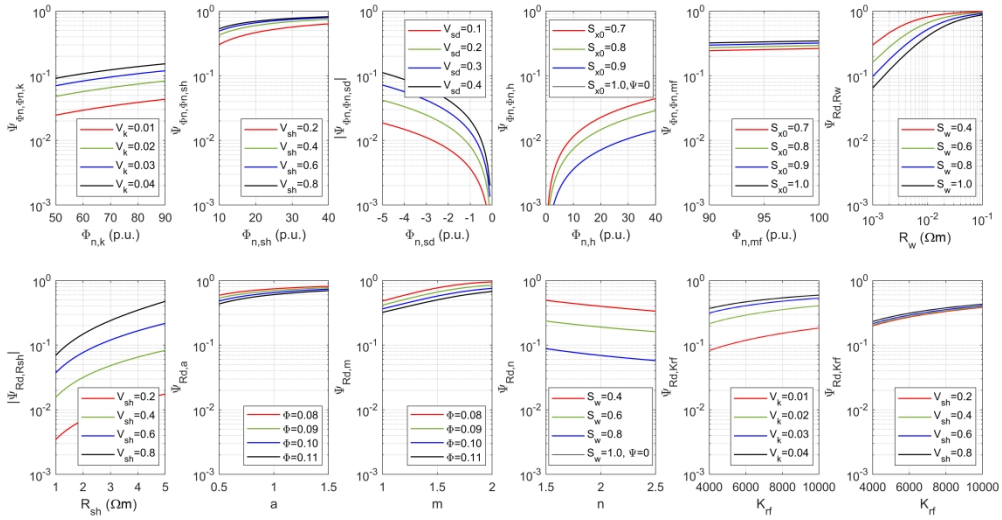

Figure 4. Parameter sensitivity functions (Ψ) measuring the extent of influence of zone parameters on neutron porosity and deep resistivity data. The independent variables: the neutron porosity of kerogen (Φn,k), shale (Φn,sh), sand (Φn,sd), hydrocarbon (Φn,h), mud-filtrate (Φn,sh); the resistivity of pore-water (Rw), shale (Rsh); tortuosity factor (a), cementation exponent (m), saturation exponent (n), kerogen resistivity factor (Krf).

1617x836mm (120 x 120 DPI)

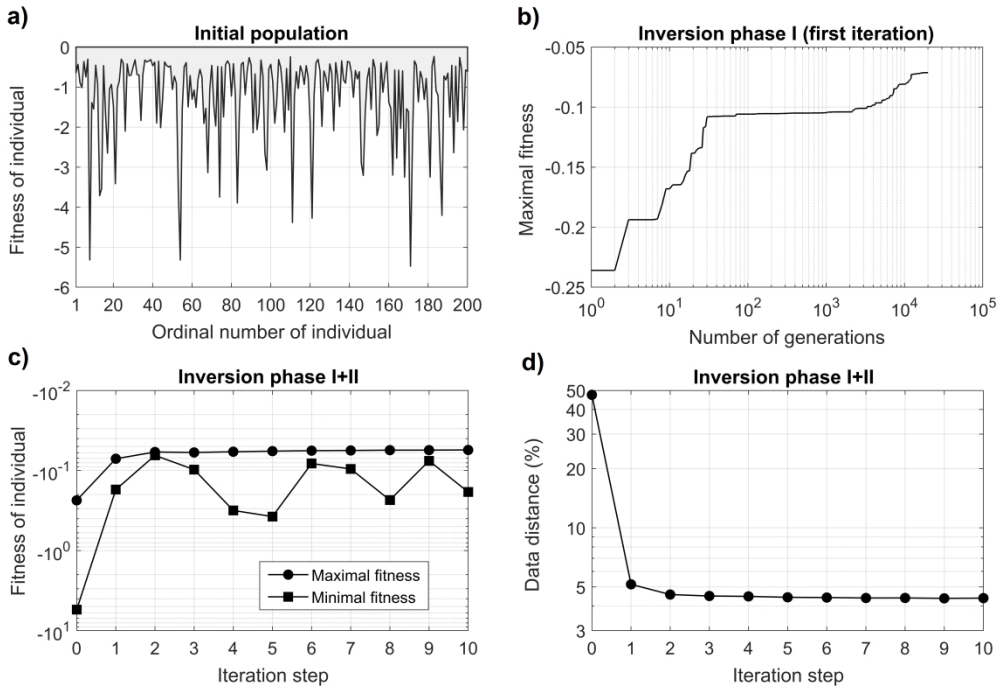

Figure 5. Convergence curves for the hyperparameter inversion procedure. Phase I represents the genetic inversion part of the two-step inversion procedure, Phase II is the linear optimization step. (a) Initial population of 200 individuals randomly generated at the beginning of the inversion procedure, (b) maximal fitness of individuals during the genetic search performed in the first loop of the inversion procedure, (c) fitness of the best zone parameter vector at each iteration when completing the two inversion phases, (d) relative data distance between the measured and calculated logs at each iteration when completing the two inversion phases.

1307x902mm (120 x 120 DPI)

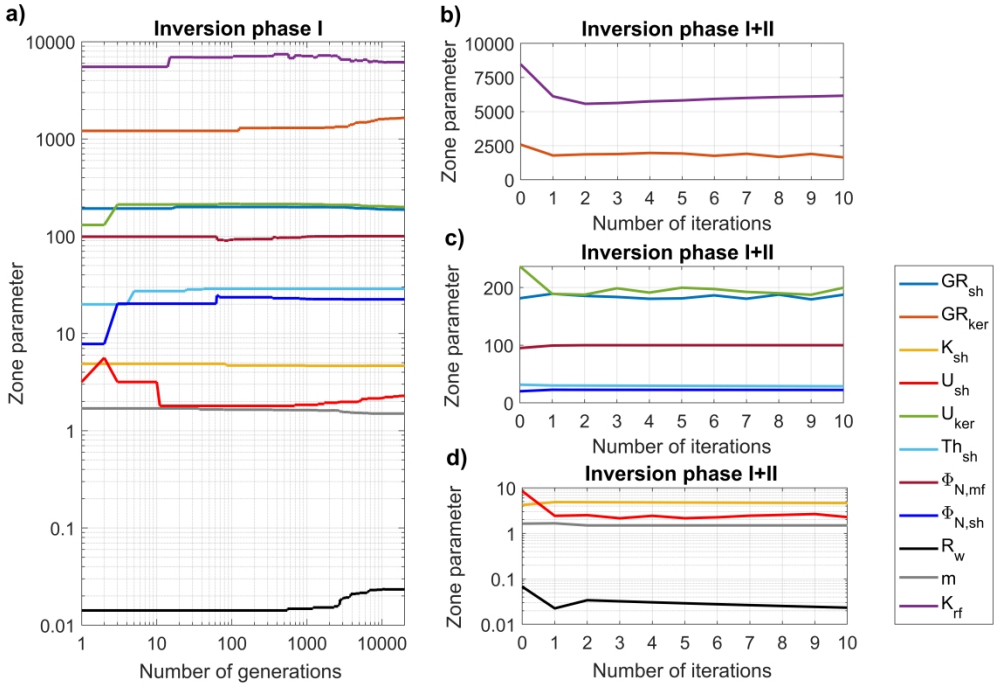

Figure 6. Change of zone parameters during the hyperparameter inversion procedure. (a) The values of zone parameters during the genetic search (phase I) in the last loop (10th iteration) of the inversion procedure. (b)-(d) The values of zone parameters estimated at each iteration when completing the two inversion phases (phase I+II).

1359x934mm (120 x 120 DPI)

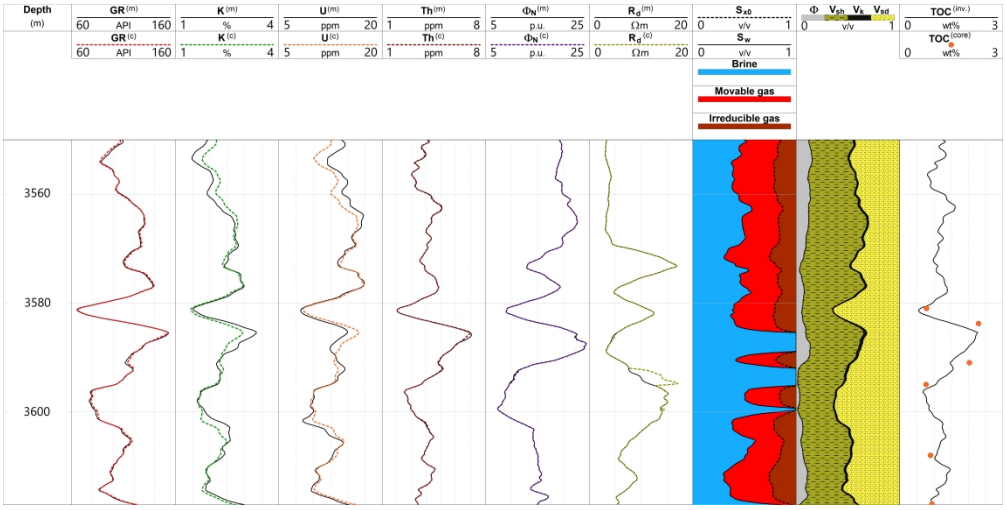

Figure 7. Hyperparameter inversion results obtained in Be-4 well, Derecske Trough, East Hungary. Measured (solid line) and calculated (dashed line) well logs: GR is natural gamma-ray intensity, K, U, Th are potassium, uranium, thorium concentration, respectively, ΦN is neutron porosity, Rd is deep resistivity. Volumetric parameters estimated (or derived) by inversion are: Φ is porosity, Sx0 is water saturation (invaded zone), Sw is water saturation (uninvaded zone), Vsh is shale volume, Vk is kerogen volume, Vsd is sand volume, TOC is total organic carbon content.

1841x921mm (120 x 120 DPI)

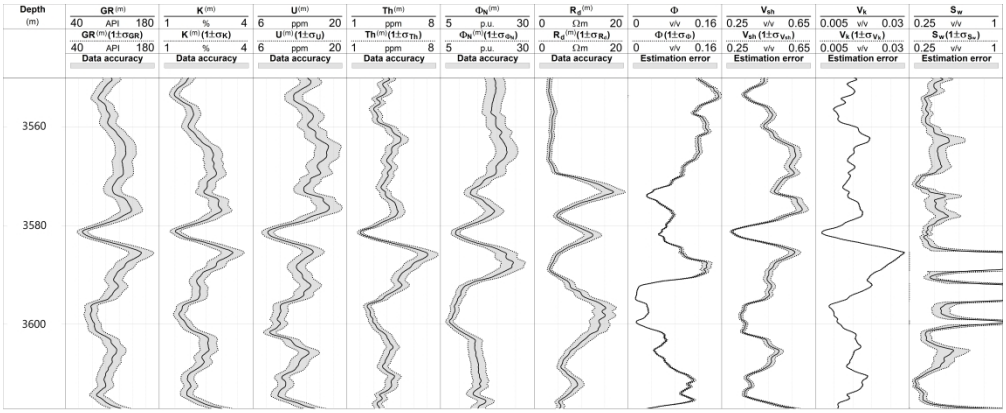

Figure 8. Uncertainty analysis of hyperparameter inversion results. Confidence intervals of measured data (data accuracy) and petrophysical properties (estimation error) are indicated with shaded regions around the well logs, σ is the standard deviation. The observed logs: GR is natural gamma-ray intensity, K, U, Th are potassium, uranium, thorium concentration, respectively, ΦN is neutron porosity, Rd is deep resistivity. Volumetric parameters estimated by inversion: Φ is porosity, Sw is water saturation in the uninvaded zone, Vsh is shale volume, Vk is kerogen volume.

2540x1032mm (96 x 96 DPI)

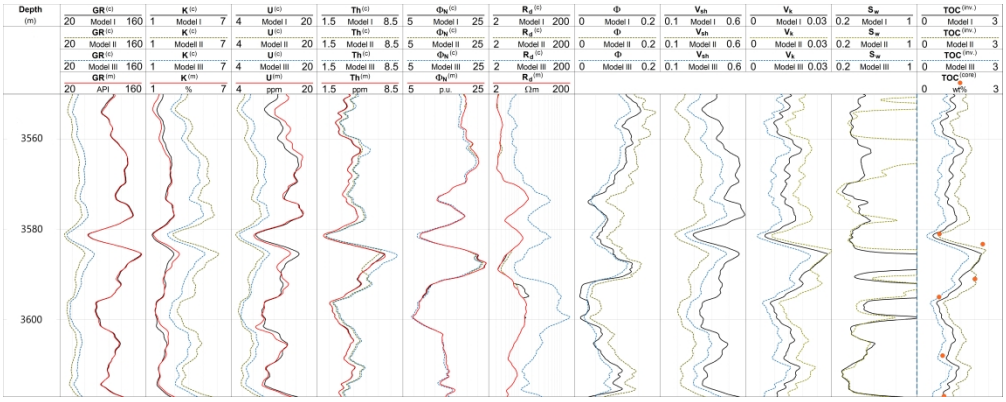

Figure 9. Hyperparameter inversion results obtained for three different petrophysical models. Zone parameters of Model I-III are given in Table 3. In tracks 1-6, measured logs are denoted by upper index (m), calculated logs are indicated by upper index (c). The input logs: GR is natural gamma-ray intensity, K, U, Th are potassium, uranium, thorium concentration, respectively, ΦN is neutron porosity, Rd is deep resistivity. Volumetric parameters estimated (or derived) by inversion: Φ is porosity, Sw is water saturation in the uninvaded zone, Vsh is shale volume, Vk is kerogen volume, TOC is total organic content.

2000x784mm (120 x 120 DPI)

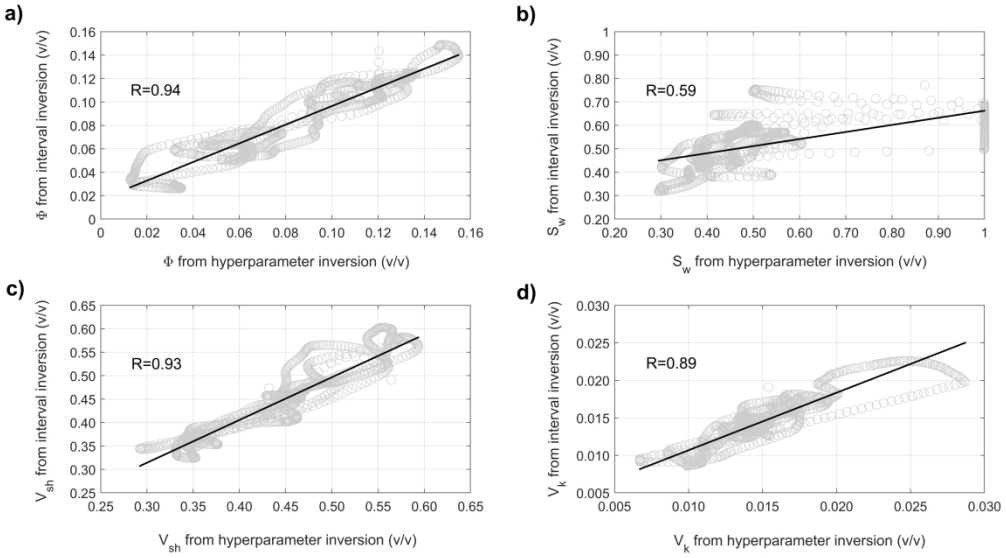

Fig. 10. Linear regression relations between the same volumetric parameters estimated by hyperparameter inversion and interval inversion, respectively. The studied parameters are (a) porosity (Φ), (b) water saturation in the uninvaded zone (Sw), (c) shale volume (Vsh), (d) kerogen volume (Vk). Quantity R is the Pearson's correlation coefficient.

318x176mm (300 x 300 DPI)

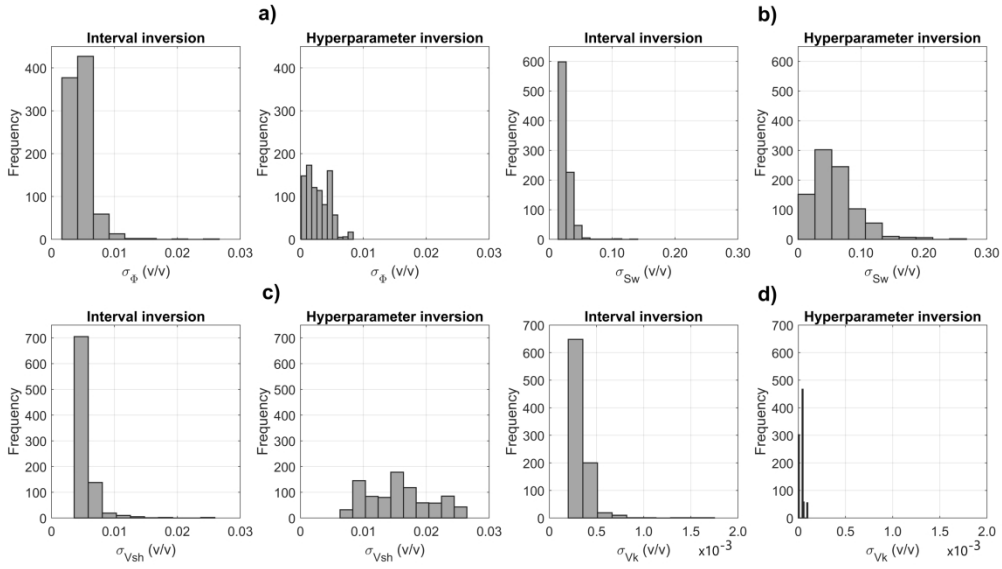

Fig. 11. Histograms of the estimation error of volumetric parameters extracted by hyperparameter inversion and interval inversion, respectively. From the model covariance matrix, we derive (a) the standard deviation of porosity (σΦ), (b) the standard deviation of water saturation in the uninvaded zone (σSw), (c) the standard deviation of shale volume (σVsh), (d) the standard deviation of kerogen volume (σVk).

325x182mm (300 x 300 DPI)