

Well Log Analysis by Global Optimization-based Interval Inversion Method

Mihály Dobróka and Norbert Péter Szabó

Abstract Artificial intelligence methods play an important role in solving an optimization problem in well log analysis. Global optimization procedures such as genetic algorithms and simulated annealing methods offer robust and highly accurate solution to several problems in petroleum geosciences. According to experience, these methods can be used effectively in the solution of well-logging inverse problems. Traditional inversion methods are used to process the borehole geophysical data collected at a given depth point. As having barely more types of probes than unknowns in a given depth, a set of marginally over-determined inverse problems has to be solved along a borehole. This single inversion scheme represents a relatively noise-sensitive interpretation procedure. For the reduction of noise, the degree of over-determination of the inverse problem must be increased. To fulfill this requirement, the so-called interval inversion method is developed, which inverts all data from a greater depth interval jointly to estimate petrophysical parameters of hydrocarbon reservoirs to the same interval. The chapter gives a detailed description of the interval inversion problem, which is solved by a series expansion-based discretization technique. Different types of basis functions can be used in series expansion depending on the geological structure to treat much more data against unknowns. The high degree of over-determination significantly increases the accuracy of parameter estimation. The quality improvement in the accuracy of estimated model parameters often leads to a more reliable calculation of hydrocarbon reserves. The knowledge of formation boundaries is also required for reserve calculation. Well logs do contain information about layer thicknesses, which cannot be extracted by the traditional local inversion approach. The interval inversion method is applicable to derive the layer boundary coordinates and certain zone parameters involved in the interpretation problem automatically. In this chapter, it is analyzed how to apply a fully automated procedure for the determination of rock interfaces and petrophysical parameters of hydrocarbon formations. Cluster analysis of well-logging data is performed as a preliminary data processing

M. Dobróka (✉) · N.P. Szabó
Department of Geophysics, University of Miskolc, H-3515 Miskolc-Egyetemváros,
Miskolc, Hungary
e-mail: dobroka@uni-miskolc.hu

step before inversion. The analysis of cluster number log allows the separation of formations and gives an initial estimate for layer thicknesses. In the global inversion phase, the model including petrophysical parameters and layer boundary coordinates is progressively refined to achieve an optimal solution. The very fast simulated re-annealing method ensures the best fit between the measured data and theoretical data calculated on the model. The inversion methodology is demonstrated by a hydrocarbon field example, which shows an application for shaly sand reservoirs. The theoretical part of the chapter gives a detailed mathematical formulation of the inverse problem, while the case study focuses on the practical details of its solution by using artificial intelligence tools.

Keywords Well-logging · Interval inversion · Global optimization · Simulated annealing · Cluster analysis · Calculation of hydrocarbon reserves · Hungary

1 Introduction

Geophysical surveying methods with their measuring and evaluation results support the exploration of the Earth and its outer environment. Borehole geophysics is abounding in observed information on the geological formations that are intersected by the drill hole. Well-logging data measured by different probes are recorded along depth in the form of well logs. The processing of open-hole logging data enables to determine some geometrical (e.g., thickness or dip of layers) and petrophysical properties such as porosity, water saturation, composition of rock matrix, and permeability that form an integral part of geological interpretation. Nowadays, there is an ever-increasing claim to the quality of well logs and interpretation results. This is especially important in oil field applications, where a precise calculation of hydrocarbon reserves should be made in complex geological environments.

The advent of inverse modeling (abbreviated as inversion)-based data processing methods was facilitated by the quick evolution of well-logging interpretation systems in the 1980s. In the early years, deterministic techniques solving linear sets of equations or using cross-plot-based graphical methods were applied. These methods gave a solution in several consecutive steps at which the petrophysical parameters were extracted one by one in different procedures (Serra 1984). It was the increased storage capacity and processor speed of computers that promoted the use of simultaneous processing of well logs. The benefit of using the data and petrophysical parameters as statistical variables was unequivocal in the improvement of the quality of interpretation results. Nowadays, the inversion methods are widely used in the petrophysical practice as they give a quick, largely automatic and reliable estimate to the vertical distributions of petrophysical parameters and their estimation errors. The biggest service companies offer inversion-based well-logging interpretation systems, e.g., Global by Schlumberger (Mayer and Sibbit 1980),

Ultra by Gearhart (Alberty and Hashmy 1984), or Optima by Baker Hughes (Ball et al. 1987). The development of these methods is strongly focused in scientific research, too.

Local inversion is the most commonly used technique for the evaluation of borehole geophysical data. Several implementations used in the oil and gas industry are well-known. They have in common that a local value of any petrophysical property is estimated to one depth point using the data measured by different probes in the same depth. In the terminology of geophysical inversion, it is a narrow type of over-determined inverse problem, where the total number of data is barely more than that of the unknown model parameters. The data and the model are connected by probe response functions that are used to calculate theoretical logs in the forward modeling phase of the inversion procedure. By assuming a petrophysical model, one can calculate theoretical well logs, which are then compared with real measurements. The actual model is progressively refined until a proper fit is achieved between the predictions and observations. Local data processing comprises a set of separate inversion runs in adjacent measuring points for the logging interval. It is a general experience that in the inversion of small number of observations, the inversion result is strongly influenced by the uncertainty of measured data. The noise of data highly affects the quality of parameter estimation; thus, the accuracy and reliability of local inversion results are relatively limited. The measurement accuracy of logging tools is prescribed that can be improved seldom with the use of any data processing method. It is a fundamental task to reduce the amount of estimation errors of inversion parameters. In one hand, one can develop more realistic probe response functions. This also means that one tends to set a model approximating the geological structure better. As a result, one can calculate such data by the response functions that are closer to the real observations. Petrophysical research deals with the development of these types of procedures that reduces the model errors. By the above contexture, it is unequivocal that another alternative to improve the quality of parameter estimation can only be facilitated by the further development of the inversion procedures. The most important requirement of the development is the improvement of accuracy and reliability of parameter estimation. For this purpose, the most essential task is the increase of data used in one interpretation procedure. In the framework of local inversion, it leads to the expansion of log types, which is of course restricted and implies additional charges. There is a more effective technique to increase the number of data without extra cost. In the so-called interval inversion procedure, all data of a longer logging interval are processed jointly to determine the characteristic values of petrophysical parameters of several rock units. As a result of the formulation of the interval inversion problem, at least one order of magnitude higher number of data than unknowns can be processed together compared with local inversion. This bears great influence on the accuracy and reliability of the extracted petrophysical parameters. The interval inversion method was introduced in Dobróka (1995), where depth-dependent probe response functions were used (instead of local ones) in the forward problem to give an estimate to the vertical distributions of petrophysical parameters for the entire logging interval. The interval inversion procedure

allows to treat increasing number of inversion unknowns without significant decrease of over-determination (data-to-unknowns) ratio. As a new feature, additional unknowns can be determined together with conventional petrophysical parameters in the same inversion procedure. In Dobróka and Szabó (2012), the possibilities of the determination of formation thicknesses were studied, where the starting model for layer-boundaries was set by external procedure. In this chapter, we suggest a fully automatic inversion strategy using a series expansion-based parameter discretization scheme to estimate the formation boundary coordinates and petrophysical parameters in one inversion procedure for a more objective calculation of hydrocarbon reserves.

2 Inverse Problem of Borehole Geophysics

In well-logging inversion, the model parameters of the geological structure are determined in the knowledge of measurement data and approximate formulae of response functions. The aim of interpretation was the lithological separation of formations and the estimation of layer thicknesses and petrophysical properties of formations such as effective porosity, water and hydrocarbon saturation, shale content, mineral volumes, and permeability to infer the quantity and quality of mineral resources. Among them, only those parameters can be determined by inversion, which are contained explicitly in the set of probe response functions and to which almost all types of data are sufficiently sensitive.

The inverse problem of borehole geophysics is classically a joint inversion problem with the particular feature that the quantities included in probe response functions can be divided into two groups. The first group comprises the so-called zone parameters, which are either constants or varying slowly over a longer depth interval (e.g., pore-water resistivity and cementation exponent). The layer parameters form the second group that are nearly constant in a given layer (e.g., porosity and mineral volume). In the practice of well-logging inversion, the zone parameters are treated as external constants that are a priori given in the inversion procedure. This simplification is compulsory in local inversion because the total number of suitable well logs is no more than 10–12, which sets a limit to the number of designated unknown quantities. If the zone parameters were treated as unknowns, an underdetermined (ambiguous) inverse problem would be encountered. In the k th local response equation

$$\varphi^{(k)} = g^{(k)}(m_1, \dots, m_P, M_1, \dots, M_L) \quad (k = 1, 2, \dots, S) \quad (1)$$

the layer parameters (m_1, \dots, m_P) are only determined by inversion, while the zone parameters (M_1, \dots, M_L) are fixed during the procedure. On the left side of Eq. (1), the calculated value of the k th logging data can be found (S is the number of applied probes). As $\varphi^{(k)}$ normally represents a nonlinear functional relationship, thus a nonlinear over-determined inverse problem is posed in the case of $S > P$.

2.1 Theory of Local Inversion

In formulating the local inverse problem, all data measured in a given depth point are collected in a column vector

$$\mathbf{d} = \{d_1, \dots, d_N\}^T, \quad (2)$$

where (d_1, d_2, d_3, \dots) represent different types of logs such as natural gamma-ray intensity, neutron porosity, and density (T is the symbol of matrix transpose). The theoretical values of the above data can be calculated by the response equations defined in Eq. (1). Let the computed data be represented by vector

$$\boldsymbol{\varpi} = \{\varphi_1, \dots, \varphi_N\}^T, \quad (3)$$

where the k th response equation is as follows:

$$\varphi^{(k)} = g^{(k)}(m_1, \dots, m_P). \quad (4)$$

The nonlinear functional relationship $g^{(k)}$ can be approximated by its Taylor series truncated at the first order

$$\varphi^{(k)} = \varphi^{(k)}(\mathbf{m}_0) + \sum_{i=1}^P \left(\frac{\partial \varphi^{(k)}}{\partial m_i} \right)_{m_0} \delta \mathbf{m}, \quad (5)$$

where the series expansion is performed around point \mathbf{m}_0 , which denotes the vector of initial model parameters. Equation (5) is expressed in vector representation

$$\boldsymbol{\varpi} = \boldsymbol{\varpi}^{(o)} + \mathbf{G} \delta \mathbf{m}, \quad (6)$$

where $\boldsymbol{\varpi}^{(o)} = \boldsymbol{\varpi}(\mathbf{m}_0)$ and $G_{ki} = \left(\frac{\partial \varphi_k}{\partial m_i} \right)_{m_0}$ is the Jacobi's (parameter sensitivity) matrix. The parameter correction vector $\delta \mathbf{m}$ is estimated by the damped least squares method, which minimizes the Euclidean norm of the following deviation vector

$$\mathbf{e} = \mathbf{d} - \boldsymbol{\varpi}^{(o)} - \mathbf{G} \delta \mathbf{m} \quad (7)$$

with a side condition that $|\delta \mathbf{m}|^2$ is minimal. The objective function of the inverse problem is as follows:

$$E = \sum_{k=1}^N e_k^2 + \lambda \sum_{i=1}^P \delta m_i^2, \quad (8)$$

where λ is a positive damping factor. With the substitution of $\delta\mathbf{d} = \mathbf{d} - \boldsymbol{\varpi}^{(o)}$, the following solution is derived

$$\delta\mathbf{m} = (\mathbf{G}^T\mathbf{G} + \lambda\mathbf{I})^{-1}\mathbf{G}^T\delta\mathbf{d}. \quad (9)$$

By solving Eq. (9), the inversion procedure is continued in a given point of the model space

$$\mathbf{m} = \mathbf{m}_0 + \delta\mathbf{m} \quad (10)$$

until a stopping criterion is met. The local inversion procedure differs from the Levenberg–Marquardt algorithm only in the consideration of a priori knowledge. Thence some criteria for the lower and upper bounds of the unknowns as well as for the sum of the specific volumes of rock constituents (material balance equation) must be fulfilled. Besides applying these constrains, another program development question is that any parameter may be set fixed in the iteration procedure. The third group of unknowns of the well-logging interpretation problem is formed by the layer boundary coordinates or layer thicknesses. Their role in local inversion is unique, because they are not contained explicitly in the probe response equations. Thus, their estimation by local inversion is out of the question. The measurement data set does contain information on the boundaries that are of great interest in oil field applications, e.g., in the estimation of hydrocarbon reserves. The determination of layer-boundaries is realized commonly in well log analysis not within the inversion procedure.

2.2 Depth-Dependent Response Functions

For the calculation of layer thicknesses and zone parameters, a new inversion strategy called interval inversion was developed. Consider the petrophysical (layer) parameters (m_1, \dots, m_P) as the function of depth. Based on Eq. (4), the k th depth-dependent response function is as follows:

$$\varphi^{(k)}(z) = g^{(k)}(m_1(z), \dots, m_P(z)). \quad (11)$$

In the general case, Eq. (11) contains also the functions of zone parameters (M_1, \dots, M_L) , which can be determined by the interval inversion method (Dobróka and Szabó 2011). The discretization of model parameters $m_1(z), \dots, m_P(z)$ can be performed by several manners. In the case of layerwise homogeneous model, a series expansion technique with proper basis functions including the coordinates of boundaries answers the purpose. Let $(B_1^{(i)}, \dots, B_{Q_i}^{(i)})$ be the series expansion coefficients of the i th model parameter $m_i(z)$. The response function in Eq. (11) takes the form as follows:

$$\varphi^{(k)}(z) = g^{(k)}\left(B_1^{(1)}, \dots, B_{Q_1}^{(1)}, \dots, B_1^{(P)}, \dots, B_{Q_P}^{(P)}, Z_1, \dots, Z_R, z\right), \quad (12)$$

where (Z_1, \dots, Z_R) denote the coordinates of layer-boundaries (Q_i is the requisite number of expansion coefficients describing the relevant model parameter). The above response function is valid in the entire interval, in which the series expansion coefficients must be chosen in such a way that the values of $\varphi^{(k)}(z)$ in each depth fit to measurement data $d^{(k)}(z)$ with the highest possible accuracy. The aim of the inversion procedure was the estimation of coefficients B , in which all data of the observed interval are inverted. This inverse problem is highly over-determined, because the number of data is several times higher than that of the unknown expansion coefficients. In local inversion, the over-determination ratio is at the best two. On the contrary, in interval inversion the same ratio may reach 50–60. Under this circumstance, the boundary coordinates (Z_1, \dots, Z_R) can be treated also as inversion unknowns to determine them with the expansion coefficients without significant reduction of the over-determination ratio. The above procedure is called interval inversion including the depth interval where the series expansion is applied for the model parameters (Dobrůka 1995). It is assumed that $d^{(k)}$ in a given depth represents a punctual data, i.e., the linear dimensions of the observed volume are smaller than the thickness of layers.

3 The Theory of Interval Inversion Method

In geophysical data processing, the term of joint inversion is used when different types of data sets are inverted together in one interpretation procedure. The data sets are measured either by different physical principles or by the same principle but in various measurement arrays. All data measured at different spread layouts carry information on the same geological structure. The theoretical values of data sets integrated into the joint inversion procedure are calculated in the knowledge of all model parameters by a proper forward modeling algorithm, that is, the data may depend on each model parameter. The more the parameters of the geological structure appear in the determination of different data sets, the more successful the solution to the inverse problem can be given. The use of such data sets that are depending only on separated groups of model parameters is unbeneficial compared with independent inversion. In the latter case, the solution will not be more accurate or reliable at all. The local inversion of well-logging data utilizes several data sets based on different physical principles (e.g., nuclear, acoustic, and electric methods), where each datum in the inversion procedure is acquired from the same depth. The observed datum does not depend on the parameters of outlying layers. In this case, therefore, the term of joint inversion can be used only in a restricted sense. It can readily be understood that in the interval inversion approach, it is easy to develop such procedures that allow to exploit all the advantages of joint inversion.

For approximating the depth variations of petrophysical parameters $m_1(z), \dots, m_P(z)$, a series expansion technique is suggested as follows:

$$m_i(z) = \sum_{q=1}^{Q_i} B_q^{(i)} \psi_q(z, Z_1, \dots, Z_R), \quad (13)$$

where $B_q^{(i)}$ are expansion coefficients, $\psi_q(z, Z_1, \dots, Z_R)$ are properly chosen (known) depth-dependent basis functions including the layer-boundaries (Q_i is the requisite number of expansion coefficients describing the i th model parameter). Combining Eqs. (12) and (13), the total number of unknowns is $\sum Q_i$, while that of the data is $\sum N_k$. Let us define the data vector of the k th well log as follows:

$$d^{(k)} = \{d_1^{(k)}, \dots, d_{N_k}^{(k)}\}^T. \quad (14)$$

The k th data in the j th depth is calculated by

$$\varphi_j^{(k)} = \varphi^{(k)}(z_j) = g^{(k)}(B_1^{(1)}, \dots, B_{Q_1}^{(1)}, \dots, B_1^{(P)}, \dots, B_{Q_P}^{(P)}, z_j) \quad (15)$$

and the vector of calculated data is as follows:

$$\boldsymbol{\varphi}^{(k)} = \{\varphi_1^{(k)}, \dots, \varphi_j^{(k)}, \dots, \varphi_{N_j}^{(k)}\}^T. \quad (16)$$

The data vector of the joint inversion problem including S number of well logs is as follows:

$$\mathbf{d} = \{d_1^{(1)}, \dots, d_{N_1}^{(1)}, d_1^{(2)}, \dots, d_{N_2}^{(2)}, \dots, d_1^{(S)}, \dots, d_{N_S}^{(S)}\}^T, \quad (17)$$

and the vector of all calculated data analogously is as follows:

$$\boldsymbol{\varphi} = \{\boldsymbol{\varphi}^{(1)}, \dots, \boldsymbol{\varphi}^{(S)}\}^T. \quad (18)$$

The series expansion coefficients in Eq. (13) represent the unknowns of the joint inversion problem, thus the combined parameter vector is as follows:

$$\mathbf{m} = \{B_1^{(1)}, \dots, B_{Q_1}^{(1)}, \dots, B_1^{(P)}, \dots, B_{Q_P}^{(P)}\}^T, \quad (19)$$

with which the forward problem can be written as follows:

$$\boldsymbol{\varpi} = \mathbf{g}(\mathbf{m}) = \left\{ g_1^{(1)}(\mathbf{m}), \dots, g_{N_1}^{(1)}(\mathbf{m}), \dots, g_1^{(S)}(\mathbf{m}), \dots, g_{N_s}^{(S)}(\mathbf{m}) \right\}^T. \quad (20)$$

The measurement data are always contaminated with some amount of noise. On the other hand, the data calculated by Eq. (20) contain modeling errors resulting from discretization and other physical simplifications. The overall error between the two quantities is defined as follows:

$$\mathbf{e} = \mathbf{d} - \mathbf{g}(\mathbf{m}), \quad (21)$$

which is not a zero vector. Inversion methods find a solution at the minimum of some norm of deviation vector \mathbf{e} . The response functions are usually nonlinear, thence the inverse problem can be solved by either global or linearized inversion methods.

In case of using linearized inversion, the starting point \mathbf{m}_0 in model space is considered not too remote from the solution \mathbf{m} . The vector of parameter corrections in Eq. (10) is sought. The calculated data can be approximated by Eq. (5). The k th element of the deviation vector defined in Eq. (21) is as follows:

$$e_k = d_k - \varphi_k(o) - \sum_{i=1}^P \left(\frac{\partial g_k}{\partial m_i} \right)_{m_o} \delta m_i, \quad (22)$$

where $\varphi_k(o) = g_k(\mathbf{m}_0)$. By introducing the notation $\delta d_k = d_k - \varphi_k(o)$ and $G_{ki} = \left(\frac{\partial g_k}{\partial m_i} \right)_{m_o}$, the previous vector is as follows:

$$\mathbf{e} = \delta \mathbf{d} - \mathbf{G} \delta \mathbf{m}. \quad (23)$$

The determination of $\delta \mathbf{m}$ ensures to reach a closer point to the solution. The optimal estimate can be extracted by an iterative method. The correction of the actual model in the l th step is as follows:

$$\mathbf{m}^{(l)} = \mathbf{m}^{(l-1)} + \delta \mathbf{m}^{(l)}, \quad \delta d_k = d_k - \varphi_k(\mathbf{m}^{(l-1)}), \quad G_{ki} = \left(\frac{\partial g_k}{\partial m_i} \right)_{m^{(l-1)}}. \quad (24)$$

If the solution is bound to the minimum of the L_p norm of the deviation vector given in Eq. (23)

$$E = \sum_{k=1}^N \left| \delta d_k - \sum_{i=1}^P G_{ki} \delta m_i \right|^p, \quad (25)$$

then the undermentioned set of conditions

$$\frac{\partial L_p}{\partial m_h} = 0 \quad (h = 1, 2, \dots, P) \quad (26)$$

must be fulfilled. As a result of derivation, a nonlinear set of equations is obtained

$$\delta \mathbf{m} = (\mathbf{G}^T \mathbf{W} \mathbf{G})^{-1} \mathbf{G}^T \mathbf{W} \delta \mathbf{d}, \quad (27)$$

which can be solved by the iteratively reweighted least squares (IRLS) method that re-calculates the diagonal elements $W_{ks} = |e_k|^{p-2} \delta_{ks}$ of the weighting matrix \mathbf{W} in each iteration step.

3.1 Basis Functions Used in Interval Inversion

The selection of basis functions is not strictly limited, but the finding of suitable ones may greatly improve the accuracy and reliability of the inversion result. In case of proper basis functions, a relatively small number of additive terms are enough to be used, because the effect of truncation in Eq. (13) is negligibly small. One should tend to reduce the number of expansion coefficients to maintain the numerical stability of the inversion procedure.

In geophysical inversion, there are several applications of using simple models. In borehole geophysics, the layerwise homogeneous model is of high importance. This situation can be described easily by substituting a combination of Heaviside basis functions into Eq. (13)

$$m_i(z) = \sum_{q=1}^{Q_i} B_q^{(i)} \psi_q(z) = \sum_{q=1}^{Q_i} B_q^{(i)} [u(z - Z_{q-1}) - u(z - Z_q)], \quad (28)$$

where Z_q is the depth coordinate of the q th layer and Q is the number of homogeneous layers. The basis function ψ_q introduced in Eq. (28) is always zero except in the q th layer, which is an element of an orthogonal sequence of functions

$$\int_0^{\tilde{z}_{\max}} \Psi_q \Psi_{q'} dz = \begin{cases} 0, & \text{if } q \neq q' \\ Z_q - Z_{q-1}, & \text{if } q = q' \end{cases} \quad (29)$$

It arises that the series expansion coefficient in the q th layer equals to the petrophysical parameter in the same layer, that is, $B_q^{(i)} = m_i(Z_{q-1} < z < Z_q)$. The series in Eq. (13) can be rewritten as follows:

$$m_i(z) = \sum_{q=1}^{Q_i} m_q^{(i)} \Psi_q(z), \quad (30)$$

where $m_q^{(i)}$ is the value of the i th parameter in the q th layer. It is obvious that the inverse problem can be solved by the smallest possible number of unknowns. On the other hand, the layer-boundaries appear in the argument of the basis function Ψ_q , which can be extracted by the interval inversion method.

The variation of petrophysical parameters within the layer can be approximated by polynomial series expansion

$$m_i(z) = \sum_{q=1}^{Q_i} B_q^{(i)} P_q(z), \quad (31)$$

where $P_q(z)$ represents some polynomial, for instance Legendre polynomials. The meaning of expansion is not demonstrative than in Eq. (30). Similarly, the selection of parameter Q is less unequivocal. The number of unknowns can be much higher than in the homogeneous case. A trade-off must be taken between the vertical resolution of model parameters and the stability of the inversion procedure. To relieve the task, in some practical cases the combination of Eqs. (30) and (31) is used to get an adequate solution. If the layerwise homogeneous model contains an inhomogeneous layer, the following series expansion can be used

$$m_i(z) = \sum_{\substack{q=1 \\ q \neq q'}}^{Q_i} m_q^{(i)} \Psi_q(z) + \sum_{u=1}^U B_u P_u(z - Z_{q'-1}), \quad (32)$$

where U is the number of additive terms used in the approximation of variation in the q' th layer. The above problem was solved by Dobróka and Szabó (2005) using a combined inversion algorithm based on the subsequent use of global and linearized optimization methods.

3.2 Layer-Thickness Determination by Interval Inversion

The greatly over-determined interval inversion method allows to treat increasing number of inversion unknowns without significant decrease of accuracy in parameter estimation. Some groups of the inversion unknowns are contained in local response functions (e.g., zone parameters); some are not included (layer thicknesses). The latter can be determined by the interval inversion of the combined data set defined in Eq. (17). Consider the model vector of the inverse problem including the layer boundary coordinates

$$\mathbf{m} = \left\{ B_1^{(1)}, \dots, B_{Q_1}^{(1)}, \dots, B_1^{(P)}, \dots, B_{Q_P}^{(P)}, Z_1, \dots, Z_R \right\}^T. \quad (33)$$

The number of unknowns is $R + \sum Q_i$ that are substituted into Eq. (12) to calculate theoretical well logs in the forward problem. The number of data is specified in Eqs. (17) and (18). The connection between model and data $\boldsymbol{\sigma} = \mathbf{g}(\mathbf{m})$ contains the layer boundary coordinates; therefore, Eq. (5) modifies as follows:

$$\varphi_k = g_k(\mathbf{m}_0) + \sum_{i=1}^P \left(\frac{\partial g_k}{\partial m_i} \right)_{m_0} \delta m_i + \sum_{r=1}^R \left(\frac{\partial g_k}{\partial Z_r} \right)_{m_0} \delta Z_r, \quad (34)$$

where $\delta Z_r = Z_r - Z_r^{(0)}$ is an element of the model correction vector $\delta \mathbf{m}$. In Eq. (23), the following Jacobi's matrix is used

$$G_{ki} = \begin{cases} \left(\frac{\partial g_k}{\partial m_i} \right)_{m_0}, & \text{if } i = 1, 2, \dots, P \\ \left(\frac{\partial g_k}{\partial Z_i} \right)_{m_0}, & \text{if } i = P + 1, \dots, P + R \end{cases}. \quad (35)$$

The minimization of the L_p norm of the deviation vector leads to the solution of the inverse problem given by Eq. (27). The iterative method gives an estimate for the series expansion coefficients and layer boundary coordinates. The determination of layer-boundaries can be made easily by using a series expansion based on Eq. (28). However, if it is required, the method can be combined with the scheme of polynomial discretization. The estimation of layer-boundaries can be performed most efficiently by using a global optimization method.

4 Global Inversion by Simulated Annealing Method

The performance of inversion methods highly depends on how successfully the optimum of the objective function defined in Eq. (25) is found. Conventional interpretation systems offer linear optimization tools that give quick and satisfactory results in case of having a suitable initial model. The weakness of these gradient-based searching methods is that they tend to find a solution at a local optimum of the objective function. This problem can be avoided by using a global optimization method, which finds the absolute optimum of the same function. There is another typical problem of linear interval inversion. In case of linear optimization, the partial derivatives with respect to depth in the Jacobi's matrix can only be determined in a rough approximation, because the difference quotient with a depth difference being equal to the distance between two measuring points. This may lead to a numerically instable inversion procedure. Global optimization does not require the computation of derivatives. For the solution of global inverse problems,

artificial intelligence tools can be used effectively. Szabó and Dobróka (2013) published previously a float-encoded genetic algorithm-based interval inversion algorithm for oil field and hydrogeological applications. In this study, a fast simulated annealing (SA) method is suggested for the determination of layer parameters and formation thicknesses.

The conventional SA method was developed by Metropolis et al. (1953). In metallurgy, the removal of work-hardening is realized by a slow cooling manipulation from the temperature of liquid alloy state. This process reduces progressively the kinetic energy of a large number of atoms with high thermal mobility, which is followed by the starting of crystallization. Theoretically, the perfect crystal with minimal overall atomic energy can be produced by an infinitely slow cooling schedule. This is analogous with the stabilization of the inversion procedure at the global optimum of the objective function. A quick cooling process causes grating defects and the solid freezes in imperfect grid at a relatively higher energy state. It is similar to the trapping of the inversion procedure in a local minimum. However, the atoms may escape from the high-energy state owing to a special process called annealing to achieve the optimal crystal grating by a slower cooling process. The SA algorithm employs this technology to search the global optimum of the objective (in the terminology energy) function such as E defined in Eq. (25). At first, the components of the model vector defined in Eq. (33) are modified properly. The modification of the i th model parameter in the l th iteration step is as follows:

$$m_i^{(l+1)} = m_i^{(l)} + b, \quad (36)$$

where $b < b_{\max}$ is a perturbation term (b_{\max} is decreased appropriately as the iteration procedure progresses). During the random seeking, the energy function $E(\mathbf{m})$ is calculated and compared with the previous one in every iteration step. The acceptance probability of the new model depends on the Metropolis criteria

$$P(\Delta E, T) = \begin{cases} 1, & \text{if } \Delta E \leq 0 \\ e^{-\Delta E/T}, & \text{otherwise} \end{cases}, \quad (37)$$

where the model is always accepted when the value of energy function is lower in the new state than that of the previous one. If the energy of the new model increased, there is also some probability of acceptance depending on the values of energy E and control temperature T . If $P(\Delta E) \geq \alpha$ fulfills, the new model is accepted, else it is rejected (α is a random number generated with uniform probability from the interval of 0 and 1). These criteria assure the escape from the local minima. Geman and Geman (1984) proved that the following cooling schedule is the necessary condition to find the global optimum

$$T(l) = \frac{T_0}{\ln(l)} \quad (l > 1), \quad (38)$$

where T_0 is a properly chosen initial temperature. The SA algorithm is very effective, but the logarithmic reduction of temperature in Eq. (38) is rather time-consuming. Several attempts were made to shorten the CPU time. Ingber (1989) proposed a modified SA algorithm called very fast simulated re-annealing (VFSA). Consider different ranges of variation for each model parameter

$$m_i^{(\min)} \leq m_i^{(l)} \leq m_i^{(\max)}. \quad (39)$$

The perturbation of the i th model parameter at iteration $(l + 1)$ is as follows:

$$m_i^{(l+1)} = m_i^{(l)} + y_i \left(m_i^{(\max)} - m_i^{(\min)} \right), \quad (40)$$

where y_i is a random number between -1 and 1 generated from a specified non-uniform probability distribution function. The global optimum is guaranteed when the decrease of the i th individual temperature follows

$$T_i^{(l)} = T_{0,i} e^{(-c_i \sqrt{l})} \quad (41)$$

Equation (41) specifies different temperature to each model parameter, where $T_{0,i}$ is the initial temperature of the i th model parameter, c_i is the i th control parameter, and P is the number of model parameters. The acceptance rule of the VFSA algorithm is the same as that used in Metropolis SA method, but the exponential cooling schedule assures much faster convergence to the global optimum than the logarithmic one suggested in Eq. (38).

5 Selection of Initial Model by Cluster Analysis

Multivariate statistical methods such as regression, factor, and cluster analyses help to find similarities between petrophysical properties of rocks, reduce problem dimensionality or explore non-measurable (latent) information from the observations, and arrange data into groups to reveal different lithological characteristics of the investigated formations. This a priori information can be useful in petrophysical modeling, facies, or trend analysis and in geophysical inversion to set an initial model as input for the inversion procedure. Clustering methods are applicable to sort data into groups in such a way that the S dimensional objects specified by well logs measured from given depths are more similar than other ones observed from different depths. From the point of the interval inversion method, it is of great importance that objects connected to the same cluster define approximately the same lithological character, while other clusters represent dissimilar ones.

Agglomerative cluster analysis builds a hierarchy from the observations by progressively merging clusters. At the beginning, we have as many clusters as individual elements. In the first step, the closest points are coupled together to form

a new cluster. In each following step, the distances between objects are re-calculated and the procedure is continued until all elements are grouped into one cluster. Several distance definitions can be used as a measure of dissimilarity between the pairs of observed objects such as Euclidean (L_2 norm based), Manhattan (L_1 norm based), or Mahalanobis (sample covariance based). During the procedure, the distances between the elements of the same group are minimized, while they are maximized between the clusters simultaneously. For the reconnection of clusters, the Ward's linkage criterion is followed that minimizes the deviances of $(x_i - C)$, where x_i is the i th object and C is the centroid (average of elements) of the given cluster (Ward 1963). The result of cluster analysis is a dendrogram that shows the hierarchy of clusters and the connections between them at different distances.

In this study, cluster analysis is used as a preliminary data processing step before inverse modeling. It is shown in Fig. 1 that clustering makes use of the complete wellbore data set originated from the entire logging interval. By finding the similarities between the well logs, the objects are grouped into clusters. The log of clusters correlates well with the lithology variation along a borehole. The change in the group number of clusters appearing on the log gives the positions of layer-boundaries, which can be read automatically by computer processing. The estimated layer boundary coordinates as important a priori information for constructing the initial model serve as input for the interval inversion procedure. In an earlier study, the layer-boundaries extracted by cluster analysis were fixed during the interval inversion procedure (Szabó et al. 2013). However, similar to the layer parameters, the layer boundary coordinates may be treated as unknowns in the interval inversion procedure.

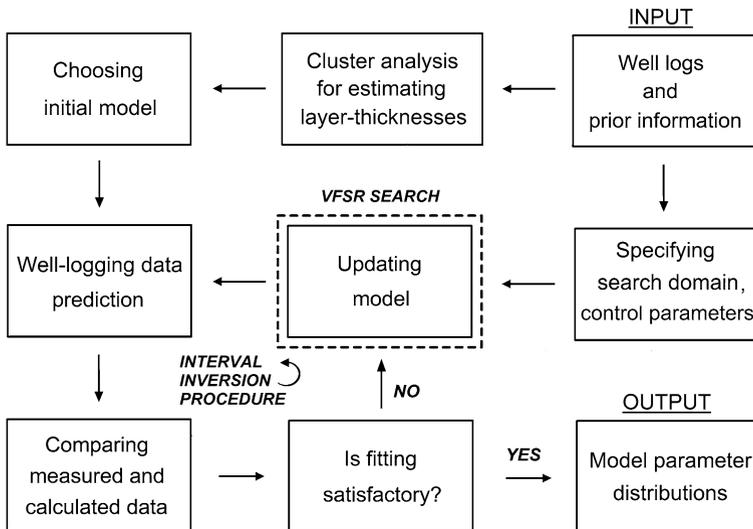


Fig. 1 Scheme of cluster analysis-assisted global inversion procedure

6 Oil field Application

6.1 Results of Cluster Analysis

In a Hungarian hydrocarbon borehole (Well No. 1), nine well logs were used for testing the interval inversion method. The following log types formed the input of clustering such as caliper (CAL), compensated neutron porosity (CN), gamma-gamma density (DEN), acoustic (primary wave) interval time (AT), natural gamma-ray intensity (GR), deep resistivity (RD), microlaterolog resistivity (RMLL), shallow resistivity (RS), and spontaneous potential (SP). In the first step of the procedure, hierarchical cluster analysis was applied to find a proper initial model for inversion processing (Fig. 1). In the processed interval, a sedimentary complex made up of seven unconsolidated shaly sandy beds were deposited. Three lithological categories were specified, namely shale, shaly sand, and sand. At this stage of interpretation, this lithological resolution was enough for finding the layer-boundaries, because the relative volumes of rock matrix and shale could be estimated in the inversion phase. The standardized Euclidean distance evaluating each datum in the sum of squares inversely weighted by the sample variance was used for measuring the distance between data objects. A hierarchical cluster tree was created by using the Ward's linkage algorithm (Fig. 2a). In Fig. 2b, the ordinal numbers of leaf nodes can be seen that were assigned to each object. Since some leaf nodes corresponded to multiple objects, the total number of nodes was 30. Three clusters can be separated if the tree is cut at centroid distance 1. The layer-boundaries can be traced out in the well log of cluster numbers. According to traditional interpretation, the inflection points of high amplitudes in the GR log indicate rock interfaces. The layer boundary coordinates can be well approximated by the steps between the clusters of sand and shale, i.e., cluster 2 and cluster 3. The depth coordinates indicated by black arrows in Fig. 2c were chosen as initial model parameters for the subsequent interval inversion procedure.

The combination of well logs usually gives useful information on lithology, petrophysical, and zone parameters. In Fig. 3, the three-dimensional cross-plots of clustered well-logging data can be seen which specify several site-specific constants for calculating data in forward modeling. These constants can be used directly in the probe response functions. For instance in Fig. 3a, the neutron porosity of sand (13 %) and shale (25 %) and the natural gamma-ray intensity of sand (45 API) and shale (140 API) can be chosen for the given hydrocarbon zone. Several observed data types show strong correlation with each other. In Fig. 3c, d, the nonlinear connection between natural gamma-ray intensity and resistivity is eye-catching. The detailed list of correlation relationships between the data is contained by the correlation matrix including the Pearson's correlation coefficients (Table 1). The highest correlations are between lithology logs (GR and CAL, SP and GR) and saturation logs (RS and RD). Porosity-sensitive logs also show strong correlations with lithology logs (DEN and CAL, SP and CAL, SP and CN, GR and DEN, CN and GR). The negative elements of the correlation matrix show inverse proportionality between the data variables (GR and SP).

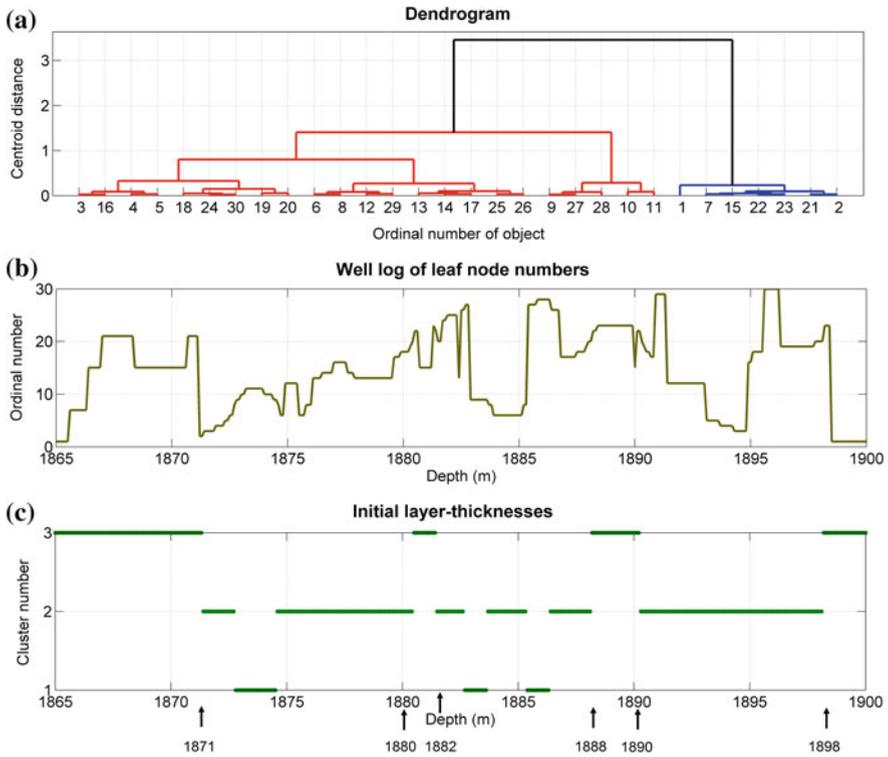


Fig. 2 Results of cluster analysis in Well No. 1. **a** Dendrogram. **b** Well log of leaf node numbers. **c** Initial layer-thicknesses

6.2 Forward Problem of Well-Logging

The mathematical relationships between the petrophysical properties and well-logging data are dominantly empirical. In the case study, the parameters of the initial model are effective porosity (POR), shale volume (VSH), water saturation of the invaded zone flooded by drilling mud (SX0), water saturation of the virgin zone occupied by original pore fluid (SW), and sand volume (VSD = 1 – POR – VSH). These parameters and other derived ones underlie the calculation of hydrocarbon reserves. The following set of response functions was used to approximate observable data

$$DENTH = POR[(SX0 \cdot DEMF) + (1 - SX0)DEHC] + VSH \cdot DESH + VSD \cdot DESD, \tag{42}$$

$$GRTH = GRSD + \frac{1}{DENTH} \left(\frac{VSH \cdot GRSH \cdot DESH}{+ VSD \cdot GRSD \cdot DESD} \right), \tag{43}$$

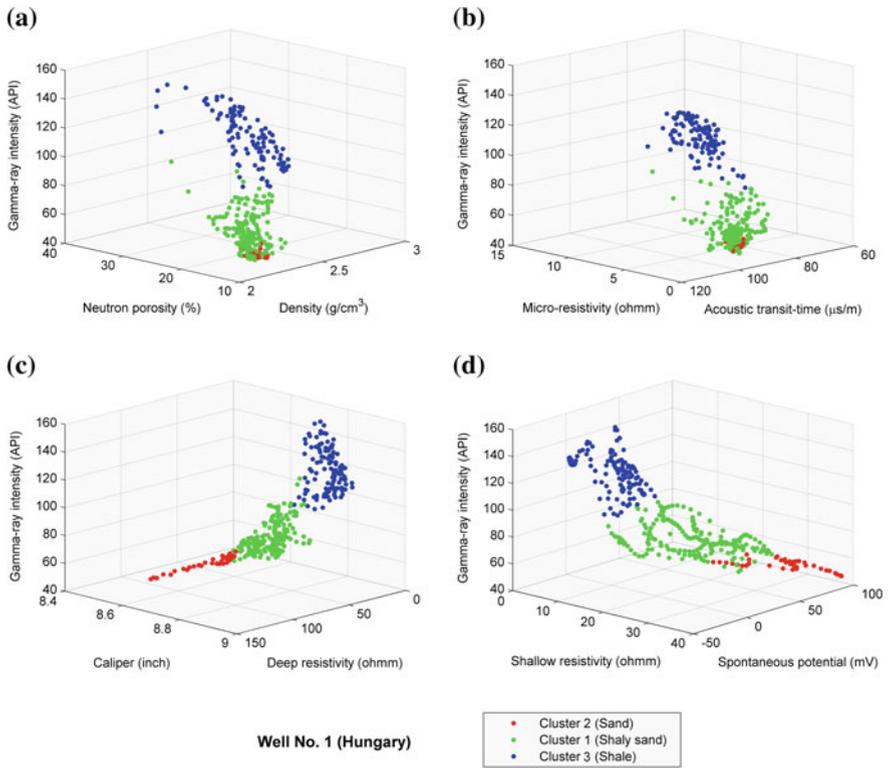


Fig. 3 Clustered data represented in the form of cross-plots in Well No. 1

Table 1 Correlation matrix of well logs measured in Well No. 1

	CAL	CN	DEN	AT	GR	RD	RMLL	RS	SP
CAL	1.0	0.67	0.89	-0.34	0.83	-0.53	0.82	-0.52	-0.74
CN	0.67	1.0	0.68	0.16	0.88	-0.59	0.47	-0.56	-0.76
DEN	0.89	0.68	1.0	-0.51	0.84	-0.58	0.82	-0.59	-0.73
AT	-0.34	0.16	-0.51	1.0	-0.01	0.16	-0.51	0.19	0.10
GR	0.83	0.88	0.84	-0.01	1.0	-0.59	0.68	-0.61	-0.81
RD	-0.53	-0.59	-0.58	0.16	-0.59	1.0	-0.41	0.92	0.68
RMLL	0.82	0.47	0.82	-0.51	0.68	-0.41	1.0	-0.41	-0.59
RS	-0.52	-0.56	-0.59	0.19	-0.61	0.92	-0.41	1.0	0.71
SP	-0.74	-0.76	-0.73	0.10	-0.81	0.68	-0.59	0.71	1.0

$$\begin{aligned} \text{CNTH} = & \text{POR}[(\text{SX0} \cdot \text{CNMF}) + (1 - \text{SX0})\text{CNHC}] \\ & + \text{VSH} \cdot \text{CNSH} + \text{VSD} \cdot \text{CNSD}, \end{aligned} \quad (44)$$

$$\begin{aligned} \text{ATTH} = & \text{POR}[(\text{SX0} \cdot \text{ATMF}) + (1 - \text{SX0})\text{ATHC}] \\ & + \text{VSH} \cdot \text{ATSH} + \text{VSD} \cdot \text{ATSD}, \end{aligned} \quad (45)$$

$$\frac{1}{\sqrt{\text{RDTH}}} = \left[\frac{\text{VSH}^{(1-\text{VSH}/2)}}{\sqrt{\text{RSH}}} + \frac{(\sqrt{\text{POR}})^{\text{BM}}}{\sqrt{\text{BA} \cdot \text{RW}}} \right] (\sqrt{\text{SW}})^{\text{BN}}, \quad (46)$$

$$\frac{1}{\sqrt{\text{RSTH}}} = \left[\frac{\text{VSH}^{(1-\text{VSH}/2)}}{\sqrt{\text{RSH}}} + \frac{(\sqrt{\text{POR}})^{\text{BM}}}{\sqrt{\text{BA} \cdot \text{RMF}}} \right] (\sqrt{\text{SX0}})^{\text{BN}}. \quad (47)$$

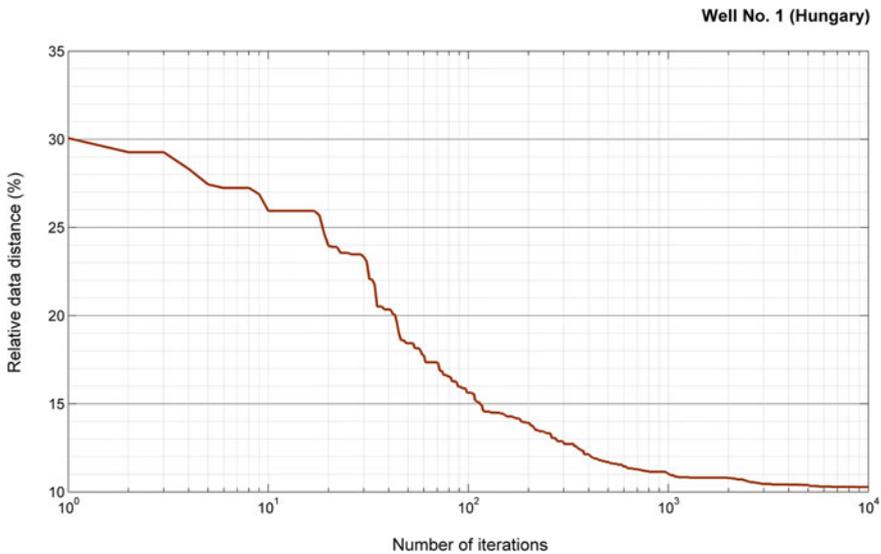
On the left side of Eqs. (42)–(47), the theoretical (TH) values of the well-logging data stand, while on the right, the layer parameters (POR, VSH, SX0, SW, and VSD) and zone parameters can be found. The latter represent the physical properties of mud filtrate (MF), water (W), hydrocarbon (HC), shale (SH), and sand (SD). They are treated as unvarying quantities known from cluster analysis (Fig. 3) or other a priori information (laboratory and well-site reports). The textural constants, such as cementation exponent (BM), saturation exponent (BN), and tortuosity factor (BA), can be estimated from the literature, laboratory data, or the interval inversion method (Dobróka and Szabó 2011). In complex reservoirs, the rock matrix may be composed of several mineral components. Depending on the interpretation problem, the relative volumes of rock constituents can also be extracted within the interval inversion procedure (Dobróka et al. 2012). By using the response Eqs. (42)–(47), an estimate for the model (layer) parameters can be given by the inversion procedure.

6.3 Results of Interval Inversion

The interval inversion procedure was performed on suitable well logs (CN, DEN, AT, GR, RD, and RS) of Well No. 1. Based on the results of cluster analysis, the following depth coordinates were chosen as initial model parameters for interval inversion: 1871, 1880, 1882, 1888, 1890, and 1898 m. The last (seventh) coordinate was the depth at the bottom of the logging interval. As a result, seven shaly sandy layers were traced out. Within three permeable intervals, the separation between DEN and CN logs confirmed the presence of hydrocarbons. In the model approximation, constant layer parameters (POR, VSH, SX0, and SW) were assumed (VSD was calculated deterministically in every iteration steps), which represented a high over-determination ratio (62 with 2100 data, 28 layer parameters, and 6 boundary coordinates) and very stable inversion procedure.

The unknowns of the inverse problem were the series expansion coefficients given in Eq. (28) and the layer boundary coordinates. The VFSSR algorithm was used to give a quick estimate to the global optimum of the L_2 -norm-based energy function. The maximal number of iteration steps was set to 10,000. The logarithmic cooling process based on Eq. (38) was applied with an initial temperature of 0.01. The lower and upper limits of layer parameters were 0 and 1, respectively. The layer-boundaries were allowed to vary within 0 and 10 m. The rate of convergence of the inversion procedure was smooth and progressive as it is seen in Fig. 4, where the root mean squared errors of the relative differences between the measured and calculated data were plotted. The data distance of the final result is influenced by the data noise and the model approximation. In Fig. 5, the change of layer thicknesses calculated from the boundary coordinates can be followed. Until the 7000th iteration step, all thicknesses had attained to their optima. Then, only the layer parameters showed considerable variation.

In the depth scale of Fig. 6, the boundary coordinates estimated by the interval inversion procedure are marked. The method distinguished the permeable and non-permeable intervals within the hydrocarbon zone and gave a proper estimate to rock interfaces as they correlate well with the layer-boundaries inferred from the GR log. The well logs of estimated layer parameters are in tracks 6–8. The pore space was divided into two separate parts filled with salty water and hydrocarbons. The movable and irreducible hydrocarbon saturation were derived from the inversion results by basic equations (Movable HC = $SX_0 - SW$, Irreducible HC = $1 - SX_0$). The hydrocarbon reserves are estimated from the movable hydrocarbon saturation, porosity, and the total volume of the reservoir rock. The latter can be estimated



from ground geophysical surveys (e.g., seismic) and multi-borehole data. A model reliable estimate of oil and gas reserves is supported by the results of the interval inversion procedure. Szabó and Dobróka (2013) made a comparison between the local and interval inversion methods. It was shown that the estimation error of porosity and water saturation can be reduced significantly by the interval inversion method. This improvement bears influence on the calculation of hydrocarbon reserves. The unit volume of rock was composed of porosity, shale content, and sand volume in different proportions along the interval (track 7). The absolute permeability shown in track 8 for each formation was calculated by the knowledge of porosity and bound water saturation (chosen as 0.1) in the hydrocarbon reservoirs (Timur 1968).

7 Conclusions

Advanced data processing methods are essential to extract reliable petrophysical information from geophysical data sets. An intensive research of inversion methods is being made worldwide in all fields of geophysics. In oilfield applications, the proper interpretation of in situ borehole logging data is especially important, because these methods lay the foundations to hydrocarbon reserve calculations. In this chapter, a new inversion methodology was presented, which is now fully automatized by giving an estimate to petrophysical parameters and layer-boundaries in a joint inversion procedure. The cluster analysis-assisted interval inversion method assures a greatly over-determined inverse problem to give accurate and reliable solution along the entire borehole. The specialty of the method is that the basis functions of series expansion can be chosen arbitrarily. The optimal set of basis functions to be in use depends on the variation of lithology and pore fluid types along a borehole. In this study, a layerwise homogeneous model was chosen to reduce the number of inversion unknowns as far as possible. This approach keeps the numerical stability of the inversion procedure in view. However, there is nothing to prevent from the improvement of the vertical resolution of the interval inversion method, but it goes with the relative increase of the number of series expansion coefficients. As the problem is highly over-determined, it can be allowed to some extent. Practically, a trade-off must be taken between the number of unknowns (resolution) and stability of the inversion (unique solution) procedure as they are inversely proportional. An adequate solution is to choose orthogonal basis functions such as Legendre polynomials in the development of series, in which case the correlation between the estimated model parameters and the parameter estimation errors is relatively the lowest. It is suggested to apply preliminary cluster analysis to find lithological similarities in the data set, which separates such intervals where the polynomial discretization can be performed most effectively. This technical solution leads to the reduction of data and model distances. The inverse problem also affected by the quality of forward problem solution. In the probe response equations, there are several zone parameters that could be chosen

properly for the given well-site. An objective solution can be given by the interval inversion procedure as the zone parameters most sensitive to data can be treated as inversion unknowns. Another strength of the method is the possibility to extend the inverse modeling to multi-borehole applications by expanding the model parameters into series of bivariate basis functions. All of the above properties confirm the feasibility of the interval inversion methods and the use of global optimization techniques preferably very fast simulated re-annealing and float-encoded genetic algorithm in petroleum geoscience applications.

Acknowledgments The first author as the leading researcher of Project No. K 109441 thanks to the support of the Hungarian Scientific Research Fund. The second author as the leading researcher of Project No. PD 109408 thanks to the support of the Hungarian Scientific Research Fund. The second author also thanks to the support of the János Bolyai Research Fellowship of the Hungarian Academy of Sciences. The authors thank the Hungarian Oil and Gas Company's (MOL) contribution to the research work and the long-term cooperation. The authors also thank Hajnalka Szegedi for improving the manuscript formally.

References

- Alberty M, Hashmy KH (1984) Application of ULTRA to log analysis. SPWLA 25th annual logging symposium, New Orleans, Paper Z, 10–13 June 1984
- Ball SM, Chace DM, Fertl WH (1987) The well data system (WDS): an advanced formation evaluation concept in a microcomputer environment. In: Proceedings of SPE eastern regional meeting, Paper 17034, pp 61–85
- Dobróka M (1995) The introduction of joint inversion algorithms into well log interpretation (in Hungarian). Scientific report, Geophysical Department, University of Miskolc
- Dobróka M, Szabó NP (2005) Combined global/linear inversion of well-logging data in layer-wise homogeneous and inhomogeneous media. *Acta Geod Geophys Hung* 40:203–214
- Dobróka M, Szabó NP (2011) Interval inversion of well-logging data for objective determination of textural parameters. *Acta Geophys* 59(5):907–934
- 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 Petrol Sci Eng* 86–87:144–152
- Dobróka M, Szabó NP, Turai E (2012) Interval inversion of borehole data for petrophysical characterization of complex reservoirs. *Acta Geod Geophys Hung* 47(2):172–184
- Geman S, Geman D (1984) Stochastic relaxation, Gibbs' distribution and Bayesian restoration of images. *IEEE Trans PAMI* 6:721–741
- Ingber L (1989) Very fast simulated reannealing. *Math Comput Model* 12(8):967–993
- Mayer C, Sibbit A (1980) GLOBAL, a new approach to computer-processed log interpretation. In: Proceedings of SPE annual fall technical conference and exhibition, Paper 9341, pp 1–14
- Metropolis N, Rosenbluth A, Rosenbluth M, Teller A, Teller E (1953) Equation of state calculations by fast computing machines. *J Chem Phys* 21:1087–1092
- Serra O (1984) *Fundamentals of well-log interpretation*. Elsevier, Amsterdam
- Szabó NP, Dobróka M (2013) Float-encoded genetic algorithm used for the inversion processing of well-logging data. In: Michalski A (ed) *Global optimization: theory, developments and applications*. Mathematics research developments, computational mathematics and analysis series. Nova Science Publishers Inc., New York

- Szabó NP, Dobróka M, Kavanda R (2013) Cluster analysis assisted float-encoded genetic algorithm for a more automated characterization of hydrocarbon reservoirs. *Intell Control Autom* 4(4):362–370
- Timur A (1968) An investigation of permeability, porosity and residual water saturation relationships for sandstone reservoirs. *Log Analyst* 9:8–17
- Ward JH (1963) Hierarchical grouping to optimize an objective function. *J Am Stat Assoc* 58(301):236–244