Geosciences and engineering : a publication of the University of Miskolc

5. évf. 8. sz. / 2016

Asfahani, Jamal - Tlas, Muhammad Interpretation of Self-Potential Anomalies by Developing an Approach Based on Linear Optimization	7-21
Balogh Gergely Pál Interval Inversion of Engineering Geophysical Sounding Logs	22-31
Bibor István - Szabó Norbert Péter Unconventional Shale Characterization Using Improved Well Logging Methods	32-50
Braun Bence Á. - Abordán Armand - Szabó Norbert Péter Lithology Determination in a Coal Exploration Drillhole Using Steiner Weighted Cluster Analysis	51-64
Czirok Lili Analysis of Stress Relations Using Focal Mechanism Solutions in the Pannonian Basin	65-84
Galsa Attila - Drahos Dezső Statistical Evaluation of Laterolog Resistivity Measurements	85-97
Kale, Utku - Dobróka Mihály An Introduction to Robust Tomography Methods	98-110
Kiss Anett - Dobróka Mihály - Turai Endre - Somogyiné Molnár Judit Laboratory Induced Polarization Data Processed with Series Expansion Inversion	111-123
Nuamah, Daniel Oduro Boatey Heavy Metal Source Identification and Analysis Using Multivariate Statistical Methods in Soils from Akuse Area, South-Eastern Ghana	124-134
Somogyiné Molnár Judit Development of New Complex Software for Investigating Acoustic Velocities under Pressure	135-146
Tóth Gyula - Völgyesi Lajos Data Processing of Qdaedalus Measurements	147-164

Geosciences and Engineering, Vol. 5, No. 8 (2016), pp. 7–21.

INTERPRETATION OF SELF-POTENTIAL ANOMALIES BY DEVELOPING AN APPROACH BASED ON LINEAR OPTIMIZATION

JAMAL ASFAHANI*–MUHAMMAD TLAS*

**Atomic Energy Commission, P. O. Box 6091, Damascus, Syria*
E-mail: cscientific@aec.org.sy (J Asfahani)

Abstract: A new practical approach is proposed for interpreting self-potential anomalies caused by simple geometric-shaped models such as a sphere, vertical cylinder or horizontal cylinder. The proposed approach is mainly based on both deconvolution and linear optimization techniques to best estimate the model parameters, e.g. the depth to the center of a buried structure, the polarization angle and the amplitude coefficient from self-potential anomaly profile. This method was tested on synthetic data sets corrupted by different Gaussian random noise levels to demonstrate the capability and the reliability of the method. The theoretical modeling results acquired show that the estimated parameter values derived by this proposed method are close to the assumed true values of parameters. The validity of this method is also demonstrated using real field self-potential anomalies taken from the United States and Turkey. A comparable and acceptable agreement is shown between the results derived by this approach and those obtained from the real field data information.

Keywords: *Self-potential anomaly. Sphere-like structure. Cylinder-like structure. Deconvolution technique. Linear optimization*

1. INTRODUCTION

The self-potential (SP) method is among the older methods applied to wide applications in sulfide and graphite exploration and in geophysical groundwater investigations. The quantitative interpretation of SP anomalies is usually carried out by approximating the causative source by simple geometrically shaped models (viz, sheet, cylinder, sphere etc). According to the simplified concept, different interpretation techniques are available in the literature for the quantitative interpretation of SP anomalies. The methods proposed by Yungul [1], Paul [2] and Bhattacharya and Roy [3] use certain characteristic points of the anomaly and hence they turned out to be less reliable in most cases. The curve-matching method proposed by Meiser [4] is cumbersome, especially when there are many parameters to be determined. The methods of least-squares involve a series of trials for minimizing the differences between the observed and the computed self-potential values [5], [6], [7]. The interpretation made by methods based on Fourier and Hilbert transforms [8]–[11] are not straightforward and subject to some inevitable errors in the estimation of the parameters, due to the inaccurate estimation of the horizontal location of

origin. Furthermore, these methods are reliable only for very long profiles. Derivative analysis methods involve higher derivative anomaly [12], [13].

In this paper, a practical interpretation method is introduced to interpret self-potential field anomalies and best estimate of model parameters values, e.g. the depth to the center of the body, the polarization angle, and the amplitude coefficient related to a buried sphere, vertical cylinder or horizontal cylinder-like structure. The method is based on the use of the deconvolution technique to avoid the local minima, where the nonlinear optimization problem, which describes the suitable simple geometric-shaped model of structure, is transformed into a linear optimization problem. This linear optimization problem is thereafter solved by the very well-known algorithm in linear optimization called the simplex algorithm of Dantzig [14], where the global minimum is definitely reached.

The reliability and capability of the proposed interpretation method is demonstrated through analyzing synthetic data sets corrupted by different white Gaussian random noise levels of 10 and 15%. The synthetic model results acquired show that the estimated parameter values derived by this proposed method are very close to the assumed true values of parameters.

The validity of this interpretation method is demonstrated thereafter through interpreting real field self-potential anomalies taken from the United States and Turkey. A comparable and acceptable agreement is obtained between the results derived by this method and those yielded by other interpretation methods. The depth obtained by such a proposed method is found to be in a good accordance with that obtained from real field data information.

2. THEORY

Different theoretical self-potential anomalies related to various geological models such as spheres and horizontal or vertical circular cylinders are treated in this research by the proposed interpretative approach.

2.1 Interpretation of self-potential anomaly due to a horizontal cylinder model

The general expression for the self-potential anomaly (V) at any point $P(x)$ along the x-axis of an arbitrary horizontal cylinder-like structure in a Cartesian coordinate system *(Figure 1)* is given according to [3], [15] as

$$V(x_i) = k \frac{x_i \cos\theta + z \sin\theta}{x_i^2 + z^2} \quad (i = 1,...,N), \tag{1}$$

where z is the depth from the surface to the center of the buried body, θ is the polarization angle between the horizontal and the axis of polarization, k is the amplitude coefficient, and $x_i (i = 1,....,N)$ is the horizontal position coordinate.

The set of *Equation (1)* consists of N nonlinear equations in function of the parameters k , θ , and z . To avoid this nonlinearity, the following proposed deconvolution technique will be used. First and for simplification, V_i is used instead of $V(x_i)$ $(i = 1,...,N)$ in the rest of this paper.

Multiplying the two sides of Eq. (1) by the term $\left(x_i^2 + z^2\right)$ and arranging them, we obtain

$$V_i x_i^2 + V_i q_1 - x_i q_2 - q_3 = 0 \quad (i = 1,...,N), \tag{2}$$

where

$$q_1 = z^2 \tag{3}$$

$$q_2 = k\cos\theta \tag{4}$$

$$q_3 = kz\sin\theta \tag{5}$$

The optimal solution (q_1, q_2, q_3) of the set of linear equations (2) can be found by solving the following nonlinear optimization problem onto the real space R^3

$$\min_{q \in R^3} f(q) = \sum_{i=1}^{N} \left(V_i x_i^2 + V_i q_1 - x_i q_2 - q_3\right)^2 \tag{6}$$

subject to $q_1 \geq 0$ and q_2, q_3 are free. With q_2, q_3 being free, the variables q_1, q_2, q_3 will be changed as

$$\left.\begin{aligned} q_1 &= p_1 \\ q_2 &= p_2 - p_4 \\ q_3 &= p_3 - p_4 \end{aligned}\right\}, \tag{7}$$

where $p_1, p_2, p_3, p_4 \geq 0$. Introducing these new variables into (6), the following nonlinear optimization problem with nonnegative variables is obtained

$$\min_{p \in R^4} \varphi(p) = \sum_{i=1}^{N} \left(V_i x_i^2 + V_i p_1 - x_i p_2 + x_i p_4 - p_3 + p_4\right)^2 \tag{8}$$

subject to $p_1, p_2, p_3, p_4 \geq 0$. The quadratic objective function $\varphi(p)$ of this nonlinear optimization problem (8) is convex onto the nonnegative orthant of the real space R^4. This mathematically means that any solution $\left(p_1, p_2, p_3, p_4\right) \in R^4$ satisfying the following optimality conditions of Karush-Kuhn-Tuker (*KKT*) [14]

$$\left.\begin{array}{l} p_i \geq 0 \quad \forall i = 1,...,4 \\ p_i \dfrac{\partial \varphi(p)}{\partial p_i} = 0 \quad \forall i = 1,...,4 \\ \dfrac{\partial \varphi(p)}{\partial p_i} \geq 0 \quad \forall i = 1,...,4 \end{array}\right\} \tag{9}$$

will be a global minima of the convex nonlinear optimization problem (8). The *KKT* optimality conditions (9) for the nonlinear optimization problem (8) are satisfied through solving the following linear optimization problem

$$\left.\begin{array}{l} \min \sum_{i=1}^{4} u_i \\ \textit{subject to } \dfrac{\partial \varphi(p)}{\partial p_i} - u_i = 0 \quad \forall i = 1,...,4 \\ p_i, u_i \geq 0 \quad \forall i = 1,...,4 \end{array}\right\} \tag{10}$$

or

$$\left.\begin{array}{l} \min \quad u_1 + u_2 + u_3 + u_4 \quad \textit{subject to} \\ \left(\sum_{i=1}^{N} V_i^{\,2}\right) p_1 - \left(\sum_{i=1}^{N} V_i x_i\right) p_2 - \left(\sum_{i=1}^{N} V_i\right) p_3 + \left(\sum_{i=1}^{N} V_i (1+x_i)\right) p_4 - u_1 = -\left(\sum_{i=1}^{N} V_i^{\,2} x_i^{\,2}\right) \\ \left(\sum_{i=1}^{N} V_i x_i\right) p_1 - \left(\sum_{i=1}^{N} x_i^{\,2}\right) p_2 - \left(\sum_{i=1}^{N} x_i\right) p_3 + \left(\sum_{i=1}^{N} x_i (1+x_i)\right) p_4 - u_2 = -\left(\sum_{i=1}^{N} V_i x_i^{\,3}\right) \\ \left(\sum_{i=1}^{N} V_i\right) p_1 - \left(\sum_{i=1}^{N} x_i\right) p_2 - N p_3 + \left(\sum_{i=1}^{N} (1+x_i)\right) p_4 - u_3 = -\left(\sum_{i=1}^{N} V_i x_i^{\,2}\right) \\ \left(\sum_{i=1}^{N} V_i (1+x_i)\right) p_1 - \left(\sum_{i=1}^{N} x_i (1+x_i)\right) p_2 - \left(\sum_{i=1}^{N} (1+x_i)\right) p_3 + \left(\sum_{i=1}^{N} (1+x_i)^2\right) p_4 - u_4 = \\ -\left(\sum_{i=1}^{N} V_i x_i^{\,2} (1+x_i)\right) \\ p_1, p_2, p_3, p_4, u_1, u_2, u_3, u_4 \geq 0 \end{array}\right\} . \tag{11}$$

In *Equation (11)* (p_1, p_2, p_3, p_4) are structural variables, while (u_1, u_2, u_3, u_4) are artificial variables (surplus variables). The artificial variables (u_1, u_2, u_3, u_4) are added only to measure the deviations between the two sides of equations (11) (requirements of the simplex algorithm) and also to define the objective function of the linear optimization problem (11). These variables (u_1, u_2, u_3, u_4) will be dropped when the optimal values of structural variables (p_1, p_2, p_3, p_4) are reached.

The linear optimization problem (11) is thereafter solved by the Simplex algorithm which starts automatically with an initial guess of parameters; zeros for (p_1, p_2, p_3, p_4) and right-hand-side values of *Equation (11)* for (u_1, u_2, u_3, u_4). The Simplex algorithm will find the optimal values of $(p_1, p_2, p_3, p_4, u_1, u_2, u_3, u_4) \in R^8$ which satisfy the *KKT* optimality conditions (9) for the convex nonlinear optimization problem (8) in a bounded number of iterations and consequently the optimal values of $(q_1, q_2, q_3) \in R^3$ by using *Equation 7*.

For more details about the simplex algorithm of Dantzig for linear optimization problems, readers are referred to [14], [16], [17].

Knowing the optimal values of q_1, q_2 and q_3, it will be easy to find the best estimate of the depth from the surface to the center of the buried horizontal cylinder body (z) by using *Equation (3)* as

$$z = \sqrt{q_1}\,. \tag{12}$$

The best estimate of the polarization angle (θ) and the amplitude coefficient (k) can be easily obtained by using simultaneously *Equations (4)* and *(5)* as

$$\theta = \arctan\left(\frac{1}{z}\frac{q_3}{q_2}\right), \tag{13}$$

$$k = \frac{1}{2}\left(\frac{q_2}{\cos\theta} + \frac{q_3}{z\,\sin\theta}\right). \tag{14}$$

2.2. Interpretation of self-potential anomaly due to a vertical cylinder model

The general expression for the self-potential anomaly (V) at any point $P(x)$ along the x-axis of an arbitrary vertical cylinder-like structure in a Cartesian coordinate system *(Figure 1)* is given according to [3] and [15] as

$$V(x_i) = k\,\frac{x_i\cos\theta + z\,\sin\theta}{\left(x_i^2 + z^2\right)^{\frac{1}{2}}} \quad (i = 1,...,N)\cdot \tag{15}$$

The notations in this expression have the same meaning as those presented in *Equation 1*. Following the same manner as performed for the horizontal cylinder model, the parameters related to the self-potential anomaly produced by a vertical cylinder can be estimated by the following equations

$$z = \sqrt{q_1}\,, \tag{16}$$

$$\theta = \arctan\frac{1}{2z}\left(\sqrt{\frac{q_3}{q_2}} + \frac{1}{2}\frac{q_4 - q_5}{q_2}\right), \tag{17}$$

$$k = \mp\sqrt{\frac{1}{3}\left(\frac{q_2}{\cos^2\theta}+\frac{q_3}{z^2\sin^2\theta}+\frac{q_4-q_5}{z\sin 2\theta}\right)}, \tag{18}$$

where q_1, q_2, q_3, q_4 and q_5 are obtained by solving the following linear optimization problem using the simplex algorithm

$$\left.\begin{aligned}
&\min \quad u_1+u_2+u_3+u_4 \quad \textit{subject to}\\
&\left(\sum_{i=1}^{N}V_i^4\right)q_1-\left(\sum_{i=1}^{N}V_i^2x_i^2\right)q_2-\left(\sum_{i=1}^{N}V_i^2\right)q_3-\left(\sum_{i=1}^{N}V_i^2x_i\right)q_4+\left(\sum_{i=1}^{N}V_i^2x_i\right)q_5-u_1=-\left(\sum_{i=1}^{N}V_i^4x_i^2\right)\\
&\left(\sum_{i=1}^{N}V_i^2x_i^2\right)q_1-\left(\sum_{i=1}^{N}x_i^4\right)q_2-\left(\sum_{i=1}^{N}x_i^2\right)q_3-\left(\sum_{i=1}^{N}x_i^3\right)q_4+\left(\sum_{i=1}^{N}x_i^3\right)q_5-u_2=-\left(\sum_{i=1}^{N}V_i^2x_i^4\right)\\
&\left(\sum_{i=1}^{N}V_i^2\right)q_1-\left(\sum_{i=1}^{N}x_i^2\right)q_2-Nq_3-\left(\sum_{i=1}^{N}x_i\right)q_4+\left(\sum_{i=1}^{N}x_i\right)q_5-u_3=-\left(\sum_{i=1}^{N}V_i^2x_i^2\right)\\
&\left(\sum_{i=1}^{N}V_i^2x_i\right)q_1-\left(\sum_{i=1}^{N}x_i^3\right)q_2-\left(\sum_{i=1}^{N}x_i\right)q_3-\left(\sum_{i=1}^{N}x_i^2\right)q_4+\left(\sum_{i=1}^{N}x_i^2\right)q_5-u_4=-\left(\sum_{i=1}^{N}V_i^2x_i^3\right)\\
&q_1,q_2,q_3,q_4,q_5,u_1,u_2,u_3,u_4\geq 0
\end{aligned}\right\}. \tag{19}$$

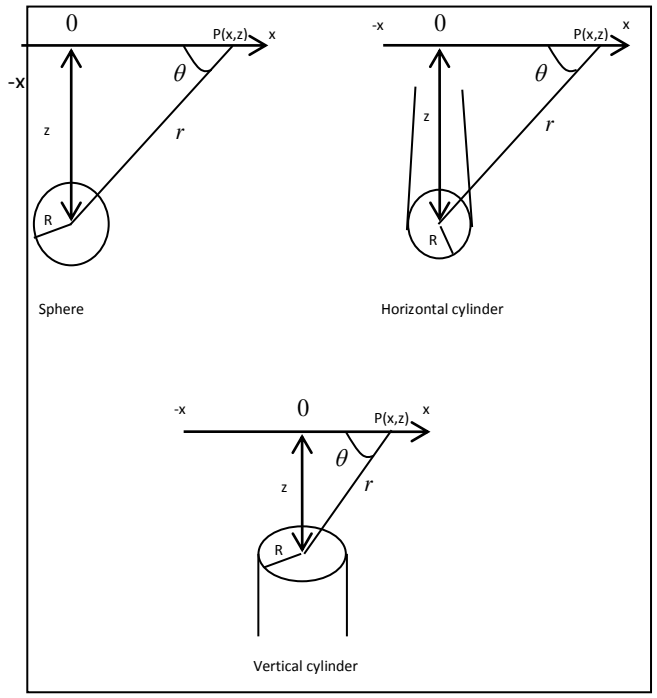

Figure 1. *Diagrams for simple geometrical structures (sphere, horizontal cylinder, and vertical cylinder)*

It is worth noting that the sign of k can be easily assigned by using the statistical criterion of preference called the Root Mean Square Error (RMSE) [18], based on the minimal value between the field data anomaly and the computed one, using the estimated values of z , θ and k . The mathematical formula of this criterion is given as

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}\left(V_i\left(Observed\right) - V_i\left(Computed\right)\right)^2}{N}},$$

where $V_i\left(Observed\right)$ and $V_i\left(Computed\right)$ $(i = 1,...,N)$ are the observed and the computed self-potential values at the point x_i $(i = 1,...,N)$.

2.3. Interpretation of self-potential anomaly due to a sphere model

The general expression for the self-potential anomaly (V) at any point $P(x)$ along the x-axis of an arbitrary sphere-like structure in a Cartesian coordinate system *(Figure 1)* is given according to [3] and [15] as

$$V(x_i) = k\,\frac{x_i\cos\theta + z\,\sin\theta}{\left(x_i^2 + z^2\right)^{\frac{3}{2}}} \quad (i = 1,...,N)\cdot \tag{20}$$

The notations in this expression have the same meaning as those presented in *Equation 1*. Following the same manner as performed for the horizontal and vertical cylinder models, the parameters related to the self-potential anomaly produced by a sphere can be estimated by the following equations

$$z = \frac{\left(q_1\right)^{\frac{1}{2}} + \left(q_2\right)^{\frac{1}{4}} + \left(q_3\right)^{\frac{1}{6}}}{3}, \tag{21}$$

$$\theta = \arctan\frac{1}{2z}\left(\sqrt{\frac{q_5}{q_4}} + \frac{1}{2}\frac{q_6 - q_7}{q_4}\right), \tag{22}$$

$$k = \mp\sqrt{\frac{1}{3}\left(\frac{q_4}{\cos^2\theta} + \frac{q_5}{z^2\sin^2\theta} + \frac{q_6 - q_7}{z\,\sin 2\theta}\right)}, \tag{23}$$

where $q_1, q_2, q_3, q_4, q_5, q_6$ and q_7 are obtained by solving the following linear optimization problem using the simplex algorithm

$$
\begin{aligned}
&\min \quad u_1+u_2+u_3+u_4+u_5+u_6 \quad \textit{subject to}\\
&3\left(\sum_{i=1}^{N}V_i^{\,4}x_i^{\,8}\right)q_1+3\left(\sum_{i=1}^{N}V_i^{\,4}x_i^{\,6}\right)q_2+\left(\sum_{i=1}^{N}V_i^{\,4}x_i^{\,4}\right)q_3-\left(\sum_{i=1}^{N}V_i^{\,2}x_i^{\,6}\right)q_4-\left(\sum_{i=1}^{N}V_i^{\,2}x_i^{\,4}\right)q_5\\
&-\left(\sum_{i=1}^{N}V_i^{\,2}x_i^{\,5}\right)q_6+\left(\sum_{i=1}^{N}V_i^{\,2}x_i^{\,5}\right)q_7-u_1=-\left(\sum_{i=1}^{N}V_i^{\,4}x_i^{\,10}\right)\\
&3\left(\sum_{i=1}^{N}V_i^{\,4}x_i^{\,6}\right)q_1+3\left(\sum_{i=1}^{N}V_i^{\,4}x_i^{\,4}\right)q_2+\left(\sum_{i=1}^{N}V_i^{\,4}x_i^{\,2}\right)q_3-\left(\sum_{i=1}^{N}V_i^{\,2}x_i^{\,4}\right)q_4-\left(\sum_{i=1}^{N}V_i^{\,2}x_i^{\,2}\right)q_5\\
&-\left(\sum_{i=1}^{N}V_i^{\,2}x_i^{\,3}\right)q_6+\left(\sum_{i=1}^{N}V_i^{\,2}x_i^{\,3}\right)q_7-u_2=-\left(\sum_{i=1}^{N}V_i^{\,4}x_i^{\,8}\right)\\
&3\left(\sum_{i=1}^{N}V_i^{\,4}x_i^{\,4}\right)q_1+3\left(\sum_{i=1}^{N}V_i^{\,4}x_i^{\,2}\right)q_2+\left(\sum_{i=1}^{N}V_i^{\,4}\right)q_3-\left(\sum_{i=1}^{N}V_i^{\,2}x_i^{\,2}\right)q_4-\left(\sum_{i=1}^{N}V_i^{\,2}\right)q_5\\
&-\left(\sum_{i=1}^{N}V_i^{\,2}x_i\right)q_6+\left(\sum_{i=1}^{N}V_i^{\,2}x_i\right)q_7-u_3=-\left(\sum_{i=1}^{N}V_i^{\,4}x_i^{\,6}\right)\\
&3\left(\sum_{i=1}^{N}V_i^{\,2}x_i^{\,6}\right)q_1+3\left(\sum_{i=1}^{N}V_i^{\,2}x_i^{\,4}\right)q_2+\left(\sum_{i=1}^{N}V_i^{\,2}x_i^{\,2}\right)q_3-\left(\sum_{i=1}^{N}x_i^{\,4}\right)q_4-\left(\sum_{i=1}^{N}x_i^{\,2}\right)q_5\\
&-\left(\sum_{i=1}^{N}x_i^{\,3}\right)q_6+\left(\sum_{i=1}^{N}x_i^{\,3}\right)q_7-u_4=-\left(\sum_{i=1}^{N}V_i^{\,2}x_i^{\,8}\right)\\
&3\left(\sum_{i=1}^{N}V_i^{\,2}x_i^{\,4}\right)q_1+3\left(\sum_{i=1}^{N}V_i^{\,2}x_i^{\,2}\right)q_2+\left(\sum_{i=1}^{N}V_i^{\,2}\right)q_3-\left(\sum_{i=1}^{N}x_i^{\,2}\right)q_4-Nq_5\\
&-\left(\sum_{i=1}^{N}x_i\right)q_6+\left(\sum_{i=1}^{N}x_i\right)q_7-u_5=-\left(\sum_{i=1}^{N}V_i^{\,2}x_i^{\,6}\right)\\
&3\left(\sum_{i=1}^{N}V_i^{\,2}x_i^{\,5}\right)q_1+3\left(\sum_{i=1}^{N}V_i^{\,2}x_i^{\,3}\right)q_2+\left(\sum_{i=1}^{N}V_i^{\,2}x_i\right)q_3-\left(\sum_{i=1}^{N}x_i^{\,3}\right)q_4-\left(\sum_{i=1}^{N}x_i\right)q_5\\
&-\left(\sum_{i=1}^{N}x_i^{\,2}\right)q_6+\left(\sum_{i=1}^{N}x_i^{\,2}\right)q_7-u_6=-\left(\sum_{i=1}^{N}V_i^{\,2}x_i^{\,7}\right)\\
&q_1,q_2,q_3,q_4,q_5,q_6,q_7,u_1,u_2,u_3,u_4,u_5,u_6\geq 0
\end{aligned}
\qquad . \quad (24)
$$

It is worth noting that the sign of k can be also assigned by using the statistical criterion of preference (RMSE), based on the minimal value between the field data anomaly and the computed one, using the estimated values of z , θ , and k .

2.4. Interpretation of a synthetic self-potential anomaly due to a horizontal cylinder model with different levels of Gaussian random noise

A synthetic self-potential anomaly $V(x_i)\,(i=1,...,N)$ due to a horizontal cylinder-like structure is generated from *Equation (1)* using the following values of model parameters: depth to the center of the structure $z=15$ units length, polarization angle $\theta^o=60$, and the amplitude coefficient $k=100\,mV.m$. This generated synthetic anomaly is perturbed by different Gaussian random noise maximum levels of 10 and 15%, using continuous uniform distribution, where two additional self-potential anomalies are regenerated. Both regenerated self-potential anomalies are consequently interpreted by the new proposed method; estimated model parameters values are shown in *Table 1* and *Figure 2*.

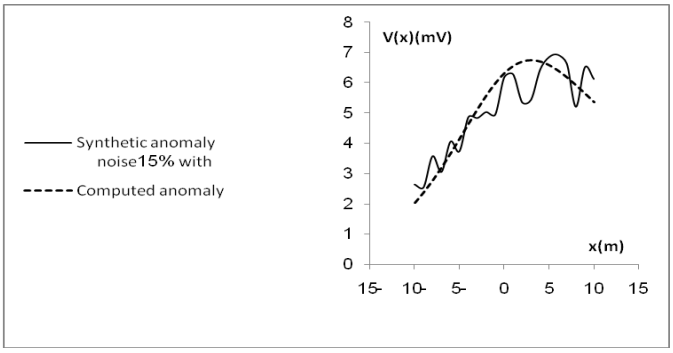

Figure 2. *Diagram for the computed anomaly and synthetic data set adding a maximum random noise of 15%*

Table 1

Interpretation of a synthetic self-potential anomaly due to a horizontal cylinder model with different maximum levels of Gaussian random noise

Model parameters	True values of model parameters	Estimated values of model parameters with maximum 10% random noise	Estimated values of model parameters with maximum 15% random noise
z (unit length)	15	14.05	11.85
θ^0	60	59.49	61.93
k $_{(mV.m)}$	100	96.97	85.12

The results presented in *Table 1* show that the estimated parameter values derived by this interpretation method, are very close to the true values of parameters, which clearly indicates the efficiency and the capability of the proposed interpretation method.

3. APPLICATION TO FIELD DATA

Two field self-potential anomalies over various geological structures are interpreted by the proposed method as follows.

3.1. Interpretation of the Malachite anomaly

Figure 3 shows a self-potential field anomaly measured over the Malachite mine, Jefferson County, Colorado, USA [19], [20]. This self-potential anomaly has been reinterpreted by the proposed method by adapting a priori the three discussed models: a horizontal cylinder, a vertical cylinder, and a sphere as sources of the SP anomaly. The estimated values of the three models parameters and the values of the statistical criterion of preference (RMSE) are explained in *Table 2*.

Table 2

Interpretation of the Malachite mine anomaly, Colorado, USA

Model parameters	Horizontal cylinder	Vertical cylinder	Sphere
z (m)	37.24	12.79	Optimization
θ^0	66.48	79.50	Problem (24) is
k	–7802.31$(mV.m)$	–231.90(mV)	not
$RMSE$ (mV)	15.57	14.95	feasible

It can be concluded from *Table 2* that the sphere source causing this anomaly cannot be accepted as a SP model, because the simplex algorithm indicates clearly that linear optimization (24) is infeasible. The minimal value of the Root Mean Square Error ($\mathrm{RMSE} = 14.95$) has been obtained for the vertical cylinder model, meaning that the self-potential anomaly could be preferably modeled as a vertical cylinder with the following parameter values

$$z = 12.79\,m, \quad \theta^o = -79.50, \quad k = 231.90\,mV$$

The depth to the center of the vertical cylinder ($z = 12.79\ m$) is found to be in very good agreement with that obtained from drilling ($z = 13.7\ m$). The computed self-potential anomaly has been drawn according to these estimated values of model parameters as shown in *Figure 3*. The comparison between field and computed

anomalies clearly indicates the close agreement between them, which attests to the capability and the validity of the proposed method.

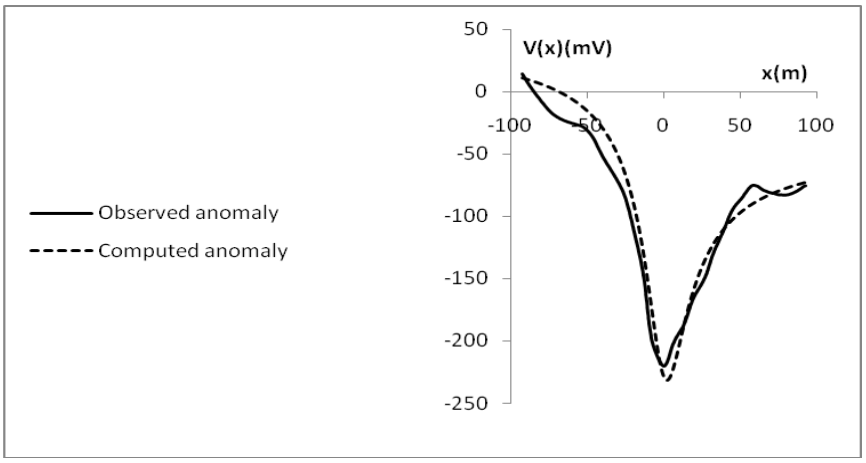

Figure 3. *Self-potential field anomaly over Malachite mine, Colorado, USA. The computed anomaly by the proposed interpretation method is also shown*

3.2. Interpretation of the Sariyer anomaly

Figure 4 shows a self-potential field anomaly measured over a sulfide ore deposit, Sariyer, Turkey [19]–[21]. This self-potential anomaly has been also reinterpreted by the proposed method by adapting a priori the three previous models; a horizontal cylinder, a vertical cylinder and a sphere as sources of the *SP* anomaly. The estimated values of parameters of the three models and the values of the statistical criterion of preference (RMSE) are shown in *Table 3*.

Table 3

Interpretation of the sulfide ore deposit anomaly, Sariyer, Turkey

Model parameters	**Horizontal cylinder**	**Vertical cylinder**	**Sphere**
z (m)	22.58	Optimization	8.88
θ^0	–82.99	Problem (19) is	–5.82
k	4695.60 $(mV.m)$	not	114550 $(mV.m^2)$
$RMSE$ (mV)	14.47	feasible	260.42

It can be concluded from *Table 3* that the source causing this anomaly cannot be modeled as a vertical cylinder, because the simplex algorithm indicates clearly that linear optimization (19) is infeasible. The minimal value of the Root Mean Square Error ($\mathrm{RMSE} = 14.47$) has been obtained for the horizontal cylinder model, mean-

ing that the self-potential anomaly could be preferably modeled as a horizontal cylinder with the following parameter values

$$z = 22.58\,m, \quad \theta^o = -82.99, \quad k = 4695.60\,mV\,.m\,.$$

The depth to the center of the horizontal cylinder ($z = 22.58\,m$) is found to be in very good agreement with that obtained from drilling ($z = 23\,m$). The computed self-potential anomaly has been drawn according to these estimated values of model parameters as shown in *Figure 4*. The comparison between field and computed anomalies clearly indicates the close agreement between them, which attests to the validity of the proposed method.

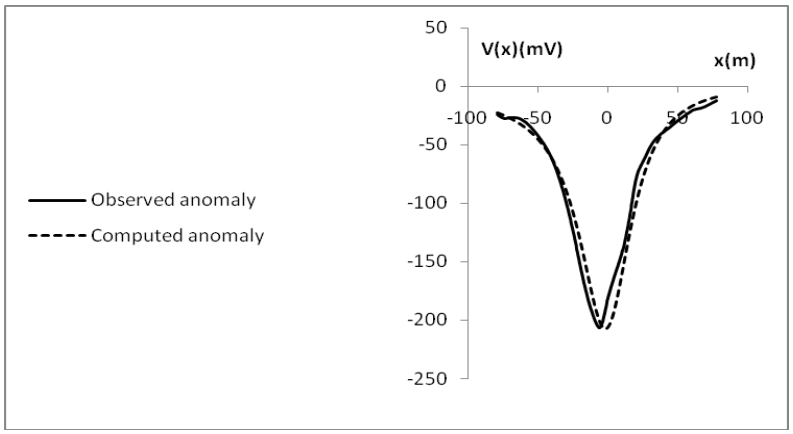

Figure 4. *Self-potential field anomaly over a sulfide ore deposit, Sariyer, Turkey. The computed anomaly by the proposed interpretation method is also shown*

3.3. Discussion

It is useful to mention that there is no loss of generality in assuming the source geometry of the SP anomaly is a priori known. There are in addition no imposed restrictions on the generality of the proposed interpretation method. When the source geometry of the self-potential field anomaly is unknown, three steps must be followed.

First, the self-potential field anomaly is interpreted by adapting the source geometry as a vertical cylinder, where the Root Mean Square Error RMSE(1) is computed between the observed field anomaly and the computed one.

Second, the self-potential field anomaly is re-interpreted by adapting the source geometry as a horizontal cylinder, where the Root Mean Square Error RMSE(2) is computed between the observed field anomaly and the computed one.

Third, the self-potential field anomaly is re-interpreted by adapting the source geometry as a sphere, where the Root Mean Square Error RMSE(3) is computed between the observed field anomaly and the computed one.

The lowest of the three values of RMSE(1), RMSE(2) and RMSE(3) is selected as the solution that exactly indicates the suitable source geometry related to the responsible SP field anomaly.

4. CONCLUSION

In this paper a practical method is proposed for interpreting self-potential anomalies for simple geometric-shaped models such as horizontal cylinder, vertical cylinder, and sphere. The proposed method is based on both the deconvolution technique to avoid local minima and on the simplex algorithm for linear optimization to best estimate the model parameters values, e.g. the depth to the center of the buried structure, the polarization angle and the amplitude coefficient from the self-potential anomaly profile.

The reliability and capability of this interpretation method were first demonstrated through testing and corrupting the synthetic data sets by different white Gaussian random noise levels of 10 and 15%. The synthetic modeling results obtained indicate that the estimated parameter values derived by this new proposed method are very close to the assumed true values of parameters.

The validity of this method is also verified through reinterpreting real field self-potential anomalies taken from the United States and Turkey. A comparable and acceptable agreement is noticed between the results derived by the proposed method and those obtained by other interpretation methods. Moreover, the depth obtained by such a method is found to be in a good accordance with that obtained from the real field data information.

This interpretation method can be easily put into a MATLAB code. Furthermore, since the method is based on the familiar algorithm in linear optimization called the simplex algorithm of Dantzig, convergence towards the best estimation of parameter values is assured and rapidly reached.

This new methodology is therefore recommended for routine analysis of self-potential anomalies in an attempt to determine the best-estimate values of parameters related to different structures under investigation.

ACKNOWLEDGMENTS

Authors would like to thank Dr. I. Othman, Director General of the Syrian Atomic Energy Commission, for his continuous encouragement and guidance to achieve this research. Two anonymous reviewers are deeply thanked for their suggestions and remarks, which considerably improve the final version of this paper.

REFERENCES

[1] YUNGUL, S.: Interpretation of spontaneous polarization anomalies caused by spherical ore bodies. *Geophysics*, 1950, 15. 237–246.

[2] PAUL, M. K.: Direct interpretation of self-potential anomalies caused by inclined sheets of infinite extension. *Geophysics*, 1965, 30. 418–423.

[3] BHATTACHARYA, B. B.–ROY, N.: A note on use of a Nomogram for self-potential anomalies. *Geophysical Prospecting*, 1981, 29, 102–107.

[4] MEISER, P.: A method of quantitative interpretation of self-potential measurements. *Geophysical Prospecting*, 1962, 10, 203–218.

[5] SHALIVAHAN–Bimalendu, B, BHATTACHARYA–Mrinal, K. SEN.; Interpretation of self-potential anomalies by nonlinear inversion. *Journal of Geophysics*, 1998, 19(4), 219–224.

[6] ABDELRAHMAN, E. S. M.–EL-ARABY, T. M.–AMMAR, A. A.–HASSANEIN, H. I.: A least-squares approach to shape determination from residual self-potential anomalies. *Pure and Applied Geophysics*, 1997, 150, 121–128.

[7] ABDELRAHMAN, E. S. M.–SHARAFELDIN, S. M.: A least-squares approach to depth determination from self-potential anomalies caused by horizontal cylinders and spheres. *Geophysics*, 1997, 62, 44–48.

[8] ATCHUTA, Rao–RAM BABU–SIVAKUMAR SINHA, G. D. J.: A Fourier transform method for the interpretation of self-potential anomalies due to two to two-dimensional inclined sheets of finite depth extent. *Pure and Applied Geophysics*, 1982, 120, 365–374.

[9] SUNDARARAJAN, N.–ARUN KUMAR, I.–MOHAN, N. L.–SESHAGIRI RAO, S. V.: Use of Hilbert transform to interpret self-potential anomalies due to two dimensional inclined sheets. *Pure and Applied Geophysics*, 1990, 133, 117–126.

[10] SUNDARARAJAN, N.–SRINIVAS. Y.: A modified Hilbert transform and its application to self-potential interpretation. *Journal of Applied Geophysics*, 1996, 36, 137–143.

[11] SUNDARARAJAN, N.–SRINIVASA RAO, P.–SUNITHA, V.: An analytical method to interpret self-potential anomalies caused by 2-D inclined sheets. *Geophysics*, 1998, 63(5), 1551–1555.

[12] ABDELRAHMAN, E. M.–AMMAR, A. A.–HASSANEIN, H. I.–HAFEZ, M. A.: Derivative analysis of *SP* anomalies. *Geophysics*, 1998, 63, 890–897.

[13] ABDELRAHMAN, E. M.–EL-ARABI. H. M.–HASSANEEN. A. R.–HAFEZ. M. A.: New methods for shape and depth determinations from *SP* data. *Geophysics*, 2003, 68, 1202–1210.

[14] PHILLIPS D. T.–RAVINDRA A.–SOLBER J. J.: Operations Research. John Wiley and Sons, Inc., 1976.

[15] TLAS, M.–ASFAHANI, J.: An approach for interpretation of self-potential anomalies due to simple geometrical structures using Fair function minimization. *Pure and Applied Geophysics*, 2012, 169.

[16] HILLIER, F.–LIEBERMAN, G. J.: Introduction to Operations Research. Holden-Day, Inc., 1986.

[17] BRADLEY, S. P.–HAX, A. C.–MAGNANTI, T. L.: Applied Mathematical Programming. Addison-Wesley, 1977.

[18] COLLINS, G. W.: Fundamental Numerical Methods and Data Analysis. Case Western Reserve University, 2003.

[19] DOBRIN, M. B.: Introduction to geophysical prospecting. Mc Graw-Hill Book Company, Inc., 1960.

[20] ABDELRAHMAN, E. M., SABER, H. S., ESSA, K. S., and FOUDA, M. A.: A least-squares approach to depth determination from numerical horizontal self-potential gradients. *Pure and Applied Geophysics*, 2004, 161, 399–411.

[21] YUNGUL, S.: Spontaneous potential survey of a copper deposit at Sariyer, Turkey. *Geophysics*, 1954, 19, 455–458.

Geosciences and Engineering, Vol. 5, No. 8 (2016), pp.22–31.

INTERVAL INVERSION OF ENGINEERING GEOPHYSICAL SOUNDING LOGS

GERGELY PÁL BALOGH

Department of Geophysics, University of Miskolc,
Miskolc-Egyetemváros H-3515 Hungary
E-mail: gfbg@uni-miskolc.hu

Abstract: A new inversion approach is presented for the interpretation of engineering geophysical sounding logs. In shallow structures, the water, sand and clay content can be estimated by the interval inversion method. Gas volume is derived from the inversion results. All penetration sounding data along the entire borehole is evaluated in one joint inversion process. The highly overdetermined interval inversion method gives accurate, reliable and stable results for the petrophysical parameters. The method can be easily upgraded by using arbitrary basis functions for the discretization, in which a series expansion method is used to describe the vertical distribution of the petrophysical parameters. The interval inversion method is numerically tested by using real penetration sounding data measured in the Bátaapáti test site.

Keywords: *engineering geophysical sounding, interval inversion, series expansion, petrophysical parameter.*

1. INTRODUCTION

Cone Penetration Tests (CPT) complemented by geophysical measurements can be effectively used for the in-situ investigation of shallow unconsolidated structures. Several soil parameters can be observed in penetration holes with a special feature that only the steel tube isolates the probe from the formation, and the measured data are transferred through the rods pushed into the ground. A Hungarian direct-push method called Engineering Geophysical Sounding (EGS) allows the measurement of different geophysical parameters such as electric resistivity, natural gamma radioactivity, bulk density and neutron-neutron intensity [1]. The EGS technique using all suitable logs gives quantitative information on the composition, porosity, shale content and water content of the unsaturated sediments. EGS data processing mostly incorporates deterministic, point-by-point inversion or factor analysis methods adapted from oilfield well-log interpretation [2], [3], [4].

In the point-by-point inversion procedure, the petrophysical properties as unknowns are determined separately at each depth using the data measured at the same point [4]. If the zone parameters are assumed constant, we have slightly more data than unknowns, thus, a marginally overdetermined inverse problem must be solved. In contrast, the interval inversion procedure is used to invert a data set of a

larger depth interval to estimate the model parameters for the same interval. In addition to the large quantity of data, the number of model parameters can be highly reduced by using a series expansion discretization technique; thus, the inverse problem will be overdetermined. In this study, the interval inversion method is applied to EGS data and tested in a Hungarian measurement area.

2. POINT-BY POINT INVERSION

The rock matrix of groundwater formations is composed of coarse and fine grain components, while their pore space is saturated with freshwater and air. The model vector of inverse problem is composed of the soil parameters given below

$$\mathbf{m} = \left[V_{cl}, V_s, V_w, V_g\right]^{\mathrm{T}}, \tag{1}$$

which contains the volumetric ratios of clay (V_{cl}), sand (V_s), water (V_w) and gas (V_g), T is the symbol of matrix transpose. In environmental studies, usually natural gamma radioactivity (*GR*), formation density (*DEN*), neutron-porosity (*NPHI*) and resistivity (*RES*) data are measured by EGS instruments. The following response equations can be used for calculating the theoretical (*TH*) EGS data [2]

$$GRTH = V_{cl}GR_{cl} + V_sGR_s, \tag{2}$$

$$DENTH = V_w\rho_w + V_{cl}\,\rho_{cl} + V_s\rho_s, \tag{3}$$

$$NPHITH = V_w\Phi_{N,w} + V_{cl}\Phi_{N,cl} + V_s\Phi_{\mathrm{N,s}}, \tag{4}$$

$$RESTH = \left(V_w + V_g + V_{cl}\right)^{-2}\left(\frac{V_{cl}/(V_w+V_{cl})}{R_{cl}} + \frac{1-[\,V_{cl}/(V_w+V_{cl})\,]}{R_w}\right)^{-1}\left(\frac{V_w+V_{cl}}{V_w+V_g+V_{cl}}\right)^{-2}, \tag{5}$$

where the physical constants of rock constituents and pore-fluids are indicated by *cl* (clay), *s* (sand), *w* (water), *g* (gas) ρ denotes mass density, Φ_N is neutron porosity, and textural parameters $a = 1$, $k = m = 2$ were chosen as default values representing the tortuosity factor, saturation exponent, and cementation exponent, respectively. For a more accurate evaluation, in unconsolidated rocks, the cementation component is used as $m < 2$ [5]. The calculated EGS data are arranged in the vector $\mathbf{d}^{(\mathrm{cal})}$. Similarly to *Equations (2)–(5)*, the measured data vector in a given depth is

$$\mathbf{d}^{(\mathrm{obs})} = \left[GR, DEN, NPHI, RES\right]^{\mathrm{T}}. \tag{6}$$

In point-by-point inversion, after the calculation of theoretical logs, an independent local inversion procedure is solved in each depth. The objective function is chosen as the weighted square of the Euclidean norm of the difference between the measured and calculated data

$$\Omega = (\mathbf{e}, \mathbf{We}) = \mathbf{e}^{\mathrm{T}}\mathbf{We} = \left(\mathbf{d}^{(\mathrm{obs})} - \mathbf{d}^{(\mathrm{cal})}\right)^{\mathrm{T}} \mathbf{W}\left(\mathbf{d}^{(\mathrm{obs})} - \mathbf{d}^{(\mathrm{cal})}\right) = \min . \tag{7}$$

The elements of weight matrix **W** are inversely proportional to the variances of the measured EGS logs (σ_d) representing the accuracy of measurements

$$\mathbf{W} = \mathrm{diag}\left(\sigma_{\mathrm{d},1}^{-2}, \sigma_{\mathrm{d},2}^{-2}, \ldots, \sigma_{\mathrm{d},N}^{-2}\right), \tag{8}$$

where N is the number of applied sounding tools. To increase the over-determination of the inverse problem, the gas volume is calculated by the material balance equation $V_g = 1 - V_w - V_{cl} - V_s$. Having four data types against three unknowns, the above presented inverse problem is overdetermined. We can use the weighted least squares method for the solution of the local inverse problem [6], which gives a solution to the model vector

$$\mathbf{m} = \left(\mathbf{G}^{\mathrm{T}}\mathbf{WG}\right)^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{Wd}^{(\mathrm{obs})}, \tag{9}$$

where **G** is the Jacobi's matrix. The local inversion method includes the calculation of the estimation errors of the volumetric soil parameters [6]. The above method is fast, but quite sensitive to data noises. To improve the accuracy of inversion estimation, we increase the over-determination ratio by using the interval inversion method.

3. INTERVAL INVERSION

Dobróka [7] developed the interval inversion method, which gives an estimate of the petrophysical properties and the layer boundaries in a highly overdetermined inverse problem. The petrophysical parameters are assumed as functions of depth and discretized using a series expansion method

$$m_i(z) = \sum_{q=1}^{Q_i} B_q^{(i)} \psi_q(z), \tag{10}$$

where $B_q^{(i)}$ is the q-th series expansion coefficient used for describing the i-th soil parameter *(*i=1,2,…M*)*, where M is the total number of petrophysical parameters,

Q_i is the number of necessary expansion coefficients, ψ_q is the q-th depth-dependent basis function. In this study, we use the Legendre polynomials as basis functions. The zero-th, first and n-th degree Legendre polynomials (P_n) are the following

$$P_0(x) = 1, \tag{11}$$

$$P_1(x) = x, \tag{12}$$

$$P_n(x) = \frac{1}{n}\{(2n-1)xP_{n-1}(x) - (n-1)P_{n-2}(x)\} \ (n \geq 2). \tag{13}$$

The above set of functions is orthogonal in the interval of –1 and 1. By using *Equations (10)* and *(13)*, a much lower number of expansion coefficients can be used than EGS data. The unknowns of the inverse problem are the series expansion coefficients. The solution is given by the minimization of the distance between the measured and calculated data using *Equation (7)*. Having estimated the expansion coefficients by the interval inversion procedure, the soil parameters are derived by *Equation (10)*. Similarly to Menke's method [6], once we know the data error, we can determine the estimation of error of soil parameters. The covariance matrix of the soil properties can be derived from the error of the series expansion coefficients. Dobróka et al. [8] defined the depth-dependent covariance using the error propagation law

$$[\operatorname{cov}\mathbf{m}(z)]_{ij} = \sum_{n=1}^{Q^{(i)}} \sum_{m=1}^{Q^{(j)}} P_{n-1}(z)(\operatorname{cov}\mathbf{B})_{\mathrm{hh}'} P_{m-1}(z), \tag{14}$$

where **B** is the vector of series expansion coefficients *(i=1,2,…,M, j=1,2,…,M, $h=n+Q_1+Q_2+\ldots+Q_{i-1}$, $h'=m+Q_1+Q_2+\ldots+Q_{j-1}$)*. The first diagonal elements of *Equation (14)* represent the variances of soil parameters. We can easily deduce the correlation coefficient of petrophysical parameters from *Equation (14)*, which characterizes the reliability of the estimated parameters for the entire processing interval

$$\operatorname{corr}(\mathbf{m}(z))_{ij} = \frac{\operatorname{cov}(\mathbf{m}(z))_{ij}}{\left[\operatorname{var}(\mathbf{m}(z))_i \operatorname{var}(\mathbf{m}(z))_j\right]^{1/2}}. \tag{15}$$

In addition to the above characteristics, we can define more metrics that provide information of the convergence of the inversion process. The deviation of the calculated and measured data can be characterized with the relative data distance

$$\hat{D}_d = \frac{1}{S}\sum_{s=1}^{S} D_{d,p} = \frac{1}{S}\sum_{s=1}^{S}\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{d_i^{(\mathrm{obs})} - d_i^{(\mathrm{cal})}}{d_i^{(\mathrm{obs})}}\right)^2}\right)_p \cdot 100\,(\%)\,, \qquad (16)$$

where N is the total number of sounding tools, S is the total number of measurement points.

4. CASE STUDY

The interval inversion method is tested using real EGS data measured in Bátaapáti, Hungary The formations contain loose clay and loessy sand saturated with air and water. Natural gamma-ray intensity *(GR)*, density *(DEN)*, neutron-porosity *(NPHI)* and resistivity *(RES)* logs are available for interval inversion. The physical variables observed in a penetration hole (Hole 6) and their errors are plotted in *Figure 1*. The accuracy of measured data is given after Drahos [4]. The standard deviation of the EGS logs are assumed as $\sigma_1 = 0.22$ kcpm, $\sigma_2 = 0.07$ g/cm^3, $\sigma_3 = 0.04$ V/V, $\sigma_4 = 2.1$ ohm-m.

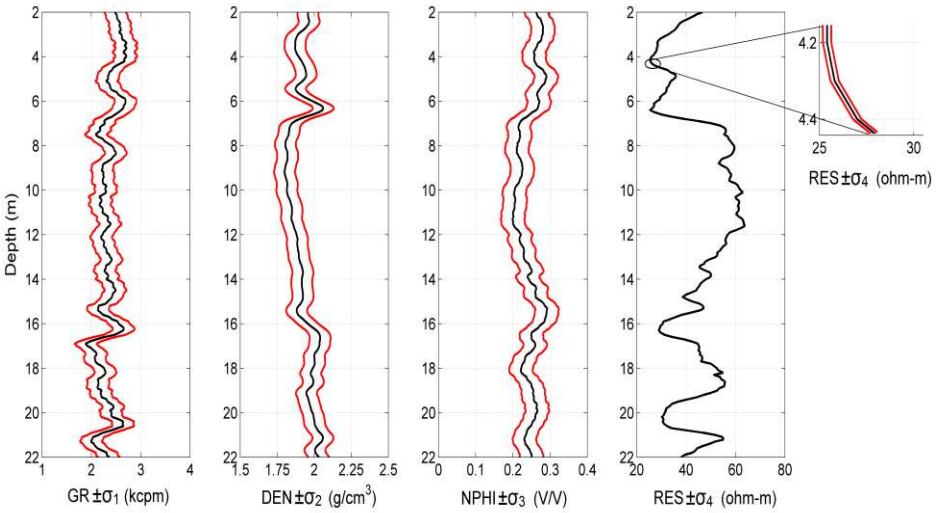

Figure 1. *EGS logs measured in Hole 6, Bátaapáti*

The model parameters *(V_w, V_s, V_{cl})* are discretized by using *Equation (10)*. The degree of Legendre polynomials for all petrophysical parameters is set to 40. The vertical coordinates were transformed into the range of −1 and 1, and after inversion were scaled back to the range of 2 and 22 (m). The data set of the whole length of the borehole was processed in one inversion procedure. The development

of convergent is continuous, thus, the inversion procedure is stable *(Figure 2)*. The correlation coefficients of the series expansion coefficients show reliable inversion results; the average correlation is 0.16, which is significantly smaller than that of traditional point-by-point inversion methods (~0.5).

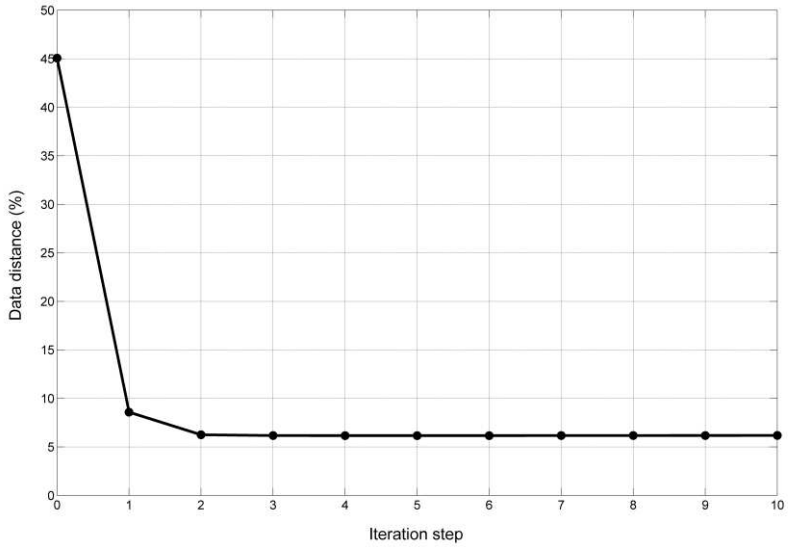

Figure 2. *Convergence of the data distance in the interval inversion procedure*

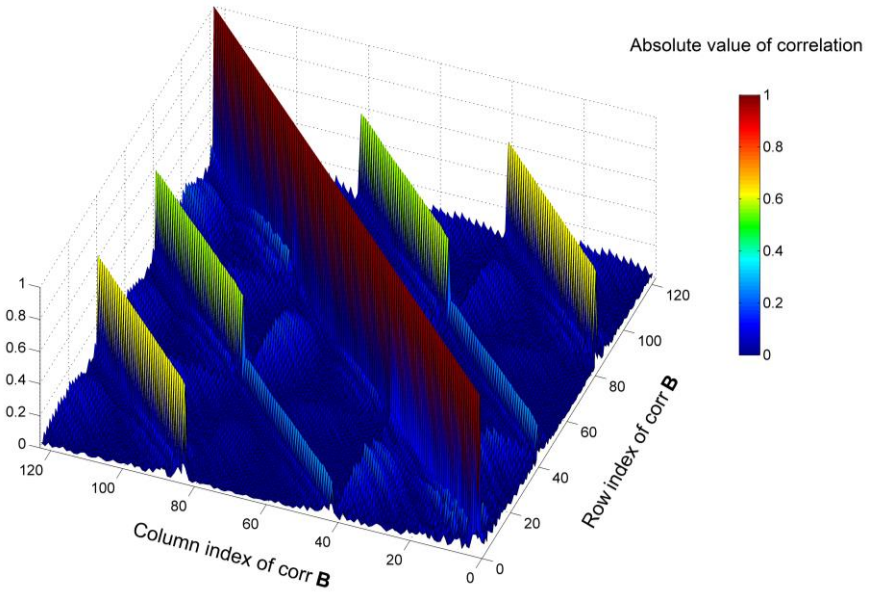

Figure 3. *Correlation matrix of series expansion coefficients as inversion unknowns*

The correlation matrix of estimated model parameters is illustrated in *Figure 3*. The indices correspond to the ordinal number in vector **B**. The mean correlation of 0.22 indicates highly reliable inversion results. The first 41 number of expansion coefficients refer to the water volume, the second to the sand content and the last to the clay volume. Except for the expansion coefficients related to the same order, the correlation between the inversion unknowns is remarkably small.

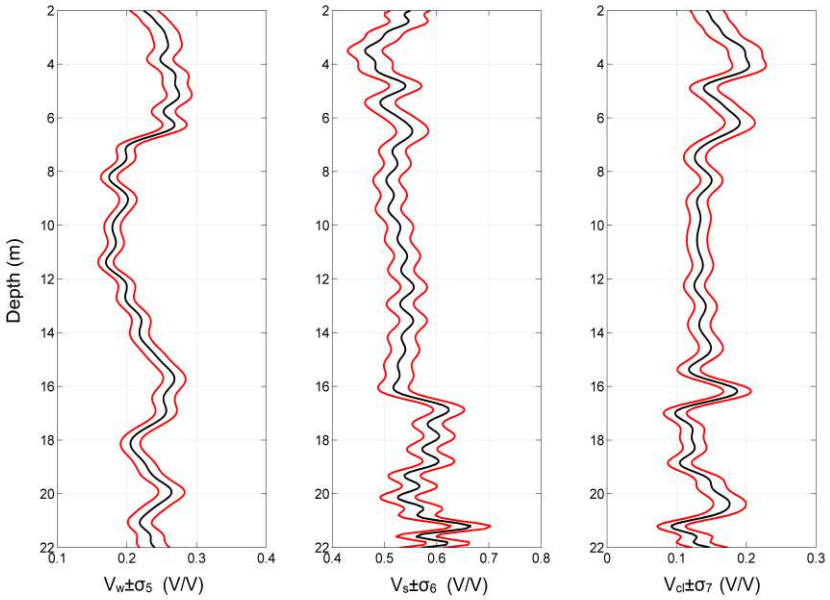

Figure 4. *Model parameters is estimated by interval inversion in Hole 6 Bátaapáti*

The interval inversion results and the petrophysical parameters such as water volume, sand volume and clay volume with their estimation errors are plotted in *Figure 4*. The well logs of the estimated soil parameters show accurate inversion results. The estimation errors are $\sigma_5 =$ 3.4% (for the water content), $\sigma_6 =$ 6.6% (for the volumetric ratio of sand) and $\sigma_7 =$ 4.1% (for the volumetric ratio of clay). The inversion tests demonstrate a highly acceptable inversion approach. For better resolution, higher degree basis functions should be used; however, it is important to keep in mind that this will reduce the overdetermination ratio and estimation accuracy.

5. CONCLUSIONS

An interval inversion method adapted from oilfield applications is proposed for the evaluation of EGS measurements. From interval inversion, it is possible to obtain a reliable estimate for the water, sand and clay content of near-surface layers. In

contrast with the point-by-point inversion procedure, the interval inversion method provides the results with greater accuracy and reliability. The interval inversion can be further improved if some soil parameters are given from multivariate statistical methods (e.g. from factor analysis), which are fixed during the interval inversion procedure. This statistical approach results in even better accuracy. Another possibility is the automated estimation of layer-boundaries and zone parameters by the interval inversion procedure; several attempts at this have been made at the Department of Geophysics (University of Miskolc) in hydrocarbon well-log analysis.

ACKNOWLEDGEMENTS

The author thanks Dr. Norbert Péter Szabó (University of Miskolc) for his scientific advice and Dr. Dezső Drahos (Eötvös Loránd University) and János Stickel (Elgoscar Ltd.) for their cooperation.

LIST OF SYMBOLS

Symbol	Description
a	tortuosity factor
$\mathbf{B}$	vector of series expansion coefficients
$B_q^{(i)}$	q-th series expansion coefficient for describing the i-th soil parameter
D_d	relative data distance
$\mathbf{d}^{(\mathrm{cal})}$	calculated data vector
$\mathbf{d}^{(\mathrm{obs})}$	measured data vector
DEN	measured formation density
$DENTH$	theoretical formation density
$\mathbf{e}$	vector of prediction errors
$\mathbf{G}$	Jacobi's matrix
GR	measured natural gamma intensity
$GRTH$	theoretical natural gamma intensity
GR_{cl}	natural gamma intensity of clay
GR_s	natural gamma intensity of sand
k	saturation exponent
m	cementation component
$\mathbf{m}$	model parameter vector
M	total number of petrophysical parameters

Symbol	Description
N	number of sounding tools applied
n	degree of applied Legendre polynomials
$NPHI$	measured neutron-porosity
$NPHITH$	theoretical neutron-porosity
P_i	i-th degree Legendre polynomials
Q_i	number of necessary expansion coefficients
RES	measured resistivity
$RESTH$	theoretical resistivity
R_{cl}	resistivity of clay
R_w	resistivity of pore water
S	total number of measurement points
V_{cl}	clay volume
V_g	gas volume
V_s	sand volume
V_w	water volume
$\mathbf{W}$	weight matrix in data space
$\Phi_{N,cl}$	neutron porosity of clay
$\Phi_{N,s}$	neutron porosity of sand
$\Phi_{N,w}$	neutron porosity of water
ρ_{cl}	mass density of clay
ρ_s	mass density of sand
ρ_w	mass density of water
$\sigma_{d,i}$	standard deviation of i-th measured data
σ_1	standard deviation of natural gamma-ray intensity
σ_2	standard deviation of density
σ_3	standard deviation of neutron porosity
σ_4	standard deviation of resistivity
σ_5	estimation errors for water volume
σ_6	estimation errors for sand volume
σ_7	estimation errors for clay volume

Symbol	Description
Ψ_q	q-th depth-dependent basis function
Ω	objective function to be minimized

REFERENCES

[1] FEJES, I.–JÓSA, E.: The engineering geophysical sounding method. Principles, instrumentation, and computerised interpretation. In: S. H. WARD (ed.): *Geotechnical and Environmental Geophysics*. 2. Environmental and groundwater, 1990, 321–331.

[2] DRASKOVITS, P., FEJES, I.: Geophysical methods in drinkwater protection of near-surface reservoirs. *Journal of Applied Geophysics,* 1994, 31, 53–63.

[3] SZABÓ, N. P.–DOBRÓKA, M.–DRAHOS, D.: Factor analysis of engineering geophysical sounding data for water saturation estimation in shallow formations. *Geophysics*, 77, 2012, WA35–WA44.

[4] DRAHOS, D.: Inversion of engineering geophysical penetration sounding logs measured along a profile. *Acta Geodetica et Geophysica Hungarica*, 40, 2005, 193–202.

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

[6] MENKE, W.: *Geophysical Data Analysis: Discrete Inverse Theory*. Elsevier, 1984.

[7] DOBRÓKA, M.: Együttes inverziós algoritmusok bevezetése a mélyfúrási geofizikai értelmezésbe. Final report. Miskolci Egyetem, Geofizikai Tanszék, Miskolc, 1995.

[8] DOBRÓKA, M.–SZABÓ, N. P.–TÓTH, J.–VASS, P.: Interval inversion approach for an improved interpretation of well logs. *Geophysics*, Vol. 81 (2), 2016, D163–D175.

Geosciences and Engineering, Vol. 5, No. 8 (2016), pp.32–50.

UNCONVENTIONAL SHALE CHARACTERIZATION USING IMPROVED WELL LOGGING METHODS

ISTVÁN BIBOR[1] –NORBERT PÉTER SZABÓ[1, 2]

[1]*Department of Geophysics, University of Miskolc,*
H-3515 Miskolc-Egyetemváros, Hungary
[2]*MTA-ME, Geoengineering Research Group, Miskolc, Hungary*
norbert.szabo.phd@gmail.com

Abstract: This paper provides an overview of the well-logging methods currently used in the evaluation of shale gas formations. The combination of natural gamma-ray spectroscopy, density, neutron, sonic and resistivity logging allows in-situ identification of unconventional reservoirs and the calculation of their total organic matter content. We improve the regression relation between the total organic content and parameter Δd estimated by the Zhao's method. The level of maturity as a key parameter should be a priori known, and is treated as an arbitrarily chosen constant during the procedure. We suggest estimating the optimal value of the maturity constant by reducing the misfit between the observed well logs and theoretical logs calculated by the Passey's model in an automated procedure. We compare the results of the improved well-log-analysis methods with core measurements. We recommend the use of multivariate statistical and inversion methods to explore the lithologic and petrophysical characteristics from independent sources, which may help in resolving the ambiguity frequently encountered with the interpretation of unconventional formations.

Keywords: *well logging, unconventional reservoir, shale gas, total organic content, level of maturity.*

1. INTRODUCTION

The world's petroleum consumption has been growing steadily since the 1950s. The continuous rise in living standards and demand entailed increased oil and gas production, which helped the oil industry to take a leading role in the energy sector. However, the Earth's hydrocarbon reserves are declining globally. To reach more resources, significant technological progress is going on today, which allows the exploitation of complex reservoirs previously thought unattainable. Conventional hydrocarbon formations such as sediments with high porosity and permeability can be easily and cheaply produced, but they are relatively small in number and contain just a small percentage of the Earth's stored hydrocarbons. The production of extra-low permeability reservoirs is more expensive and technically challenging; however, the price of hydrocarbons and the continuous rise of new technologies have enabled their efficient operation. Unconventional reservoirs are of great im-

portance, because they comprise a significant portion of the Earth's hydrocarbon reserves. The first exploited unconventional reservoir types were heavy oils and oil sands. However, extensive production can be primarily connected to shale gas formations found in the USA. After the successful production of the Barnett Shale, several North American shale reservoirs have been utilized, e.g. the Wolfcamp, Ford-Eagle, Haynesvillem, Fayetteville, Marcellus, Bakken, Woodford, Montney and Niobrara formations [1].

Unconventional reservoirs constitute a complex geological and petrophysical system. They have a number of subtypes, e.g. shale gas, shale oil, heavy oil, gas hydrates, coal-bed methane and tight gas sand [2]. Two large groups of unconventional reservoirs can be distinguished based on their relation to the source rock. The first group includes those formations where the reservoir is different from the source rock, and the original conventional reservoir is transformed into a non-conventional one. The second group practically consists of source rock reservoirs. During their generation, the sediment with high organic matter content is buried and comes under heat. With time and sufficient pressure and temperature conditions, the organic material changes to kerogen, then to hydrocarbon. During the evolution of hydrocarbon, the pore space is filled up and its pressure increases. The pressure increase may reach a critical value when the rock cracks and loses its hydrocarbon content, thereby feeding the migration routes. If the pore pressure does not exceed the critical level, the resulting hydrocarbons remain in the source rock, creating unconventional reservoirs [1]. These types of reservoirs are typically made up of fine-grained rocks, amongst which pore size of micro- and nanometer order is reached [3].

In this study, we focus on the exploration of shale reservoirs. The composition of shale gas formations may vary between wide ranges. They contain mainly quartz and clay minerals and carbonates. The relative amounts of these constituents vary widely, but normally good quality shale reservoirs generally have high silica content. The permeability of reservoirs is low (< 0.1 mD) and the gas and oil are found in small pore spaces, bound to the remaining organic matter or on the surface of clay minerals. Their production requires increasing porosity, which is implemented by using horizontal drilling technologies and high volume hydraulic fracturing [1].

Significant shale gas reserves are found in China, Argentina, the United States, Canada and Australia. Among them, China has the largest amount of stocks, namely 1,115 trillion cubic feet. In Europe, the biggest shale gas potential has been confirmed in Poland, where the resource amounts to 148 trillion cubic feet [4]. In this paper, we present the theory and application of the well-logging methods currently used to detect shale reservoirs of high organic content. We study the possible further development of these methods using an improved regression analysis and parameter optimization, and give preliminary results for two case studies.

2. PETROPHYSICAL MODELING OF SHALE RESERVOIRS

Unconventional reservoirs are complex structures; thus, the determination of their petrophysical properties represents one of the major challenges for well log analysts. Traditional well-logging methods used in the exploration of conventional reservoirs often do not work, and the relation between the measured quantities and petrophysical parameters is not always consistent or known. However, in order to produce hydrocarbons economically, the knowledge of these parameters is of prime importance. To reduce the ambiguity of the interpretation, multidisciplinary approaches are normally followed, in which the results of well log analysis are validated using laboratory data and other information originating from geological, petrophysical, geophysical and geotechnical sources.

The petrophysical properties of shale reservoirs to be determined are the Total Organic Content (TOC), porosity Φ, pore-size distribution, pore geometry, hydrocarbon and water saturation, fractional volumes of matrix-forming minerals, clay volume, permeability, elastic parameters, and stress conditions. Shale formations are normally easily identified; however, it is a more complicated issue to determine whether they contain hydrocarbons. Among the above listed parameters, the best indicator is TOC, which specifies the amount of organic matter in the formation, by weight. PASSEY et al. [5] distinguished three possible rock types, i.e. organic matter-free rock, organic matter rich immature rock, or mature source rock. The first type consists of the rock matrix and water occupying the pore space *(Figure 1a)*. In the general case, the rock matrix is made up of several minerals, the relative amounts of which have to be determined. The organic matter in immature rocks is located dispersedly between the rock particles *(Figure 1b)*. With increase in compaction, in many cases, the particles of organic matter are arranged in horizontal lamellae. During the process of maturation, the organic material, formed into kerogen, becomes sufficiently ductile to be pressed into the pore space. As the maturation progresses, some of the organic material is converted into liquid or gaseous hydrocarbon which migrates into the pores, displacing the originally stored water *(Figure 1c)*.

ZHAO et al. [6] suggested a slightly different petrophysical model for organic matter-free rocks by taking into account the clay content of the formation *(Figure 1d)*. For organic matter-rich rocks, the previous model is supplemented by a fourth phase representing the kerogen *(Figure 1e)*. The two models are basically the same, describing the organic content as part of the rock matrix. Thus, the fundamental task in organic matter-rich formations is the estimation of total organic content, the matrix volume and hydrocarbon-saturated porosity.

It must be mentioned that there are more sophisticated rock models, in which the matrix is subclassified into non-clay minerals, silt and clay, and the water volume is further divided into mobile, capillary and clay-bound water, and into free hydrocarbons and absorbed gas. An improved interpretation workflow is presented by HOLMES et al. [7], which is applied successfully to the biggest shale gas reservoirs in the USA.

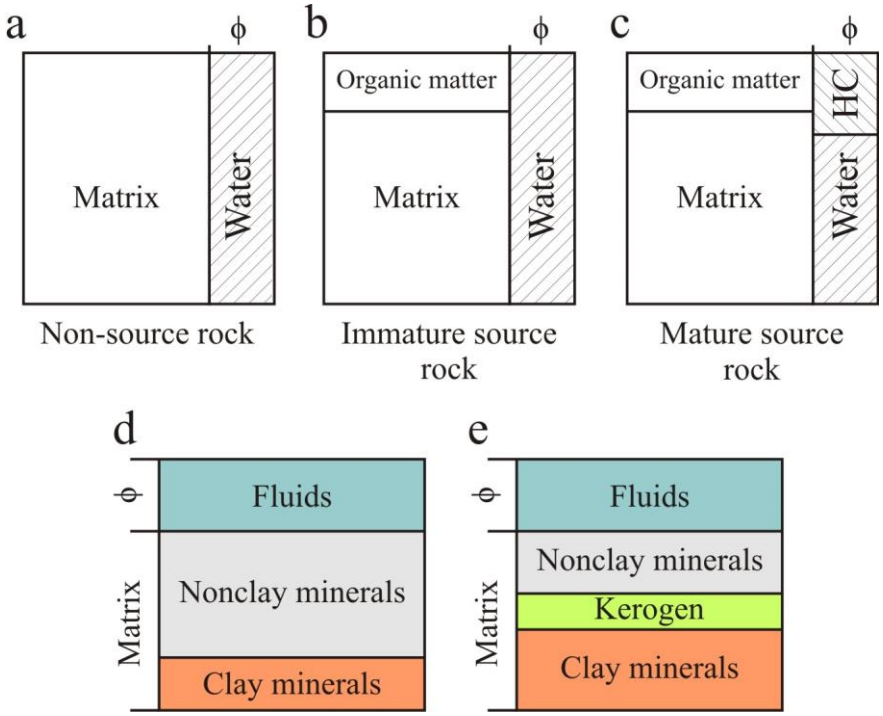

Figure 1.
Petrophysical models used for the interpretation of unconventional reservoirs: rock models (a)–(c) (PASSEY et al. [5]) and (d)–(e) (ZHAO et al. [6])

3. WELL LOGGING IN SHALE RESERVOIRS

The in-situ identification of unconventional reservoirs is based on the use of well logging methods. BEERS [8] first studied natural radioactivity and its possible links with organic content. This is because in sedimentary rocks containing organic matter the concentration of uranium increases [9], [10], [11]. High natural gamma-ray intensity *(GR)* values are mostly associated to the uranium content, while the clay content does not necessarily increase [1]. Uranium-rich sediments are formed in marine conditions, where the uranium is bound to planktonic matter and other marine deposits. The uranium concentration is governed by four factors, i.e. uranium content at the age of deposition, type of the deposited organic matter, chemical conditions at the water-sediment boundary and the velocity of sedimentation [10]. In lacustrine sediments, due to the lack of uranium ions, the increased uranium concentration is not typical.

Since the resistivity of both organic matter and hydrocarbons is high, we measure high apparent resistivities in unconventional shales [9]. In the analysis of resistivity logs, the composition of the rock matrix must be taken into account, e.g. py-

rite is also a common building block in organic-rich layers, the amount of which must be considered. A linear relation was observed between the amount of organic material and pyrite concentration [12]. The clay volume in shale reservoirs may reach 30–40 v/v. If the salinity of pore fluid is low and the matrix does not contain conductive minerals, the clay conductivity should be added to the rock conductivity.

The three most commonly used porosity measurements are acoustic, compensated density, and compensated neutron logs. All three methods respond to the presence of organic matter because of its physical properties. Depending on its composition, the organic material has an average density of 1 g/cm^3, sonic transit-time of 180 μs/ft, and hydrogen index of 0.67 [13]. The organic matter in mature reservoirs is indicated by high porosities. The presence of organic matter, due to its low P-wave velocity, gives high acoustic-interval times [14], which is also affected by the mineral composition, porosity, and type of pore-filling fluid. The bulk density of shale gas formations is normally low, which is due to the low density of the organic matter. The neutron-porosity response is mostly influenced by the hydrogen content of organic matter, pore water, and hydrocarbons and the clay volume. Other advanced logging technologies can also be used for lithology identification (e.g. spectral *GR* and photoelectric absorption index logs), improved estimation of porosity and permeability (e.g. nuclear magnetic resonance log), and mechanical characterization (e.g. full waveform sonic log).

4. EVALUATION OF SHALE GAS RESERVOIRS

Several well-logging methods can be used for the determination of the organic material content. SCHMOKER [15] suggested the use of the density log with the condition that the heavy minerals like pyrite must be taken into account for the interpretation. The correlation between the natural gamma-ray intensity and organic content also allows the suitable determination of TOC [16]. Spectral gamma-ray logging is used to estimate the uranium, thorium, and potassium content of the formation. Since there is a strong correlation between the TOC and uranium content, the well logs of uranium versus thorium and uranium versus potassium ratios are also applicable [17]. The difference between the potassium/thorium and uranium logs shows an empirical correlation to the organic matter content [14]. Another alternative is the use of the carbon-oxygen-ratio log, which gives the total carbon content of the formation [18]. This log should be corrected to the amount of calcium, which provides a possibility to distinguish inorganic and organic carbon sources. The combination of different log types provides a more reliable estimation of organic matter content. Most of these techniques are based on the simultaneous processing of porosity and resistivity logs [16], [19], [20]. The following subsections present the interpretation methods used in this paper and give an overview of those that can be used in future studies.

4.1 The Δlog*R* method

The Passey method and its industrial software implementation form one of the most widespread well logging methods for the estimation of TOC. It is based on the combination of sonic and resistivity measurements, both sensitive to the presence of organic matter. Sonic data responds to its low density and acoustic traveltime, while resistivity values are increased because of its high resistivity [5]. We plot the two well logs on the same track in a reverse scale. The two curves are appropriately scaled so that one segment of 100 μs/ft on the linear scale of the acoustic log is equal to two decades on the logarithmic scale of the resistivity log. For describing the difference between the acoustic traveltime Δt and resistivity R curves, the following variable was introduced:

$$\Delta \log R = \log_{10}\left(\frac{R}{R_{base}}\right) + \frac{1}{50}\left(\Delta t - \Delta t_{base}\right), \tag{1}$$

where R_b and Δt_b denote the resistivity reading and acoustic traveltime observed at the baseline, respectively. In fine-grained organic-matter-free (water saturated) shaly formations, both curves run along the baseline, while they are separated from each other in organic matter-rich shale reservoirs [5]. The separation between the two curves is directly proportional to the TOC. The method assumes a nearly linear connection between quantities Δlog$R$ and TOC. This connection holds only when the constant value of Level of Organic Maturity (LOM) is known; LOM is normally measured in the laboratory. An estimate can be given to the value of TOC by using

$$\mathrm{TOC} = \Delta \log R \cdot 10^{(a - b \cdot \mathrm{LOM})}, \tag{2}$$

where the default values of regression constants are a = 2.297 and b= 0.1688 [5]. The TOC is estimated in weight percent, which must be corrected for the original organic carbon content of the shale (it is ~0.8%). If the sonic log is not available, the following empirical formulae can be used analogously to *Equation (1)*

$$\Delta \log R_N = \log_{10}\left(\frac{R}{R_{base}}\right) + 4\left(\Phi_N - \Phi_{N,base}\right), \tag{3}$$

$$\Delta \log R_D = \log_{10}\left(\frac{R}{R_b}\right) - 2.5\left(\rho_b - \rho_{b,base}\right), \tag{4}$$

where ΔlogR_N and ΔlogR_D are the logarithmic deviation of the density and neutron log from the resistivity log, respectively, Φ_N is the neutron porosity, $\Phi_{N,base}$ is the

baseline value of neutron porosity, ρ_b is the bulk density and $\rho_{b,base}$ is the bulk density observed at the baseline. The crossover between the properly scaled well logs in *Equations. (3)–(4)* is proportional to the quantity of organic matter [2].

Several authors have proposed modification of the $\Delta \log R$ method. SONDERGELD et al. [11] observed that the traditional method tends to overestimate the value of TOC in over-matured reservoirs. They introduced an empirical correction factor, which was used to multiply the right side of *Equation (2)*. WANG et al. [21] pointed out that the mineral composition of shale reservoirs may vary between wide limits, which may result in erroneous conclusions in the evaluation of reservoirs with complex lithology. PASSEY et al. [5] modified *Equation (2)* by incorporating the acoustic traveltime of the rock matrix and cementation factor. Instead of using LOM in the Passey model, in some cases, the maximum temperature of hydrocarbon generation *(T_{max})* is recommended. Further improvements can be made by incorporating natural gamma-ray intensity readings, too [21]. RENCHUN et al. [17] changed the original equation of the Passey model to comprise simultaneously the differences of resistivity, density, and acoustic logs.

4.2 The Δd method

The method introduced by ZHAO et al. [6] is based on the joint use of the compensated neutron, density, and natural gamma-ray logs. In the frame of the method, a clay indicator is calculated from the porosity logs, which are then compared to the natural gamma-ray intensity log. Once the clay volume is known, the residual signal provides information on the organic matter content. The method is feasible in reservoirs that show increased radioactivity but low potassium content. Since the neutron log is influenced by the clay-bound porosity and the density is not, the difference between them gives information on the clay volume. The clay indicator is calculated as

$$I_{cl} = \Phi_{N,a} - \Phi_{D,a}\,, \tag{5}$$

where $\Phi_{N,a}$ (v/v) and $\Phi_{D,a}$ (v/v) are the apparent neutron and density porosity, respectively. The *GR* and I_{cl} logs are plotted on the same track. In a formation with low organic content, the two logs overlap, while in organic-rich intervals they are separated from each other. The degree of separation is measured by the following derived parameter

$$\Delta d = \frac{GR - GR_{\min}}{GR_{\max} - GR_{\min}} - \frac{I_{cl} - I_{cl,\min}}{I_{cl,\max} - I_{cl,\min}}\,, \tag{6}$$

where GR_{min} and GR_{max} are the extreme values of the gamma-ray log, I_{min} and I_{max} are those of the clay indicator. According to ZHAO et al. [6], the parameter Δd de-

fined in *Equation (6)* is linearly proportional to TOC, where the regression coefficients must be estimated for the given measurement area.

4.3 Advanced methods for shale gas evaluation

Substantial expansion has occurred in the use of Nuclear Magnetic Resonance (NMR) logging for the exploration of conventional hydrocarbon formations. The application of NMR, which is used to measure the effective porosity, has begun in the case of unconventional reservoirs as well. HOOK et al. [22] suggested the combined use of NMR and compensated density logs. The organic matter is displayed as part of the total porosity on the density log, while the NMR porosity does not indicate organic matter. The TOC can thus be inferred from the difference between the two porosity logs. The simultaneous use of NMR and sonic logs was suggested by JACOBI et al. [23]. WOLANSKI et al. [24] added spectral gamma-ray logs for the evaluation of TOC, which was estimated as the difference between the NMR-derived and geochemical log-derived porosities.

Artificial intelligence methods have an increasingly important role in the determination of organic matters. Neural network approaches have been efficiently used for the calculation of TOC in several cases [25], [26], [27], [28]. The use of neural networks assumes a complex nonlinear relation between the well logs and TOC. The effectiveness of the problem solution depends largely on the teaching process, which establishes the connection between the well logs as input and TOC as desired output in the study area. The output is then estimated by optimizing the model parameters of the network, i.e. weight coefficients that define the contribution of input signals at each neuron organized in a multilayered network structure.

Multivariate statistical methods are also a powerful tool in the interpretation of well logs. WAWRZYNIAK-GUZ et al. [29] applied a factor analysis technique for the characterization of Paleozoic shale gas formations. The study confirmed the great heterogeneity of the investigated formations and revealed different correlations between the petrophysical properties derived from well logs and extracted factors responsible for most of the variance of the observed data. SZABÓ and DOBRÓKA [30] published an improved factor analysis method for the robust estimation of reservoir shaliness in shaly sand and carbonate reservoirs. The same methodology was successfully used in fractured dolomite aquifers to calculate the hydraulic conductivity as a related quantity to shale volume along the borehole [31]. Normally high *GR* values and low resistivity indicate low porosity shales. In the study in question, high permeability is related to argillaceous shales with high porosity. This is because of the oriented texture of the rock upon which the clay formations are considered permeable. However, this does not follow from neither the *GR* nor the Φ_N log. The resistivity tools with different penetration in the fractured formation indicate the permeable rock. The joint interpretation of nuclear and resistivity logs effectively reveals the reservoir zones by means of factor analysis. The improvement of the resistivity forward problem is also of high importance to better

model the response of the resistivity tools and to solve the related inverse problem more accurately [32], [33], [34].

Cluster analysis may also help in the identification and classification of shale reservoirs. SZABÓ et al. [35] developed a hierarchical cluster analysis assisted global inversion technique to estimate the layer thicknesses and zone parameters in conventional hydrocarbon reservoirs. The rock matrix of unconventional reservoirs is generally made up of several minerals, the relative amounts of which should be estimated as accurately as possible. The models illustrated in *Figure 1* include several other parameters to be determined, too. Local (depth-by-depth) inversion techniques traditionally used in well-log analysis can give an estimate for fewer petrophysical parameters than data measured in the point. In case of complex reservoirs, there are many more petrophysical and zone parameters than data, which leads to an ambiguous (underdetermined) inverse problem. To increase the data-to-unknowns ratio and stability, we suggest the use of the interval inversion method, which inverts a data set of a longer depth interval jointly. By using a series expansion-based discretization approach, the inverse problem is solved for a far lower number of inversion unknowns than data to reduce the estimation error of the model parameters. An application to multimineral hydrocarbon formations was published by DOBRÓKA et al. [36].

5. CASE STUDIES

5.1 Estimation of total organic content

We improve the Δd method and test it on the same well-logging data set originally used by ZHAO et al. [6], which was collected in the Sichuan Basin, South-West China. The investigated formation is mainly made up of calcareous shale with a thickness varying between 20–120 m. We analyze the natural gamma-ray intensity *(GR)*, compensated density *(DEN)* and neutron-porosity *(CEN)* logs for estimating the TOC, which is approximately 1–3% according to core measurements. After calculating the apparent density- and neutron-porosities, we calculate the values of Δd along the well. For this, we use the following zone parameters: GR_{min} = 108, GR_{max} = 550, $I_{cl,min}$ = –0.22, $I_{cl,max}$ = 0.8. We study the relation between the TOC measured on core samples and parameter Δd estimated by *Equation (6)*. By plotting the TOC and Δd values, *Figure 2* shows that the relation is much more nonlinear than linear, as originally suggested by [6]. For a more accurate interpretation, especially in the domain of large Δd values, we suggest the use of the regression model

$$\mathrm{TOC} = \alpha\left(1 - \beta\, \mathrm{e}^{-\gamma \Delta d^{\eta}}\right), \tag{7}$$

where α, β, γ, η are site-specific regression coefficients which should be determined at the well site. The optimal values of regression coefficients are estimated

by the minimization of the misfit between the observed TOC values and those calculated by *Equation (7)*. We use the Marquardt algorithm for solving the least squares inverse problem [37]. For the Sichuan Basin, we obtain the following regression coefficients: $\alpha = 4.29$, $\beta = 0.85$, $\gamma = -55$, $\eta = 3$.

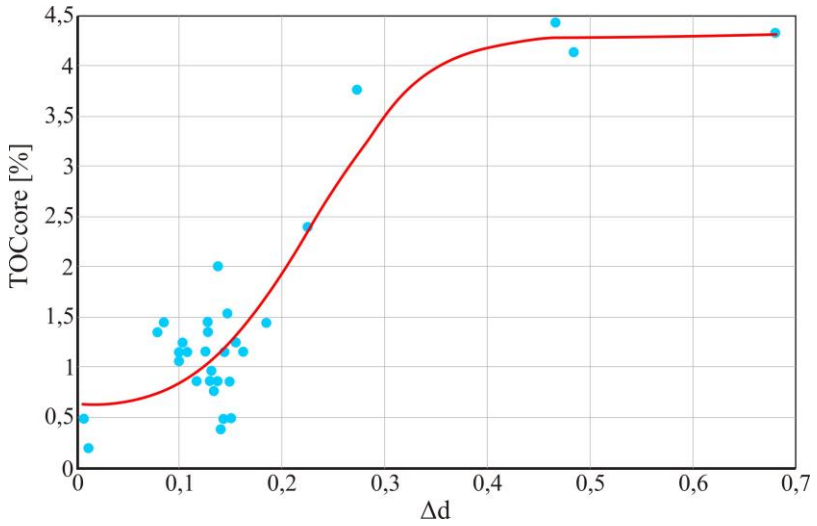

Figure 2.
Nonlinear regression relation between total organic matter content and parameter Δd estimated by *Equation (6)* in the Sichuan Basin

The result of well log analysis is illustrated in *Figure 3*. The separation between the *GR* and I_{cl} curves is directly proportional to the amount of organic matter in the shales. A potential shale gas zone is indicated by the well logs between 710 and 805 m. The highest values of TOC are around 790–800 m, where it increases up to 5%. The reservoir is also confirmed by high *GR*, low *DEN* and *CN* values.

5.2 Determination of maturity level

We test the Δlog*R* method using acoustic *(AC)*, natural gamma-ray intensity *(GR)* and deep resistivity *(RES)* logs measured in the Barnett shale, Texas, USA [7]. The Barnett Shale is a siliceous shale, calcareous and dolomitic sedimentary formation, which is rich in silicates (30–50%) but poor in clay minerals (<30%). Its average depth is 8,500 feet, and the thickness varies between 50 and 400 feet. The estimated gas content of the formation is significant, it is approximately $5.78 \cdot 10^9$ m^3 [38].

The method allows the estimation of TOC by using *Equation (2)*, which requires preliminary knowledge of the level of maturity. We use TOC data determined in the laboratory for calculating the optimal value of LOM. We introduce

the data distance as the root mean square error between the measured TOC values and the calculated ones

$$RMS = \left[\frac{1}{N} \sum_{i=1}^{N} \left(TOC_i^{meas} - TOC_i^{calc} \right)^2 \right]^{1/2} \times 100(\%). \tag{8}$$

where N is the number of core data. The RMS versus LOM relation for Barnett shale is given in *Figure 5*.

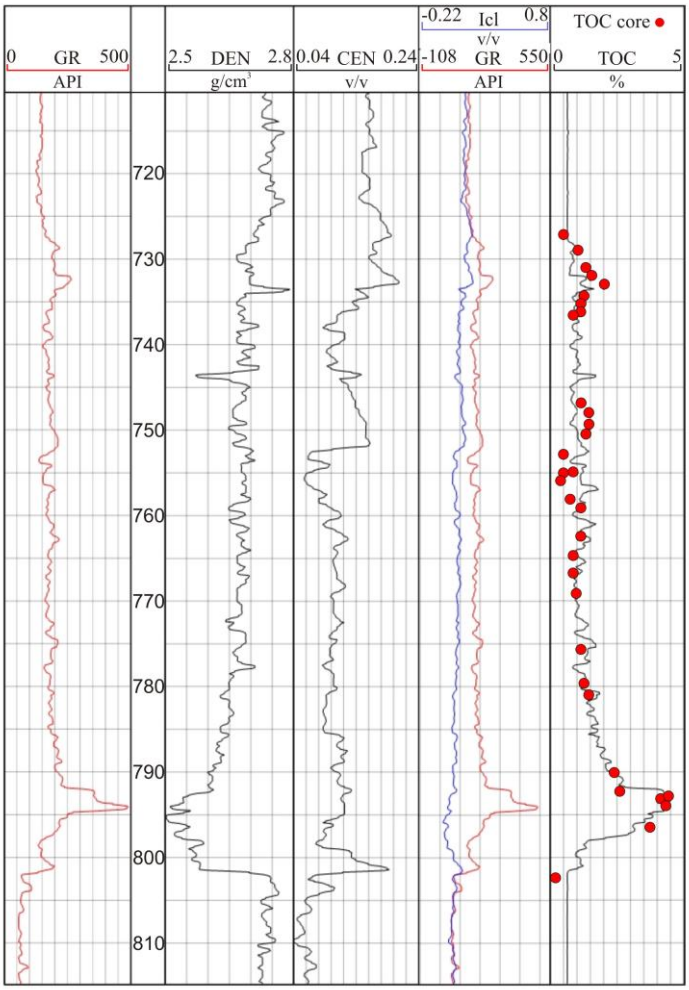

Figure 3.
Total organic matter content estimation by the improved Δd method in the Sichuan Basin

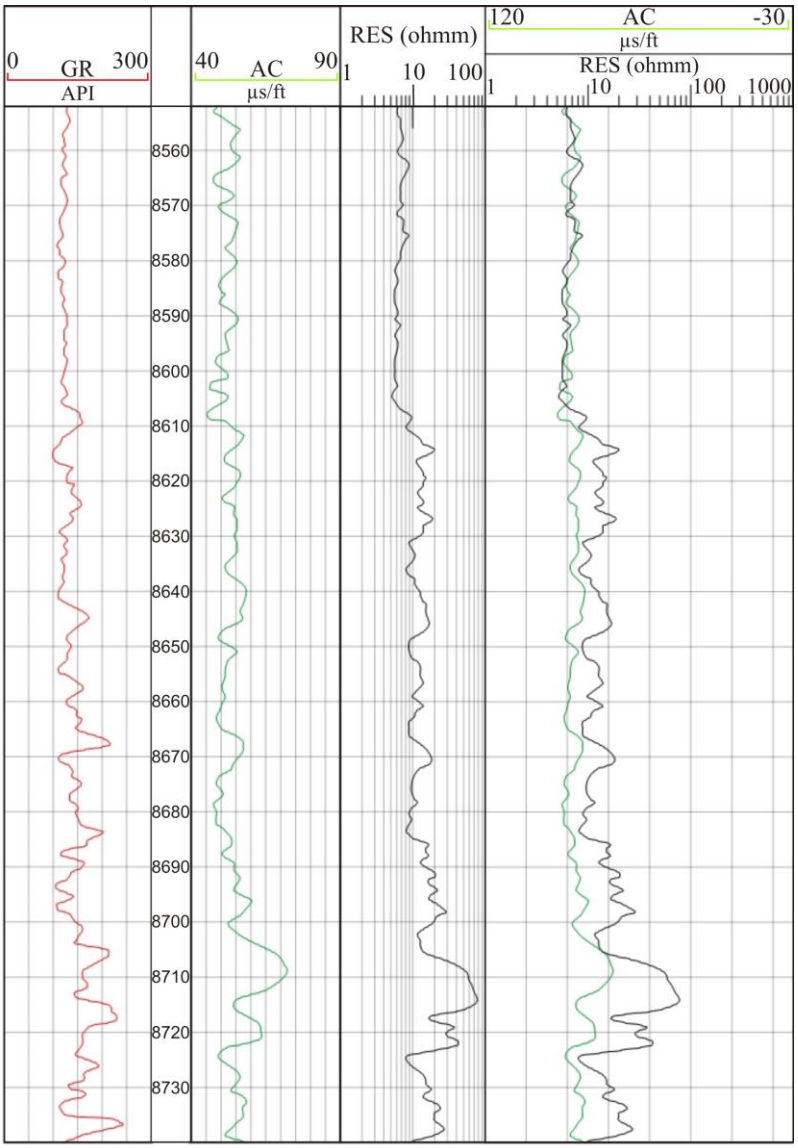

Figure 4.
Application of the Δlog*R* method on well logs measured in the Barnett Shale

The optimal value of LOM is given at the minimum of the RMS curve, which should be used in *Equation (2)* to estimate the TOC. By reading the baseline values on the sonic (80 µs/ft) and resistivity (6 ohmm) logs *(Figure 4)*, we calculate the well log of parameter Δlog*R* using *Equation (1)*. The TOC is calculated at different values of LOM *(Figure 6)*. The figure confirms that the best fit between the meas-

ured and calculated TOC is given by LOM = 8.4. The result agrees well with data in the literature, e.g. CRAIN [39] suggests a LOM ≅ 8.5 for shale gas reservoirs.

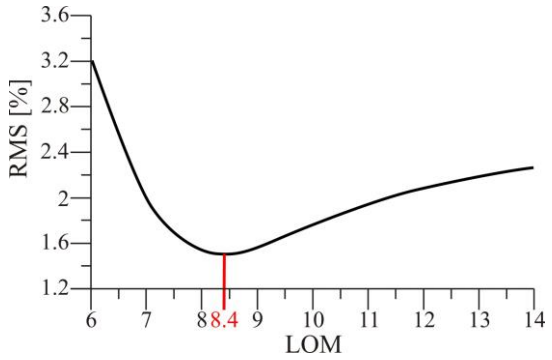

Figure 5.
Determination of maturation level by the minimization of data prediction error

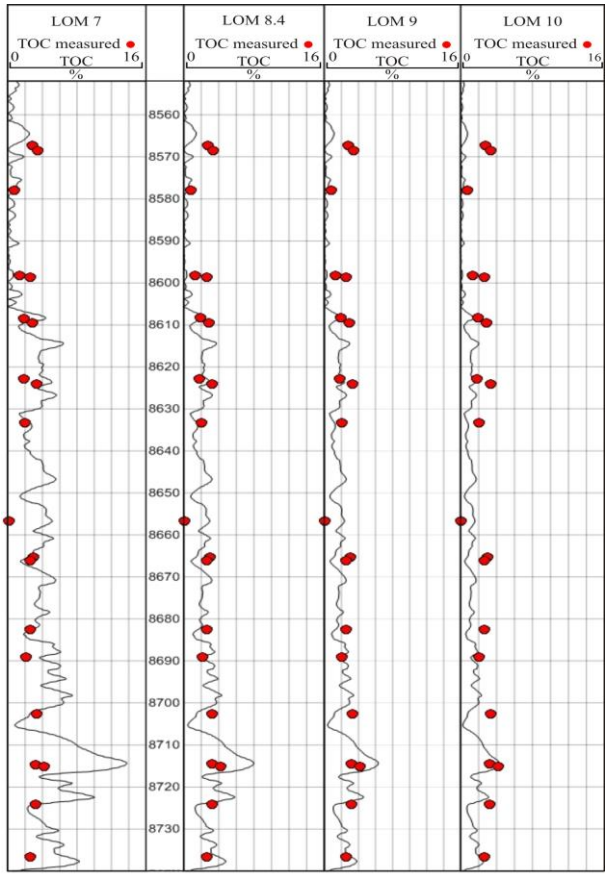

Figure 6.
Observed and calculated organic matter content at different maturity levels in the Barnett Shale

6. CONCLUSIONS

In this paper, an overview is given on the evaluation of shale gas reservoirs using well-logging methods. We suggest two improved methods for a more accurate estimation of total organic matter content. A nonlinear regression model is assumed between the organic matter content and parameter Δd suggested previously by ZHAO et al. [6]. The level of maturity should be optimized for a better estimation of organic content. We suggest reducing the misfit between the TOC values calibrated by lab measurement and predicted TOC by the Passey model. We recommend the use of multivariate statistical and inversion methods to explore the lithologic and petrophysical characteristics from independent sources, which may help in resolving the ambiguity frequently encountered with the interpretation of unconventional formations. A robust form of cluster and factor analysis may help in the identification and reserve calculations. Factor analysis explores the connections between the well-logging data set and the petrophysical properties of shale gas formations. By the interval inversion of well logs, it becomes possible to an estimate for the petrophysical parameters of unconventional reservoirs. The added advantage of the inversion approach is that it can give an estimate also for the layer thicknesses and zone parameters, including their estimation errors.

ACKNOWLEDGEMENTS

This research was supported by the GINOP-2.3.2-15-2016-00010 "Development of enhanced engineering methods with the aim at utilization of subterranean energy resources" project in the framework of the Széchenyi 2020 Plan, funded by the European Union, cofinanced by the European Structural and Investment Funds. Special thanks for providing the data set go to Mr. Peiqiang Zhao (Sichuan Basin) and Mr. Michael Holmes (Barnett Shale).

LIST OF SYMBOLS

Symbol	Description	Unit
a	Regression constant	–
AC	Acoustic traveltime data	μs/ft
b	Regression constant	–
CEN	Neutron-porosity data	v/v
DEN	Compensated density data	g/cm^3
GR	Natural gamma-ray intensity data	API
GR_{min}	Minimal values of the gamma-ray log	API
GR_{max}	Maximal values of the gamma-ray log	API
I_{cl}	Clay indicator	v/v
$I_{cl,min}$	Minimal values of the clay indicator	v/v
$I_{cl,max}$	Maximal values of the clay indicator	v/v

Symbol	Description	Unit
LOM	Level of Organic Maturity	–
NMR	Nuclear Magnetic Resonance log	–
R	Resistivity curves	ohmm
R_b	Resistivity reading at the baseline	ohmm
RES	Deep resistivity data	ohmm
T_{max}	Maximum temperature of hydrocarbon generation	°C
TOC	Total Organic Content	%
TOC^{meas}	Measured value of the Total Organic Content	%
TOC^{calc}	Calculated value of the Total Organic Content	%
α	Site-specific regression coefficient	–
β	Site-specific regression coefficient	–
γ	Site-specific regression coefficient	–
Δt	Acoustic interval-time data	μs/ft
Δt_b	Acoustic traveltime reading at the baseline	μs/ft
$\Delta \log R_N$	Logarithmic deviation between the neutron and resistivity logs	–
$\Delta \log R_D$	Logarithmic deviation between the density and resistivity logs	–
η	site-specific regression coefficient	–
ρ_b	Bulk density data	g/cm^3
$\rho_{b,base}$	Bulk density at the baseline	g/cm^3
Φ	Porosity	v/v
Φ_N	Neutron porosity data	v/v
$\Phi_{N,base}$	Baseline value of neutron porosity log	v/v
$\Phi_{N,a}$	Apparent neutron-porosity data	v/v
$\Phi_{D,a}$	Apparent density-porosity data	v/v

REFERENCES

[1] MA, Y.: Unconventional Resources from Exploration to Production. In: MA, Y.–HOLDITCH S. (ed.): *Unconventional Oil and Gas Resources Handbook*. Elsevier, Waltham, 2016, 3–52.

[2] BJØRLYKKE, K.: Unconventional Hydrocarbons: Oil Shales, Heavy Oil, Tar Sands, Shale Oil, Shale Gas and Gas Hydrates. In: BJØRLYKKE, K. (ed.): *Petroleum Geoscience – From Sedimentary Environments to Rock Physics*. Springer-Verlag, Berlin, 2015, 581–590.

[3] MEHDI, L.–REZA, R.: Petrophysical Evaluation of Gas Shale Reservoirs. In: REZA, R. (ed.): *Fundamentals of Gas Shale Reservoirs*. Wiley, New Jersey, 2015, 117–138.

[4] STARK, M.–ZHAO, S.–PEREIRA, P.–HEYES, N.–DENG, D.–BLANCO, F.–CHUDZIAK, J.–DAVID, M.–NOUWAILATI, K.–KORNELIUS, S.: *International Development of Unconventional Resources: If, where and how fast?* Accenture PLC, Dublin, 2014.

[5] PASSEY, Q. R.–CREANEY, S.–KULLA, B. J.–MORETTI, F. J.–STROUD, J. D.: A Practical Model for Organic Richness from Porosity and Resistivity Logs. *AAPG Bulletin*, 1990, 74(12), 1777–1794.

[6] ZHAO, P.–MAO, Z.–HUANG, Z.–ZHANG, C.: A new method for estimating total organic carbon content from well logs. *AAPG Bulletin*, 2016, 100(8), 1311–1327.

[7] HOLMES, M.–HOLMES, A.–HOLMES, D.: A Petrophysical Model for Shale Reservoirs to Distinguish Macro Porosity, Free Shale Porosity, and TOC. *AAPG ACE*, 2012.

[8] BEERS, R. F.: Radioactivity and Organic Content of Some Paleozoic Shales. *AAPG Bulettin*, 1945, 29(1), 1–22.

[9] AUTRIC, A.–DUMESNIL, P.: Resistivity, Radioactivity And Sonic Transit Time Logs To Evaluate The Organic Content Of Low Permeability Rocks. *The Log Analyst*, 1985, 26(03), 36–45.

[10] SCHMOKER, J.: Determination of Organic-Matter Content of Appalachian Devonian Shales from Gamma-Ray Logs. *AAPG Bulletin*, 1980, 65(7), 1285–1298.

[11] SONDERGELD, C. H.–NEWSHAM, K. E.–COMISKY, J. T.–RICE, M. C.–RAI, C. S.: Petrophysical Considerations in Evaluating and Producing Shale Gas Resources. *SPE Unconventional Gas Conference*, Pittsburgh, Pennsylvania, USA, 2010, 23–25 February.

[12] WITKOWSKY, J.–GALFORD, J.–QUIREIN, J.–TRUAX, J.: Predicting Pyrite and Total Organic Carbon from Well Logs for Enhancing Shale Reservoir Interpretation. *SPE Eastern Regional Meeting*, Lexington, Kentucky, USA, 2012, 3–5 October.

[13] MENDELZON, J. D., TOKSÖZ, M. N.: Source Rock Characterisation Using Multivarivariate of Log Data. *SPWLA 26th Annual Logging Symposium*, Dallas, Texas, USA, 1985, 17–20 June.

[14] FERTL, W. H.–CHILINGAR, G. V.: Total organic carbon content determined from well logs. *SPE Formation Evaluation*, 1988, 3(02), 407–419.

[15] SCHMOKER, J.: Determination of Organic Content of Appalachian Devonian Shales from Formation-Density Logs. *AAPG Bulletin*, 1979, 63(9), 1504–1509.

[16] MEYER, B.L.–NEDERLOF, M.H.: Identification of source rocks on wireline logs by density/ resistivity and sonic transit time/resistivity cross plots. *AAPG Bulletin*, 1984, 68(2), 121–129.

[17] RENCHUN, H.–YAN, W.–SIJIE, C.–SHUAI, L.–LI, C.: Selection of logging-based TOC calculation methods for shale reservoirs: A case study of the Jiaoshiba shale gas field in the Sichuan Basin. *Natural Gas Industry*, 2015, 2(2–3), 155–161.

[18] HERRON, S. L.–TENDRE, L.: Wireline Source-Rock Evaluation in the Paris Basin. In: HUC, A. Y. (ed.): *Deposition of Organic Facies*. AAPG Studies in Geology 30, 1990, 57–71.

[19] DELLENBACH, J.–ESPITALIE, J.–LEBRETON, F.: *Source rock logging: Transactions of the 8th European SPWLA Symposium*. 1983.

[20] CARPENTIER, B.–HUC, A.Y.–BESSEREAU, G.: Wireline Logging and Source Rocks – Estimation of Organic Carbon Content by the CARBOLBG@ Method. *The Log Analyst*, 1991, 32(03), 279–297.

[21] WANG, P.–CHEN, Z.–PANG, X.–HU, K.–SUN, M.–CHEN, X.: Revised models for determining TOC in shale play: Example from Devonian Duvernay Shale, Western Canada Sedimentary Basin. *Marine and Petroleum Geology*, 2015, 70, 304–319.

[22] HOOK, P.–FAIRHURST, D.–RYLANDER, E.–BADRY, R.–BACHMAN, N.–CRARY, S.–CHATAWANICH, K.–TAYLOR, T.: Improved precision magnetic resonance acquisition; application to shale evaluation. *SPE Annual Technical Conference and Exhibition 30*, Denver, Colorado, USA, 2011.

[23] JACOBI, D. J.–BREIG, J. J.–LECOMPTE, B.–KOPAL, M.–HURSAN, G.–MENDEZ, F.–BLIVEN, S.–LONGO, J.: Effective geochemical and geomechanical characterization of shale gas reservoirs from the well bore environment: Caney and the Woodford Shale. *SPE Annual Technical Conference and Exhibition*, New Orleans, Louisiana, USA, 2009, 4–7 October.

[24] WOLANSKI, K.–OCHALIK, S.–BARANKIEWICZ, P.–KOZLOWSKI, M.–VICAN, K.–BULLER, D.–BALLIET, R.: A Baltic Basin Shale Reservoir Integrated Evaluation Using a Wireline Geochemical, NMR, and Acoustic Dipole Workflow. *SPWLA 56th Annual Logging Symposium*, 2015, 18–22 July.

[25] HUANG, Z.–WILLIAMSON, A. M.: Artificial neural network modelling as an aid to source rock characterization. *Marine and Petroleum Geology*, 1995, 13(2), 277–290.

[26] KHOSHNOODKIA, M.–MOHSENI, H.–RAHMANI, O.–MOHAMMADI, A.: TOC determination of Gadvan Formation in South Pars Gas field, using artificial intelligent systems and geochemical data. *Journal of Petroleum Science and Engineering*, 2011, 78(1), 119–130.

[27] OUADFEUL, S.–ALIOUANE, L.: Total Organic Carbon Prediction in Shale Gas Reservoirs from Well Logs Data Using the Multilayer Perceptron Neural Network with Levenberg Marquardt Training Algorithm: Application to Barnett Shale. *Arabian Journal for Science and Engineering*, 2015, 40(11), 3345–3349.

[28] SHI, X.–WANG, J.–LIU, G.–YANG, L.–GE, X.–JIANG, S.: Application of extreme learning machine and neural networks in total organic carbon content prediction in organic shale with wire line logs. *Journal of Natural Gas Science and Engineering*, 2016, 33, 687–702.

[29] WAWRZYNIAK-GUZ, K.–JARZYNA, J. A.–ZYCH, M.–BAŁA, M.–KRAKOWSKA, P. I.–PUSKARCZYK, E.: Analysis of the heterogeneity of the Polish shale gas formations by factor analysis on the basis of well logs. *78th EAGE Conference and Exhibition*, Extended Abstracts, 2016, paper Tu SBT3 07, 1–5.

[30] SZABÓ, N. P.–DOBRÓKA, M.: Robust estimation of reservoir shaliness by iteratively reweighted factor analysis. *Geophysics*, 2017, 82(2), D69–D83.

[31] SZABÓ, N. P.: Hydraulic conductivity explored by factor analysis of borehole geophysical data. *Hydrogeology Journal*, 2015, 23, 869−882.

[32] GALSA, A.–HEREIN, M.–DRAHOS, D.–HEREIN, A.: Effect of the eccentricity of normal resistivity borehole tools on the current field and resistivity measurement. *Journal of Applied Geophysics*, 2016, 134, 281–290.

[33] FALCO, P.–NEGRO, F.–SZALAI, S.–MILNES, E.: Fracture characterisation using geoelectric null-arrays. *Journal of Applied Geophysics*, 2013, 93, 33–42.

[34] SZALAI, S.–SZARKA, L.–PRÁCSER, E.–MÜLLER, I.–TURBERG, P.: Geoelectric mapping of near-surface karstic fractures by using null-arrays. *Geophysics*, 2002, 67, 1769–1778.

[35] SZABÓ, N. P.–DOBRÓKA, M.–KAVANDA, R.: Cluster analysis assisted float-encoded genetic algorithm for a more automated characterization of hydrocarbon reservoirs. *Intelligent Control and Automation*, 2013, 4, 362–370.

[36] DOBRÓKA, M.–SZABÓ, N. P.–TÓTH, J.–VASS, P.: Interval inversion approach for an improved interpretation of well logs. *Geophysics*, 2016, 81, D163–D175.

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

[38] MONTGOMERY, L. S.–JARVIE, M. D.–BOWKER, A. K.–POLLASTRO, M. R.: Mississippian Barnett Shale, Fort Worth basin, north-central Texas: Gas-shale play with multi–trillion cubic foot potential. *AAPG Bulletin*, 2005, 89(2), 155–175.

[39] CRAIN, E., R.: *Crain's Petrophysical Handbook*. https://spec2000.net/index.htm, 2016. 09. 12.

Geosciences and Engineering, Vol. 5, No. 8 (2016), pp.51–64.

LITHOLOGY DETERMINATION IN A COAL EXPLORATION DRILLHOLE USING STEINER WEIGHTED CLUSTER ANALYSIS

BENCE Á. BRAUN[1]–ARMAND ABORDÁN[1,*, 2]–NORBERT P. SZABÓ[1, 2]

[1]*Department of Geophysics, University of Miskolc*
**armand.abordan@yahoo.com*
[2]*MTA-ME, Geoengineering Research Group,*
3515, Miskolc-Egyetemváros, Hungary

Abstract: For a successful exploration project, the determination of lithology is crucial. For determining the lithology of a coal exploration drillhole, a non-hierarchical cluster analysis is presented in which the most commonly used measure of distance, the Euclidean distance, is altered by Steiner weights. Using the Steiner weighted Euclidean distance makes the method more robust, thus affected less by outliers. We compare the results of the improved method of cluster analysis (MFV-CA) with the interpreted borehole geophysical logs, mud logs and the traditional cluster analysis based on the Euclidean distance (TCA). The comparative study shows that the new method gives better resolution and is less affected by outliers than the Euclidean distance based cluster analysis, and may give additional information to the interpretation of geophysical well logs and mud logs.

Keywords: *cluster analysis, most frequent value, Steiner weight, well-logging methods, coal exploration*

1. INTRODUCTION

The exploration and production of hydrocarbons, fresh water, ores, and coals are becoming more important as humankind is advancing. The main goal of applied earth sciences nowadays is to find, accurately locate and estimate the volume and quality of these raw materials in the subsurface. For this purpose, many geophysical and geological methods can be used, two of which are utilized in this paper, i.e. well logging and geostatistics. The combination of these tools for data processing and interpretation is already well known in earth sciences. In this paper, a novel approach for the modification of a multivariate geostatistical method is presented.

Cluster analysis, a frequently used multivariate statistical method, is applied to classify objects based on a set of measured variables into a number of different groups by placing similar subjects in the same group. In this paper, the traditional cluster analysis (TCA), which is based on the Euclidean distance, is modified by incorporating the most frequent value (MFV) method [1]. Modifying the Euclidean distance with Steiner weights makes the procedure robust and therefore less affected by outlying data. The improved statistical method is tested and compared with the interpreted well logs, mud log and the TCA on two well logs measured from a coal

exploration drillhole. In both cases, the cluster analysis modified by Steiner weights (MFV-CA) proves to be superior to the conventional cluster analysis.

2. TRADITIONAL METHODS OF CLUSTER ANALYSIS

Cluster analysis is a multivariate statistical method that aims to group data objects into groups based only on the information found in the data that describe the objects and their relationships. The goal is to collect the objects into groups in a way that the objects within a group are alike and are very different from objects outside the group. The greater the homogeneity within a group and the bigger the difference between groups, the better the clustering is. The objects are grouped based on some defined distance metric.

Based on the above definition, it can be seen that if the distance between two objects is small then they are similar and should be placed in the same group; if the distance between two objects is great then they are different and cannot be put into the same group. If the distance between two objects is too great (outliers), then the non-resistant nature of the method should be taken into consideration, meaning that the method is greatly affected by outliers, because in this case data with different orders of magnitude may greatly influence the result of the estimation [2]. A datum can be considered an outlier if it is orders of magnitude greater or smaller than the average value range of the dataset and it shows a sudden change in the dataset such as a Dirac delta function. The goal of clustering is to have minimal distances within groups and maximal distances between groups [3].

Let the vectors $\boldsymbol{x}^{(i)}$ *and* $\boldsymbol{x}^{(j)}$ denote two multivariate observations from a population with p random variables $X_1,\ldots,X_p$. In well log analysis, X_i denotes a physical variable measured in the borehole by the i-th logging tool. In a more detailed form, the i-th and j-th observations are $\mathbf{x}^{(i)} = \left\{x_1^{(i)}, \ldots, x_p^{(i)}\right\}^{\mathrm{T}}$ and $\mathbf{x}^{(j)} = \left\{x_1^{(j)}, \ldots, x_p^{(j)}\right\}^{\mathrm{T}}$, which represent two so-called objects in the data space, respectively. In order to group the objects into clusters a measure for the similarity of elements needs to be defined. To determine the similarity between two objects, distance measures can be used. The TCA uses the Euclidean distance

$$D_e\left(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}\right) = \sqrt{\{(\mathbf{x}^{(i)} - \mathbf{x}^{(j)})^{\mathrm{T}}(\mathbf{x}^{(i)} - \mathbf{x}^{(j)})\}}. \tag{1}$$

By weighting it with the covariance matrix, we get the Mahalanobis distance

$$D_m\left(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}\right) = \sqrt{\{(\mathbf{x}^{(i)} - \mathbf{x}^{(j)})^{\mathrm{T}}\mathbf{S}^{-1}(\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\}}, \tag{2}$$

where $\mathbf{S} = \mathbf{C}^{\mathbf{T}}\mathbf{C}/(n-1)$ is the covariance matrix derived from the standardized data matrix $\mathbf{C}$.

There are several different methods that can be used for cluster analysis. The main difference among the different types is whether the set of clusters is hierarchical or non-hierarchical. The main property of hierarchical clustering is that it permits clusters to have sub-clusters. It also has the advantage over non-hierarchical clustering that prior to the analysis it does not require the specific number of clusters to be created during the process. The main drawback of hierarchical clustering is that it needs a lot of computing power; therefore, it is suitable only for smaller data sets. It can further be divided into agglomerative and divisive methods. In case of the agglomerative method each object starts as an individual cluster, then the two closest are combined repeatedly until all objects are in the same cluster. After calculating all cluster solutions, the optimum number of clusters can be chosen. On the contrary, the divisive method starts with one cluster and with the same strategy as the above method in a reverse order separates objects until all are separated.

The non-hierarchical or partitioning clustering is fundamentally different from hierarchical clustering. This kind of clustering requires the specific number of clusters to be created prior to the process. It is done by a partitioning algorithm and the number of objects in each cluster is computed during the process. It has the feature of handling big data sets well; however, the result is affected by the initial selection of centroids.

The SSE (Sum of the squared error) can be used for estimating the optimal number of clusters. It measures the distance of each data point to the closest centroid and then calculates the total sum of squared errors

$$SSE = \sum_{k=1}^{K} \sum_{i=1}^{n_k} d^2 \left(\mathbf{c}_k, \mathbf{x}_k^{(i)} \right), \tag{3}$$

where d gives the distance between the k-th vector belonging to the k-th cluster (with n_k elements) and k-th centroid $\boldsymbol{c}_k$ ($k = 1,2,..,K$). The centroid is calculated by the mean of objects forming a cluster

$$\mathbf{c}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} \mathbf{x}_k^{(i)} \quad . \tag{4}$$

The value of SSE also describes scatter, which reaches a limit with the increasing number of clusters [2]. For the optimal solution, the smallest SSE value and the smallest number of clusters should be chosen because choosing a greater value hardly adds any more information but it makes the interpretation more difficult.

From the non-hierarchical clustering methods, the most commonly used is K-means clustering *(Figure 1)*. A prototype-based technique attempts to find a user-specified number of clusters *(K)*, which are represented by their centroids. The first step is to assign K initial centroids, where K is a user-specified parameter then each point is assigned to the closest centroid and each collection of points assigned

to a centroid is a cluster. The centroids are recalculated based on the points that are assigned to the clusters. We repeat this until all centroids remain the same, as seen in *Figure 1* [3].

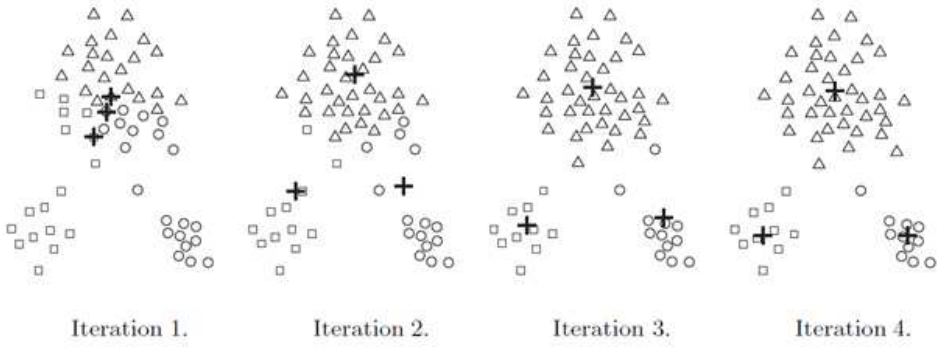

Figure 1
K-means algorithm used to find three clusters in the sample data

3. MOST FREQUENT VALUE BASED CLUSTER ANALYSIS

The most frequent value method is based on the weighting of data, calculating the weighted average. The weighted average is in most cases more advantageous than the arithmetic means because of its robustness (being less affected by outliers). The goal of cluster analysis is to have minimal distance within the clusters and maximal among the clusters, therefore it is necessary to weight data more that are closer to each other and weight less those which are farther [4].

The next equation is assumed to have high values of n and we have a symmetric density function. The weighted average is calculated with the $\varphi(x)$ symmetric weight function, where T denotes the point where the $\varphi(x)$ weight function has its maximum value

$$M_n = \frac{\sum_{i=1}^{n} x_i \varphi_i}{\sum_{i=1}^{n} \varphi_i}, \tag{5}$$

$$\varphi_i = \frac{\varepsilon^2}{\varepsilon^2 + (x_i - M_n)^2}, \tag{6}$$

where $(x_i - M_n)$ denotes the differences within the cluster, φ is a characteristic weight-function and ε is the dihesion, the shape parameter of the weight function.

The M_n is called the most frequent value, which can be calculated iteratively from the combination of equations (5) and (6), because the weights are independent of M_n

[1]. Thus during the calculation of M_n we have to define M_n iteratively in such a way that it satisfies the equation

$$M_n = \frac{\sum_{i=1}^{n} \frac{\varepsilon^2}{\varepsilon^2 + (x_i - M_n)^2} x_i}{\sum_{i=1}^{n} \frac{\varepsilon^2}{\varepsilon^2 + (x_i - M_n)^2}}. \tag{7}$$

For the first iteration (k=1) M_n is substituted with the mean of the data and the dihesion ε is calculated from *Equation (8)*, the sequential number k of the iterations and the clusters beyond possibility

$$\varepsilon_1 \leq \frac{\sqrt{3}}{2}(\max(x_i) - \min(x_i)). \tag{8}$$

In the beginning, we choose a greater value for ε and then even the outlying data are assigned a relatively large weight. In the following iteration steps, the values of ε and M_n are derived from each other based on *Equations (9)* and *(10)* [1]

$$\varepsilon^2{}_{k+1} = \frac{3\sum_{k=1}^{n} \frac{(x_i - M_k)^2}{[\varepsilon_k^2 + (x_i - M_k)^2]^2}}{\sum_{i=1}^{n} \frac{1}{[\varepsilon_k^2 + (x_i - M_k)^2]^2}}, \tag{9}$$

$$M_{n,k+1} = \frac{\sum_{k=1}^{n} \frac{\varepsilon_{k+1}^2}{[\varepsilon_{k+1}^2 + (x_i - M_{n,k})^2} x_i}{\sum_{k=1}^{n} \frac{1}{[\varepsilon_{k+1}^2 + (x_i - M_{n,k})^2}}. \tag{10}$$

At the end of the procedure the small value of ε ensures that data close to M_n contribute to the result with larger weight and outliers with smaller weight or not at all. The least-squares procedure works with 100% efficiency if the error distribution is Gaussian. This is not surprising because the least-squares estimate is the best when the distribution is Gaussian. Unfortunately, its efficiency diminishes sharply to zero for error distributions having longer tails. Therefore, the least-squares principle should not be applied for any type of distribution other than the Gaussian type. In contrast, the MFV procedure is very highly efficient (>90%) regardless of the distribution type. Therefore, the general high robustness of the MFV procedure is proven [5].

Based on tests the MFV weighting improves the result of the TCA, and being a robust statistical procedure it can be applied very effectively for lithology determination. As a starting point for the Steiner weights, first we take the Euclidean distance as in *Equation (1)* and then we modify it, which gives the robust algorithm of MFV-CA as seen in *Equation (12)*. The values of the Steiner weights that are computed in the most frequent value method are given by the following equation

$$W_{ii}^{(MFV)} = \frac{\varepsilon^2}{\varepsilon^2 + (x_i - c_k)^2}, \tag{11}$$

where W_{ii} is the *i*-th diagonal element of the weight matrix, which gives the difference between the *i*-th data and *k*-th centroid, and ε is calculated in an inner iteration cycle.

Applying the Steiner weights in the TCA will improve its performance. Combining *Equations (1)* and *(11)* we get *Equation (12)*, the Euclidean distance altered by the most frequent value based weight function

$$D_{MFV}\left(\mathbf{x}_k^{(i)}, \mathbf{c}_k\right) = \left(\mathbf{x}_k^{(i)} - \mathbf{c}_k\right)^{\mathrm{T}} \mathbf{W}^{(MFV)} \left(\mathbf{x}_k^{(i)} - \mathbf{c}_k\right), \tag{12}$$

where k = 1,2,3...K. As a result, data farther from the centroid will affect the formation of clusters with smaller weight, while the contribution of the elements closest to M_n is the greatest.

4. Applied well logging tools and the lithology

To test the MFV-CA procedure and compare it to the TCA, a coal exploration drillhole in Hungary is used. Based on the mud log and the interpreted well logs the dominant rock types of the area are sandstone, siltstone, mudstone, marl, coal, carbonaceous clay, marly sandstone, silty sandstone, and diabase.

In the investigated drillhole the layer thicknesses vary from a few tens of cm to 25 m. The well logs were measured at 0.1 m intervals. The tools and the measurement interval were selected based on the highly varying lithology and layer thicknesses. The following well logging tools were selected as an input for the cluster analysis: spontaneous potential (SP), natural gamma-ray intensity (GR), shallow apparent resistivity (RES10), medium apparent resistivity (RES40), gamma-gamma (density) (DEN) and neutron porosity (NPOR). With these tools the detection of coal is relatively unambiguous, since coals show high density-porosity, high neutron-porosity, high resistivity and low natural gamma radiation, which increases with shale content [6]. Lithological boundaries are usually picked manually or with software, but cluster analysis automatically provides the layer boundaries based on the created clusters.

5. Interpretation of well logs with Steiner-weighted cluster analysis

5.1. Test on well log-1

Based on the SSE diagram *(Figure 2)*, the optimal number of clusters is 4; for more than 4 clusters the algorithm created non-existing and transitional rock types. *Figure 3* contains the interpreted well logs, mud log, TCA and the MFV-CA.

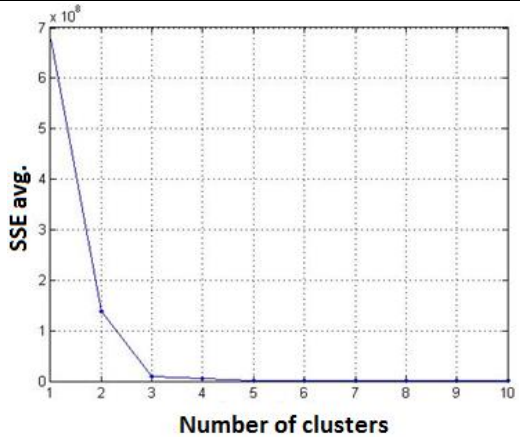

Figure 2
Optimal number of clusters for well log 1

The brightness index (BI) of coals indicate their rank; the brighter the coal the higher its carbon content and the lower its ash content is [7]. It can be seen that in the interval of 576–577.5 m the interpreted well logs and the mud log defined sandstone; however, both types of cluster analysis defined siltstone. In the depth range of 577.5–587.5 m all methods defined the same rock type, sandstone. In the interval of 587.5–588.5 m most probably coaly mudstone is present, based partly on the interpreted well logs and the cluster analysis with the modified Euclidean distance; however, the mud log suggests siltstone with interbedded coal layers and the cluster analysis based on the Euclidean distance suggests coaly siltstone. The area between 588.5–592.5 m is defined as coal by the mud log and the cluster analysis, while the well logs suggest carbonaceous clay at some intervals. In the 592.5–595.5 m interval the well logs defined coal, although both the cluster analysis and the mud log suggest interbedded carbonaceous clay layers. This reinterpretation changes the predicted lithology of this coal seam, thus the extent of the economically extractable coal seam is reduced. This might be acceptable for a coal seam of this magnitude, but in the case of a bigger coal seam or an oil field these reinterpretations can greatly influence the volume estimations and therefore the economic decisions. Therefore, it is not advisable to rely only on the interpretation of well logs and the mud log; the results of the cluster analysis should also be considered, because it numerically interprets the lithology and thus can either confirm or confute the original estimates. In the depth range of 595.5–597 m all methods suggest the presence of siltstone. The diabase starting from 597 m is interpreted as siltstone by the cluster analysis. Defining the diabase by cluster analysis is somewhat uncertain because the properties of this layer are not described well.

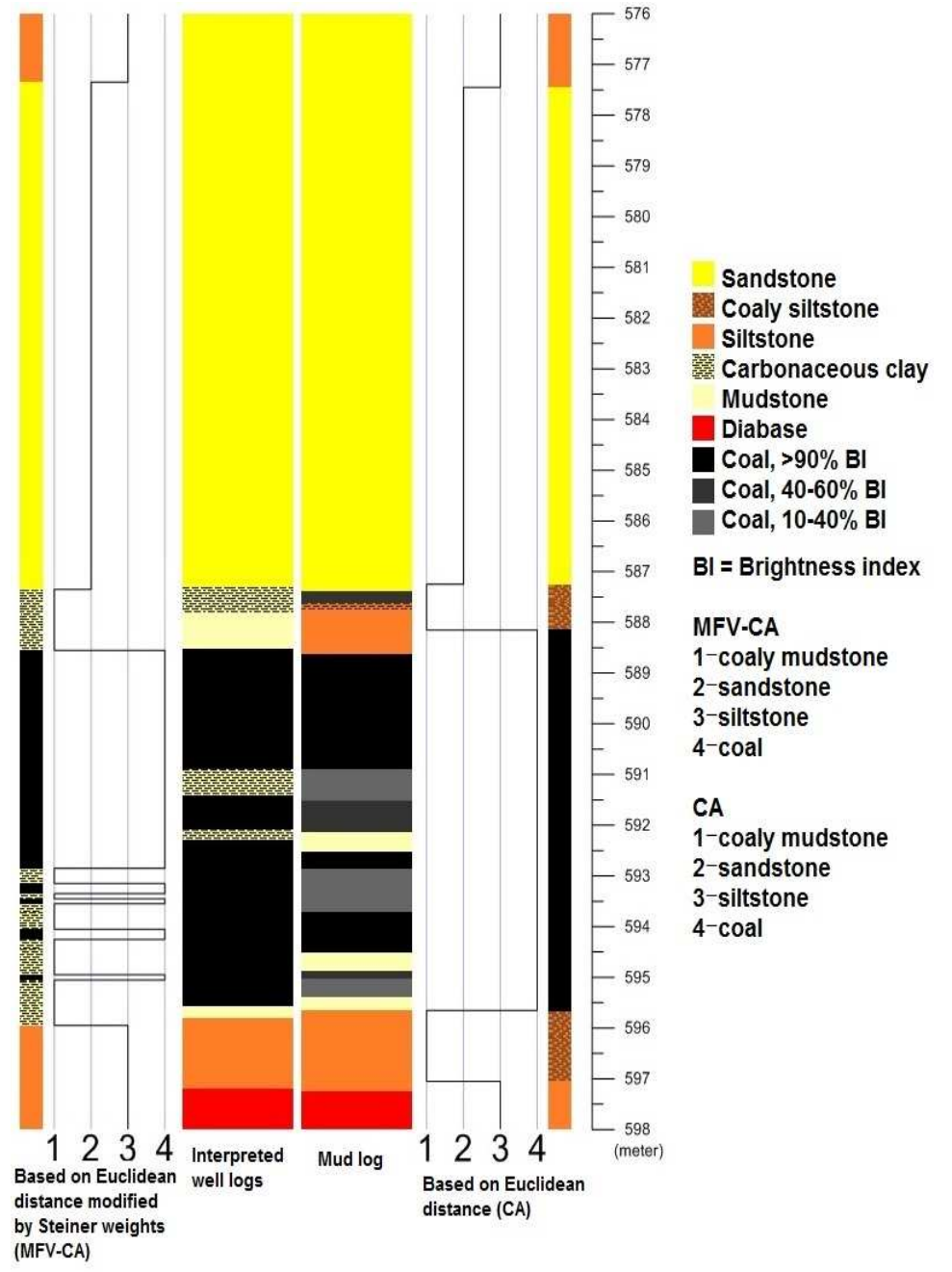

Figure 3
Results of cluster analysis for well log 1

5.2. Test on well log-2

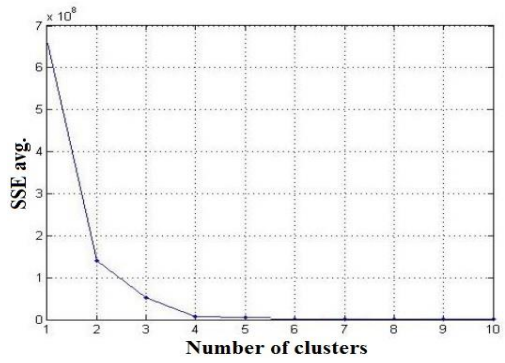

Figure 4

Optimal number of clusters for well log 2

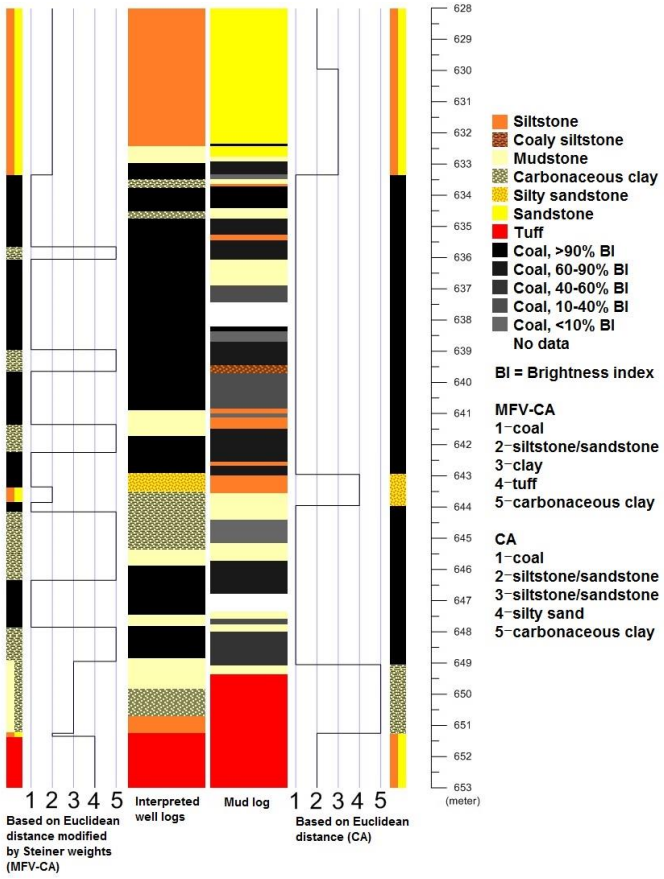

Figure 5

Results of cluster analysis for well log 2

As shown in *Figure 4*, five clusters were created during the clustering for well log 2. *Figure 5* contains the interpreted well logs, mud log, the TCA and the MFV-CA. In the interval of 628–632.5 m the rock type is either siltstone or sandstone (it is uncertain which cluster denotes which rock type; there might be grain size transition between the two). Down to 649 m, there is a thick coal seam, which is interbedded with up to 1 m thick carbonaceous clay layers, according to the cluster analysis. The interpreted well logs show fewer carbonaceous clay layers, but suggests interbedded mudstone layers up to 1.5 m thick. Based on the mud log and the cluster analysis, more carbonaceous clay layers intersect this thick coal seam, therefore the whole seam cannot be economically extracted. The tuff layer at the bottom of the log is defined by all methods but at slightly different depths. The description of the mud log only correlates at some intervals with the cluster analysis, which is probably caused by the lower rank coals in the interval that are grouped to carbonaceous clay layers.

6. Comparison of traditional cluster analysis and Steiner-weighted cluster analysis

The cluster analysis modified by Steiner weights works well in case of shorter intervals where diverse layers and thin layering is present. The SSE value can usually be used effectively to estimate the number of clusters to be created. If it proves to be insufficient, an additional cluster can be added. The shallow apparent resistivity tool at most intervals read a monotonic trend, therefore it did not affect the clustering algorithm much. The lower rank of coals is closely related to the grouping of clusters into carbonaceous clay. Where similar rock types follow each other (clay–carbonaceous clay–clay) the cluster analysis based on the Euclidean distance modified by Steiner weights exerts an adequate smoothening effect. Both clustering algorithms have difficulties differentiating sandstone and siltstone (this can be due to grain size boundary or a badly graded rock). When processed with the Euclidean distance, thicker layers are smoothed out completely and the separation of thin interbedded layers is not possible, only the depth range of the layer is acceptable. CPU time for both methods, even for big data sets, is negligible.

The main goal of these methods is to determine the lithology precisely with the minimal number of clusters. The presented cluster analysis with the modified Euclidean distance gives a more robust estimation, which later can be used as a starting model for an inversion procedure. Weighting cluster analysis based on the Euclidean distance with the MFV procedure allows us to determine the lithology with a minimal number of clusters and with good accuracy and resolution. This new procedure makes it easier to accurately map the layers or reservoirs in the subsurface, and therefore better volume estimates can be made, which has significant economic importance.

7. TEST OF ROBUSTNESS

In *Figure 6*, four different log intervals can be seen. In these logs two outliers each were added to the data set *(Table 1)*, for both the classical cluster analysis with the Euclidean distance and for the one weighted by the Steiner weights. During robustness testing we define how much a data point needs to be altered for the clustering algorithm to handle the object as a different cluster.

Comparing the results, it can be seen that the cluster analysis based on the Euclidean distance modified by the Steiner weights is more robust and has better noise rejection capability than the classical cluster analysis using the Euclidean distance, because it groups an object to a new cluster only with greater outliers. These tests verify the more robust nature of the non-hierarchical cluster analysis using the Euclidean distance modified by the Steiner weights presented in this paper. This exerts a smoothening effect on the logs where thin and similar rocks are next to each other (e.g. clay – carbonaceous clay – clay) and where an outlier is present. This smoothing is not applicable for extreme values.

Table 1

Statistics of well logs

	GR (1)	**N. POR. (2)**	**RES40 (3)**	**SP (4)**	
	Natural gamma ray	Neutron porosity	Medium apparent resistivity	Spontaneous potential	
Median	36.000	54.500	9.590	43.630	
Modus	–	55.630	9.590	41.630	
Average	36.383	52.098	9.681	43.558	
Limit of outlier	83	162	28	132	• = Euclidean distance modified by Steiner weights
Limit of outlier	55	127	21	102	• = Euclidean distance

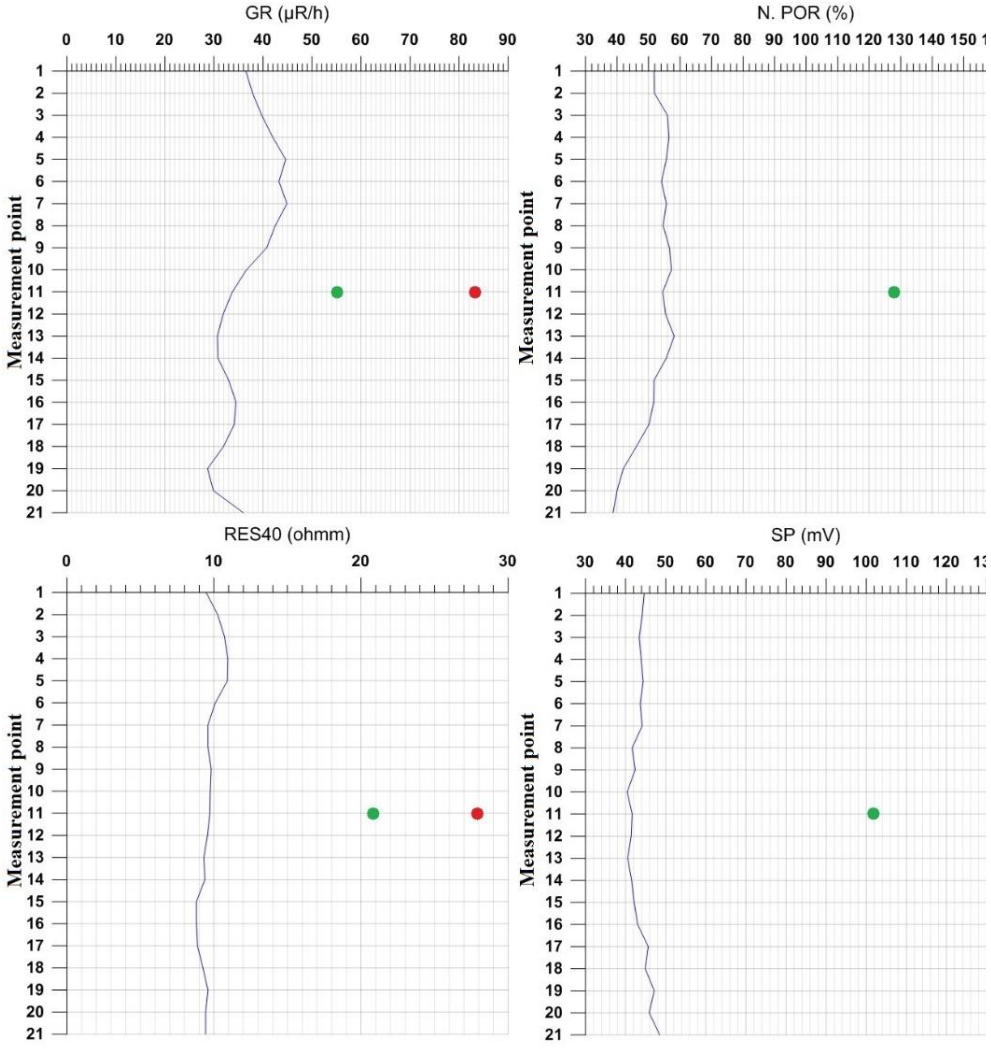

Figure 6
Upper value of outliers for the four analyzed logs

8. CONCLUSION

According to our tests, the non-hierarchical cluster analysis based on the Euclidean distance modified by the Steiner weights (MFV-CA) presented in this paper provides highly acceptable results. It is efficient, robust and can determine the lithology with good resolution. It also has good noise rejection capability, as shown in *Figure 6*. The classical cluster analysis based on the Euclidean distance (TCA) can also pro-

vide acceptable results; however, its accuracy is worse and its lithology determination is less exact. The MFV-CA method as an additional tool for lithology determination, supplementing mud log and well log analysis, can be applied to verify and specify the results of the other two methods numerically. In the future we are planning to follow up our research by applying the MFV-CA method to unconventional reservoirs.

ACKNOWLEDGEMENTS

The authors are thankful for the support and data provided by Geo-Log Geophysical & Environmental Ltd.

LIST OF SYMBOLS

Symbol	Description	Unit
$\boldsymbol{x}^{(i)}$	vector of *i*-th multivariate observation	–
$\boldsymbol{x}^{(j)}$	vector of *j*-th multivariate observation	–
X_i	*i*-th observed well logging parameter	–
D_e	Euclidean distance	–
D_m	Mahalanobis distance	–
$\mathbf{S}$	covariance matrix	–
$\mathbf{C}$	standardized data matrix	–
SSE	sum of the squared error	–
d	distance between the cluster elements and its centroid	–
$\mathbf{c}_k$	the centroid of the *k*-th cluster	–
n_k	number of objects in the *k*-th cluster	–
$\varphi(x)$	weight function in MFV procedure	–
M_n	empiric most frequent value	–
ε	dihesion	–
k	number of iterations	–
K	number of clusters	–
ε_1	initial value of dihesion	–
$W_{ii}^{(MFV)}$	*i*-th Steiner weight coefficient	–
D_{MFV}	Steiner weighted Euclidean distance	–
SP	spontaneous potential log	mV
GR	natural gamma-ray intensity log	API
RES10	resistivity log with shallow penetration	ohmm
RES40	resistivity log with medium penetration	ohmm
DEN	gamma-gamma (density) log	g/cm^3
NPOR	neutron porosity log	%

REFERENCES

[1] STEINER, F.: *The Most Frequent Value: Introduction to a Modern Conception of Statistics.* Akadémia Kiadó, Budapest, Hungary, 1991.

[2] STEINER, F.: *The Fundamentals of Geostatistics* (in Hungarian). Tankönyvkiadó, Budapest, Hungary, 1990.

[3] TAN, P. N.–STEINBACH, M.–KUMAR, V.: *Introduction to Data Mining*. Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA, 2005.

[4] REIMAN, C.–FILZMOSER, P.–GARRETT, R. G.–DUTTER, R.: *Statistical Data Analysis Explained*. John Wiley & Sons Ltd., 2008.

[5] SZŰCS, P.–CIVAN, F.–VIRAG, M.: Applicability of the most frequent value method in groundwater modeling. *Hydrogeology Journal*, January 2006, Volume 14, Issue 1, 31–43.

[6] SERRA, O.: *Fundamentals of Well-log Interpretation*. Elsevier, Amsterdam, 1984.

[7] DUTCHER, R. R.: *Field Description of Coal*. ASTM International. ASTM STP 661, Baltimore, 1978.

Geosciences and Engineering, Vol. 5, No. 8 (2016), pp.65–84.

ANALYSIS OF STRESS RELATIONS USING FOCAL MECHANISM SOLUTIONS IN THE PANNONIAN BASIN

LILI CZIROK[1]

[1] *Kitaibel Pál Environmental Science Doctoral School,*
University of West Hungary,
cziroklili@gmail.com

Abstract: An important inversion problem in seismology is the determination of principal stress patterns, which can be resolved by stress inversions, using focal mechanism solutions (FMS). In the course of my research, I carried out a linear, iterativ stress inversion in MATLAB (STRESSINVERSE, [1]) and analyzed the stress field in the Pannonian Basin, especially in Hungary and the peripheral areas (e.g. the Mur-Mürz-Zilina Zone). Because of the variability of principal stress trajectories (directions of the maximum horizontal compression) and distribution of the epicenters, I had to use several subareas for calculations. In general, my results confirm those in previously published studies, but differences occur between the results and also a priori information.

1. INTRODUCTION

The contemporary stress field of the Pannonian Basin presents a great deal of variability, both horizontally and vertically. This can be traced back to the extension in the early Miocene and tectonic inversion in the early Pliocene. These processes eventuated in the bending and plastic deformation of the lithosphere, as well as the genesis and/or reactivation of faults. Nowadays, the most important driving force is the "Adria-push", in which the Adriatic microplate rotates in a counter-clockwise direction and drifts north-northeast simultaneously. The recent stress field is dominated by compressional and strike-slip-like features [2], [3], [4], [5], [6].

The World Stress Map Project (WSMP) was created to study stress relations around the world in 1986. Recently, this project has become the most important for the analysis of stress fields. Its database contains borehole breakouts, focal mechanisms, results of overcorings and space geodetic data (GPS in particular). All data represent the directions of the maximum horizontal compression (S_{hmax}). It is possible to study the stress tectonic regimes too, mostly based on focal mechanisms solutions (FMS). Unfortunately, the distribution of these directions is diverse, so other methods for the analysis of stress patterns can be helpful too, for example numerical modelling [4], [5], [6].

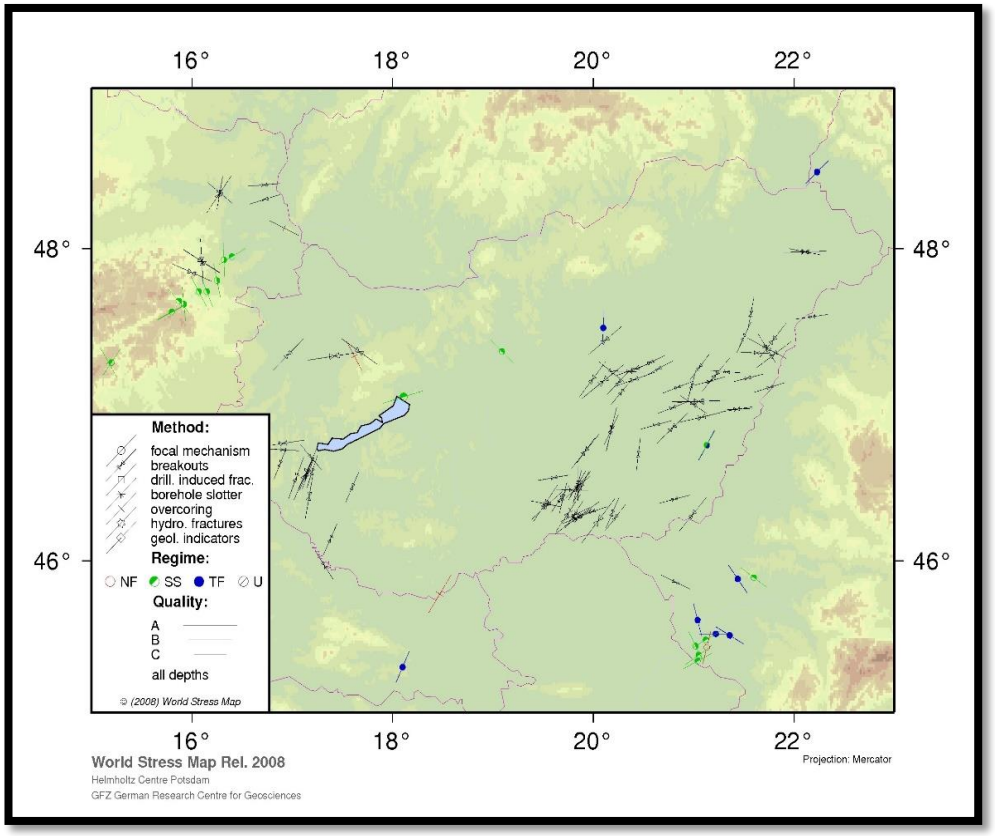

Figure 1

The Sh_{max}-directions and tectonic regimes in Hungary and inside the Pannonian Basin. This map was created by the CASMO software, which is accessible on the website of WSMP [7]. NF – normal fault; SS- strike-slips; TF – thrust fault; U - undefined

In *Figure 1*, the S_{hmax}-directions and tectonic regimes are visualized for Hungary and the surrounding regions. It is noticeable that, although the obtained directions and the locations of measurements are diffuse, the dominant tectonic regimes are thrust faults and strike-slips.

In *Figure 2*, the recent stress field of the Pannonian Basin (gray lines) is presented with the S_{hmax}-directions of the WSMP database (red, yellow lines) along with published data of FMS from Hungarian National Seismological Bulletins (blue lines, [8], [9], [10], [11], [12]).

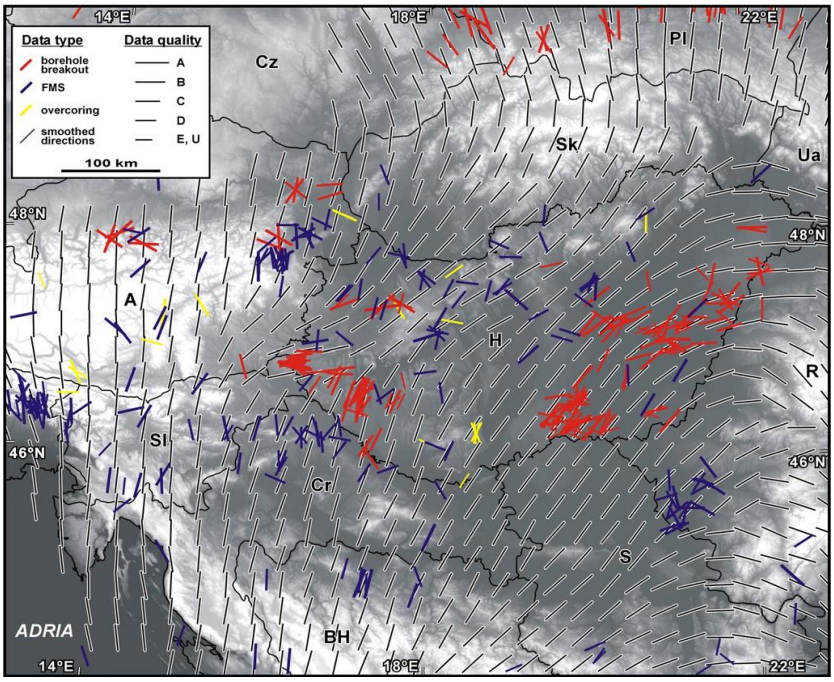

Figure 2

Smoothed stress trajectories (gray lines) shown together with the WSMP database (red and yellow lines) and focal mechanisms (blue lines) [6]

The results of numerical modelling are the grey lines, which mark the smoothed directions of maximum horizontal compression. These lines are more uniform and regular than measurement data, but they differ from them slightly in some places as well. Based on this map, the orientation of the stress field is N-S on the west side, NE-SW in the middle of Hungary, E-W in eastern areas, and forms a fan-shaped pattern [3], [4], [5], [6].

Because the stress field generates earthquakes, analysis using the data on seismic events (focal mechanisms solutions) is the best method for the description of stress relations and tectonics. The data of FMS have three components: strike, dip and rake. The solutions are determined by the first arrivals of P-waves (+ or – polarities) or by full waveform inversions. The given focal mechanisms can be divided into 4 quadrants, two compressional and two dilatational.

The most important aims of my research were the verification and specifying of well-known studies about tectonic stress relations in the Pannonian Basin, particularly in Hungary and the surroundings areas. I used published focal mechanism solutions until 2015. I collected these data from domestic and international databases (e.g. [8]–[18]) to carry out stress inversions using FMSs for the determination of tectonic stresses.

2. DATA

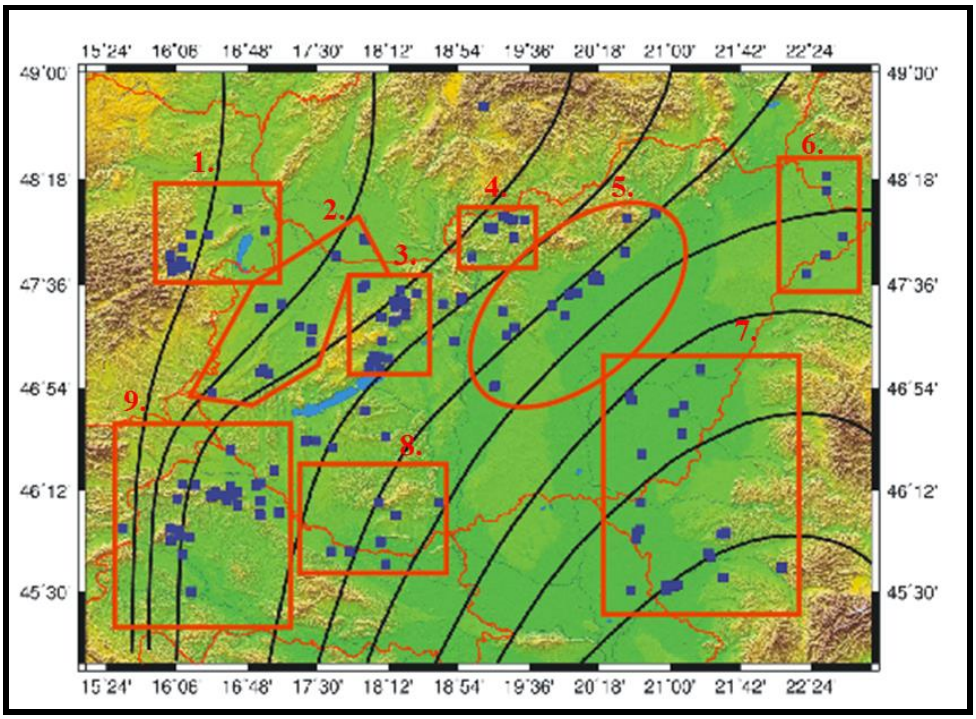

Figure 3

Data points (blue dots) and created subareas (red outlines) for stress inversions. The black curves indicate the stress trajectories of S_{hmax}-directions [6]. 1.: Mur–Mürz–Ziline Zone (MMZ), 2.: Kisalföld, 3.: Dunántúl Highlands, 4.: Region of Nógrád–Gömör, 5.: Jászság, 6.: Nyírség, 7.: Southern part of Alföld and Bánság, 8.: Southern part of Dunántúl (Mecsek), 9.: NW Croatia

Most solutions were collected from Hungarian Earthquake Bulletins [8]–[18], Hungarian National Seismological Bulletins [19], [20] and professional reviews [21], [22], [23], but I also downloaded some FMSs from the website of the European Mediterrean Seismological Centre (http://www.emsc-csem.org/) [24]. These data were estimated using polarities or full waveform inversion.

After the collection of seismic events, solutions had to be selected depending on their reliability. For example, if for one specific earthquake both methods were available, I used the full waveform inverted solution because of its better reliability [25]. This was also the case when the polarities were diffusely distributed or there were few measurements available.

In total 160 focal mechanisms could be utilized for my calculations. Before the estimations, I had to create several subareas because the distribution of events was uneven, so the orientation of principal stress trajectories varied. This partition is presented in *Figure 3.*

3. APPLIED INVERSION METHOD

Three criteria must be met to determine the stress tensor using stress inversion:

1. The stress field of the studied area is homogeneous,
2. The earthquakes evolve on existing faults with variable orientations,
3. The deviation of the fault slip vector from the shear stress vector is minimal (Wallace-Bott hypothesis).

If these criteria are satisfied, the researcher can estimate the stress tensor. The direct outcomes of such an estimation are the orientations of three principal stresses (σ_1, σ_2, σ_3) and the shape ratio R, which characterises the relationship of the principal stresses:

$$R = \frac{\sigma_1 - \sigma_2}{\sigma_1 - \sigma_3} \tag{1}$$

A number of methods have been developed for the inversion of stress fields, for example Gephart [26], Angelier [27], Michael [28], [29] and Vavrycuk [1] contributed greatly to the existing techniques. For the stress tensor estimation presented in this paper, Vavrycuk's method was used in a MATLAB environment. His method is a linear, iterative stress inversion.

All stress inversions aim to find the best fitting stress tensor ($\bar{\bar{\Sigma}}$) (2) to the input focal mechanisms. This stress tensor is a symmetric tensor so it has only six independent components.

$$\bar{\bar{\Sigma}} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix} \tag{2}$$

This inversion is linear, meaning that the data and model vectors are in a linear relationship with each other.

The model vector $\bar{\boldsymbol{m}}$ includes the elements of the stress tensor, and it can be described as follows:

$$\bar{\mathbf{m}} = \begin{pmatrix} \sigma_{11} \\ \sigma_{12} \\ \sigma_{13} \\ \sigma_{22} \\ \sigma_{23} \end{pmatrix} \tag{3}$$

We suppose the stress tensor does not have an isotropic component, so we can express σ_{33} using relation (3):

$$\sigma_{33} = -(\sigma_{11} + \sigma_{22})$$

Otherwise, a constant value should be added to all principal stresses, but this value cannot be determined.

The data vector includes the components of the unit slip vector of each seismic event:

$$\bar{\mathbf{d}} = \begin{pmatrix} s_{11} \\ s_{12} \\ s_{13} \\ \cdots \\ s_{K1} \\ s_{K2} \\ s_{K3} \end{pmatrix} \tag{4}$$

where K marks the number of events, s_{ik} represents the k-th component of the slip vector belonging to the i-th event.

According to the Wallace-Bott hypothesis, the differences between the slip and shear stress vectors have to be minimal or zero. In order to specify the shear stress ($\mathbf{t_s}$), the stress tensor $\left(\bar{\bar{\Sigma}}\right)$ and normal vector ($\mathbf{n}$) are supposed to be known. At first, the stress tensor is multiplied by the normal vector to receive the stress vector ($\mathbf{t}$), after that the normal ($\mathbf{t_n}$) and shear stress vectors [*Equations* of *(5)–(8)*] can be calculated [29], [1]:

$$\mathbf{t}_i = \bar{\bar{\Sigma}}\mathbf{n} \tag{5}$$

$$\mathbf{t_{in}} = (\mathbf{tn})\mathbf{n} \tag{6}$$

$$\mathbf{t_{is}} = \mathbf{t}_i - \mathbf{t_{in}} \tag{7}$$

$$\mathbf{t_s} = \bar{\bar{\mathbf{G}}}\bar{\mathbf{m}},$$

where

$$\bar{\bar{G}}_i = \begin{pmatrix} n_{i1} - n_{i1}^3 + n_{i1}n_{i3}^2 & n_{i2} - 2n_{i1}^2 n_{i2} & n_{i3} - 2n_{i1}^2 n_{i3} & n_{i1}n_{i3}^2 - n_{i1}n_{i2}^2 & -2n_{i1}n_{i2}n_{i3} \\ n_{i2}n_{i3}^2 - n_{i2}n_{i1}^2 & n_{i1} - 2n_{i2}^2 n_{i1} & -2n_{i1}n_{i2}n_{i3} & n_{i2} - n_{i2}^3 + n_{i2}n_{i3}^2 & n_{i3} - 2n_{i2}^2 n_{i3} \\ n_{i3}^3 - n_{i3}n_{i1}^2 - n_{i3} & -2n_{i1}n_{i2}n_{i3} & n_{i1} - 2n_{i3}^2 n_{i1} & n_{i3}^3 - n_{i3}n_{i2}^2 - n_{i3} & n_{i2} - 2n_{i3}^2 n_{i2} \end{pmatrix} \tag{8}$$

Equation (8) can be derived by some algebraical transformations from *Equation (7)*, [27], [1]. The $\bar{\bar{G}}$ matrice is a kernel matrix, its components being the normal vectors.

If we write *Equation (8)* and apply the Wallace-Bott hypothesis, we receive a linear system of equations, which comprises 3*K relations:

$$\bar{\bar{\mathbf{G}}}\bar{\mathbf{m}} = \bar{\mathbf{d}} \tag{9}$$

This problem is over-determined so the Least Squares Method has to be applied to solve it:

$$\bar{\bar{\mathbf{G}}}^{\mathbf{T}}\bar{\bar{\mathbf{G}}}\bar{\mathbf{m}} = \bar{\bar{\mathbf{G}}}^{\mathbf{T}}\bar{\mathbf{d}} \tag{10},$$

where $\bar{\bar{\mathbf{G}}}^{\mathbf{T}}$ is the transpose of $\bar{\bar{\mathbf{G}}}$.

The inversion method of Vavrycuk [1] includes a fault instability constraint to select real fault slip from input data and estimate fault orientations; therefore, it needs more than one iteration step. In the course of iterations, the inversion compares the given auxiliary nodal planes, and with the help of the Mohr-Coulomb criterion (11) selects the most unstable nodal plane as the real fault plane *(Figure 4)*.

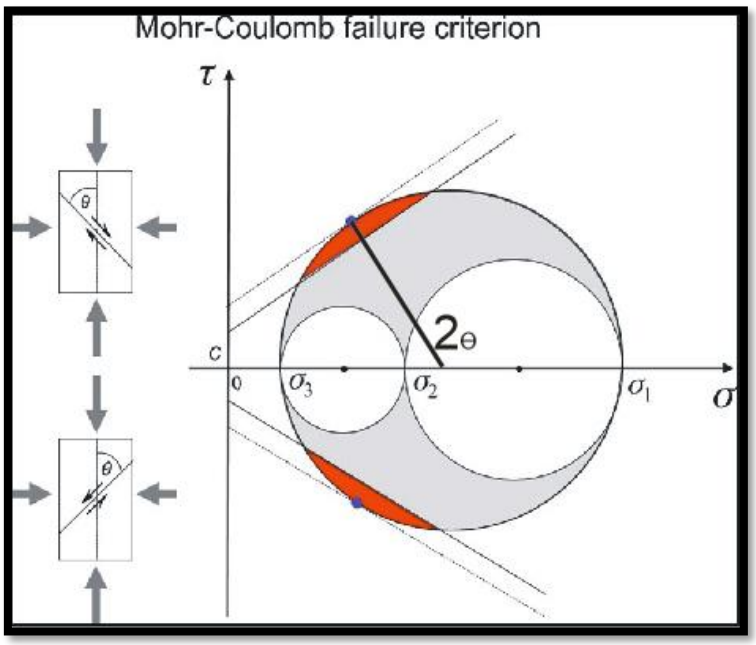

Figure 4
Mohr-Coulomb diagram. Red areas indicate the possible orientations of fault planes and the blue dots are the real fault planes [1]

According to the Mohr-Coulomb criterion, if inequality (11) is true then displacement occurs:

$$\Delta\tau = \tau - \tau_c \geq 0, \tag{11}$$

where τ is the shear stress on the active fault, τ_c is the critical shear stress:

$$\tau_c = C + \mu(\sigma_n - p). \tag{12}$$

In *Equation (12)*, C is the cohesive force, μ is the friction coefficient, p is the pore pressure and σ_n is the normal compressional stress. Because the $\Delta\tau$-values shall be compared, the pore pressure and the cohesive force are negligible and the most important quantity is the friction coefficient, which is between 0.6 and 0.8. The fault instability can be estimated with the help of the following ratio:

$$I = \frac{\tau - \mu * (\sigma - \sigma_1)}{\tau_c - \mu * (\sigma_c - \sigma_1)} \tag{13}$$

τ and σ are the effective shear and normal traction along the analysed fault plane. τ_c and σ_c denote the effective shear and normal traction along the optimally oriented fault plane. The I-values vary between 0 and 1. The nearer the I value approaches 1, the more unstable the fault becomes. In *Figure 5*, fault instability is defined with the help of normal and shear stress [27].

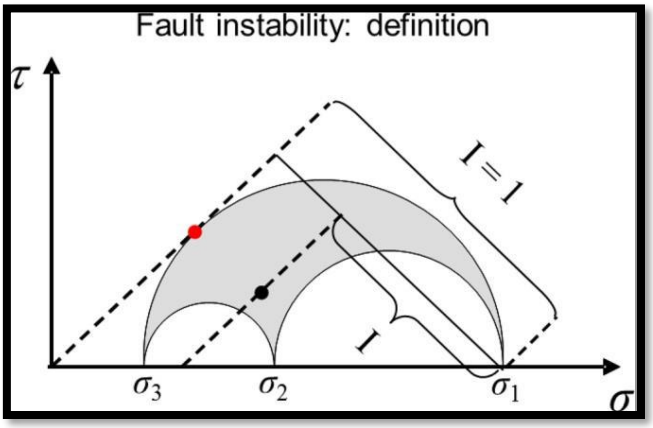

Figure 5
The definition of fault instability. The black dot is a studied fault plane and the red points show the most unstable fault plane (I = 1). The grey part outlines the largest Mohr circle [1].

Consequently, this stress inversion method can select the real fault plane regarding the fault instability constraint and thus it defines an optimal stress tensor. Generally, it needs 3–4 iteration steps. At first, the inversion calculates the principal stress orientations and the shape ratio without the known fault planes, after that it determines the real fault planes with the Mohr-Coulomb criterion. These steps are repeated until the result converges to the best fitting solution. In the cases discussed here, 10 to 60 iteration steps were required, depending on the reliability of the used solutions and the number of data points. Several settings have to be tried to obtain the optimized solution. For this judgment I had to rely on a priori information [4], [5], [6] and my personal experience. Such bias was unavoidable because of the risk of overfitting.
After solving the inverse problems, analysis of the reliability of the method and the results had to be conducted. For that, realizations produced by the STRESS-INVERSE algorithm were used. For these estimations, the number of realizations and the mean deviation needed to be set. In general, 100 realizations were produced for each inversion, the mean deviation being 5 degrees, with the exception of modifications due to the properties of the given focal mechanism solutions. Thereafter, the programme generated further solutions from those available with a Gauss distribution of 5 degrees.

4. RESULTS

Illustrations of the results were produced by STRESSINVERSE in four figures, showing a stereogram of P (compressional) and T (tensional) axes, a stereogram of principal stress axes with their realibility, a histogram of R values, and a diagram of Mohr circles.

In this paper, the outcome of the calculations are presented in detail only for the vicinity of Komárom and Berhida, while other subareas are illustrated on comprehensive maps and via numerical results in tables.

4.1. Dunántúl Highlands

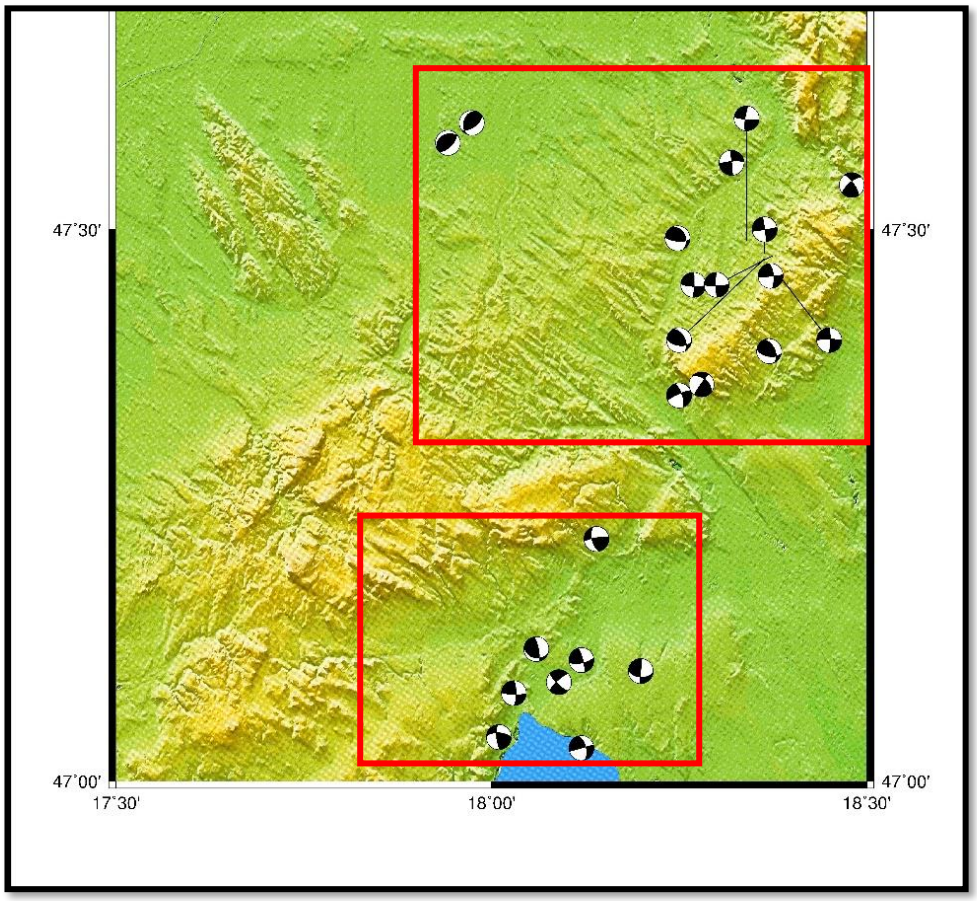

Figure 6

"Beachballs" of seismic events between Komárom and Berhida. The black indicates compressional quadrants, the white indicates dilatational quadrants. The red rectangles outline subareas for the inversions.

In *Figure 6*, the "beachballs" of focal mechanisms are illustrated within the range of Komárom and Berhida. I carried out my estimations using 23 such solutions.

These solutions have to be interpreted with respect to the P axes which are perpendicular to dilatational quadrants. Provided that the direction of the σ_1 and P axes are the same, S_{hmax} vectors can be determined. Therefore, we can assume what kind of tectonic regime can connect to the earthquakes. In this area, the dominant tectonic regimes are the thrust faults and strike-slips. The main orientation of stress field is near NE-SW. Two subareas between Komárom and Berhida were distinguished based on the distribution of epicenters (red rectangles), and solutions for each of these subareas were also derived.

First, results for the whole region are presented in *Figure 7*.

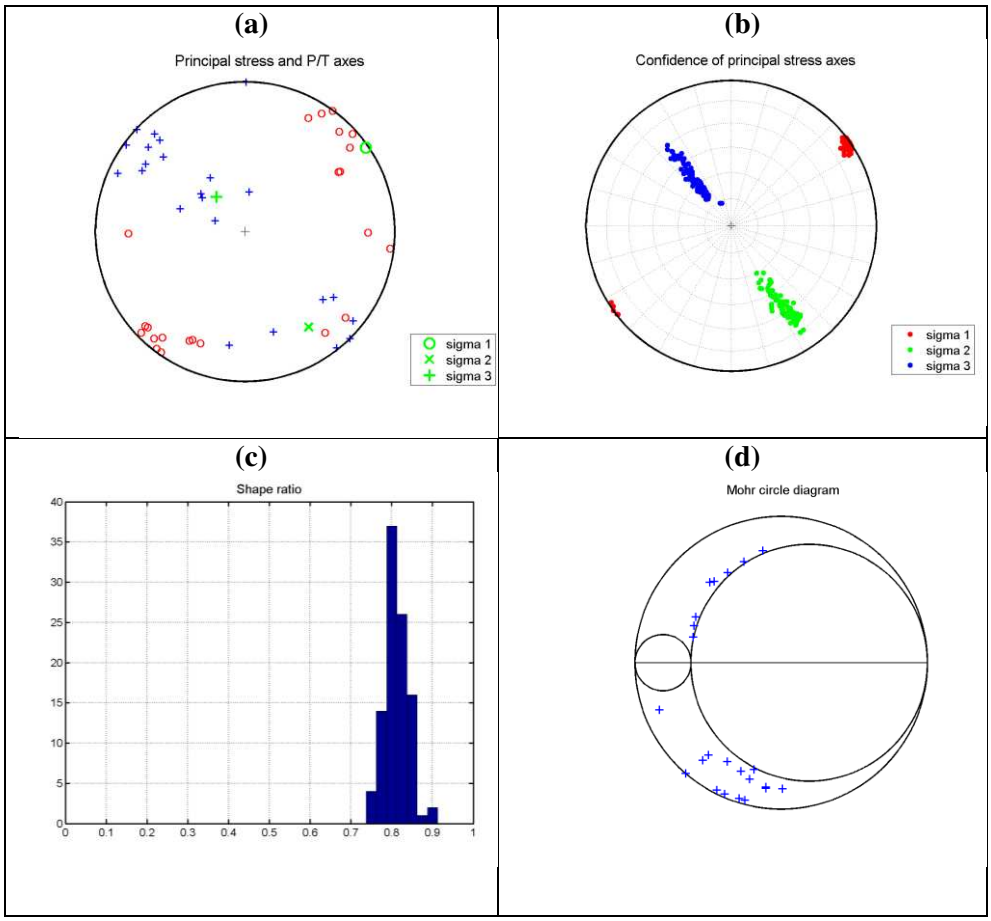

Figure 7

The obtained results using linear, iterative inversion in MATLAB: (a) stereogram of P and T axes, (b) stereogram of principal stress axes with their uncertainties, (c) histogram of shape ratios, (d) Mohr circles

The stereogram of P axes (red circles) and T axes (blue crosses) demonstrate the input data. The main orientation of the P axes is NE-SW but E-W directions occur as well. On the stereogram of principal stress axes, the azimuth of σ_1 (red points) is approximately 55°, so it correlates with the orientations of P axes rather well. The plunges of σ_1, σ_2 (green dots) and σ_3 (blue dots) are roughly 2°; 24° and 66°, namely the axes of σ_3 is the nearest to the vertical position.

The histogram of shape ratios is generated based on the realizations. The maximum of the histogram indicates the best R-value, which is 0.8 in this area.

The Mohr circle diagram represents the reliability of the outcomes and the stress inversion. Blue crosses were assigned to the possible orientations of fault slips (one slip direction belongs to each seismic event). If these crosses were located near the largest Mohr circle, that is between σ_1 and σ_3, their corresponding direction could be considered unstable and identified as that of the real fault plane.

4.2. Northern and southern subareas of Dunántúl Highlands

For the northern part of the region, 15 focal mechanisms could be selected for calculation. In *Figure 8* only the resulting stereogram of principal stresses and the histogram are illustrated.

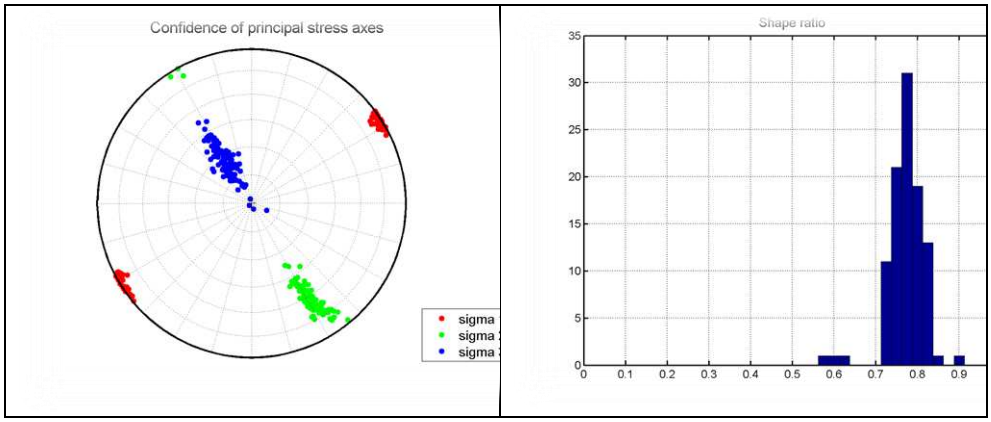

Figure 8

Estimated results on the northern part, the stereogram of principal stress axes and the histogram of R-values

The azimuth of the σ_1-axis is 241.21° and the plunges of σ_1, σ_2 and σ_3 are 0.3°; 24.83° and 65.17°, thus the σ_3-axis is the nearest to vertical position. The best shape ratio is equivalent to 0.78.

In the southern subarea, the situation differs from the one in the north because here, the plunge of σ_2 is found to be the largest. This is clearly visible in *Figure 9*. Otherwise the results are similar to those in the northern part.

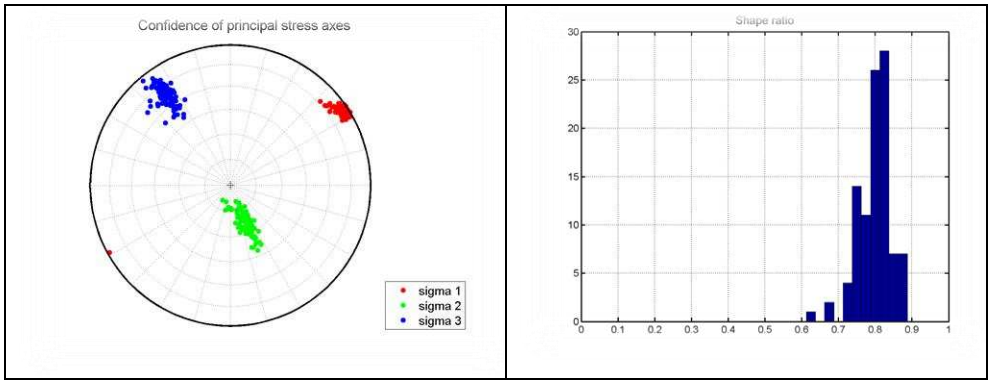

Figure 9

The calculated stereoplot of principal stress axes and histogram of shape ratios

The azimuth of the σ_1-axis is 57.17°, the plunges of σ_1, σ_2 and σ_3 are 4.3°, 66.64°; and 22.9° respectively. The value of the best fitting shape ratio is 0.83.

Both stress fields in the northern and southern parts have similar directions, but the tectonics seems to be different: thrust fault is the typical tectonic regime in the North, whereas strike slip dominates the South.

4.3. Further results

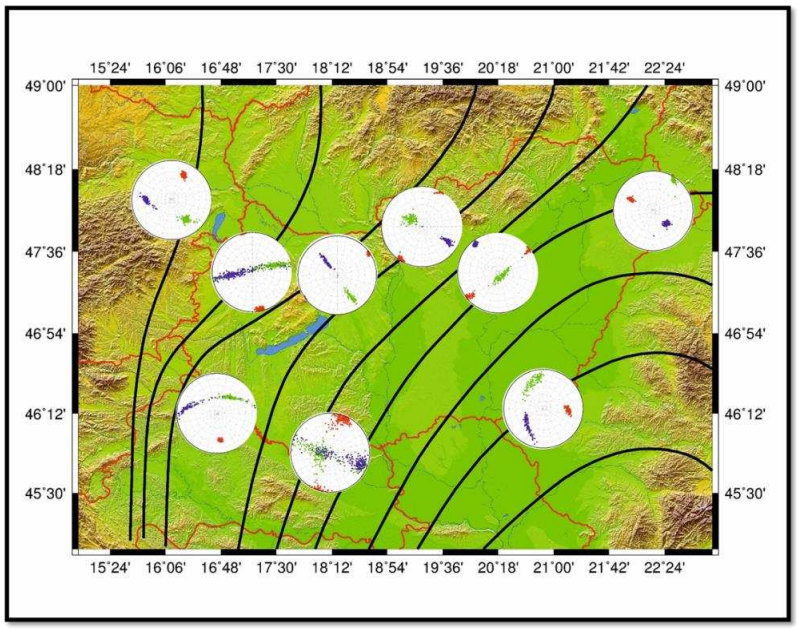

Figure 10

Estimated stereograms of principal stress axes and stress trajectories (black curves, [6])

In the next map *(Figure 10)* stereograms of principal stress axes in all studied regions are presented together with the calculated stress trajectories [6].

In most of the regions, the σ_1 and σ_3 axes are mostly horizontal and σ_2 is vertical, for example in Jászág or in the region of Nógrád–Gömör. There are subareas where the plunge of the σ_3-axis exceeds 50°, for example in Nyírség or in the range of Dunántúl Highlands. Furthermore, in some areas, the axes assigned to σ_2 and σ_3 run into one another (NW Croatia) indicating the uncertainty because of the smaller amount or inaccuracy of utilized data.

Fortunately, the orientations of σ_1 directions are clearly visible everywhere, thus the directions of S_{hmax} could be determined. The direction of maximum horizontal compression is verifiable based on the azimuths of σ_1, which we can describe as the stress field's orientation. The numerical results are presented in *Table 1*. I calculated the values in *Table 1* using STRESSINVERSE. After the estimations, these can be viewed together with the plots.

Table 1

Estimated orientations of σ1, σ2, σ3 and R values

	Plunge			**Azi-muth**	**R**
Area	**σ_1(°)**	**σ_2(°)**	**σ_3(°)**	**σ_1(°)**	
MMZ	27	42.5	36	25	0.68
Kisalföld	6	53	36	170	0.92
Dunántúl Highlands	2	24	66	55	0.80
Northern of Dunántúl Highlands	0.3	24.83	65.17	241.21	0.78
Southern of Dunántúl Highlands	4.3	66.64	22.9	57.17	0.83
NW Croatia	34	48	20	172.5	0.94
S Dunántúl	10.5	72	15	17	0.85
Nógrád-Gömör	2	56	30	30	0.75
Jászság	3	83	6	45	0.37
Nyírség	38	8	51	298	0.72
S Alföld	35	15	51	95	0.70

5. DISCUSSION

After carrying out estimations, the recent stress field could be characterised based on the presented data. I could classify regions based on the plunges of principal stress axes and determine the direction of maximum horizontal compression with the help of the azimuth of σ1 axes. I investigated the tectonic style further using the R' values.

The R' values come from shape ratios depending on which principal axes have the largest plunges. Three relations can be given [4]:

1) If $\boldsymbol{\sigma_1}$ is vertical: R' = R
2) If $\boldsymbol{\sigma_2}$ is vertical: R' = 2-R
3) If $\boldsymbol{\sigma_3}$ is vertical: R' = 2+ R

Values of R' vary from 0 to 3, from clear extension to radial compression. In *Table 2*, I summarize the vertical principal stress axes determined from the estimations, the R' values and the characteristic tectonic styles of the studied regions.

Table 2
Determined vertical principal axes, R and R' values and typical tectonic styles

Area	σ_v	R	R'	Tectonics
MMZ	$\boldsymbol{\sigma_2}$	0.68	**1.32**	**clear strike-slip**
Kisalföld	$\boldsymbol{\sigma_2}$	0.92	**1.08**	**transtension**
Dunántúl Highlands	$\boldsymbol{\sigma_3}$	0.80	**2.80**	**radial compression**
N of Dunántúl Highlands	$\boldsymbol{\sigma_3}$	0.78	**2.78**	**radial compression**
S of Dunántúl Highlands	$\boldsymbol{\sigma_2}$	0.83	**1.17**	**transtension**
NW Croatia	$\boldsymbol{\sigma_2}$	0.94	**1.06**	**transtension**
S Dunántúl	$\boldsymbol{\sigma_2}$	0.85	**1.15**	**transtension**
Nógrád-Gömör	$\boldsymbol{\sigma_2}$	0.75	**1.25**	**clear strike-slip**
Jászság	$\boldsymbol{\sigma_2}$	0.37	**1.63**	**clear strike-slip**
Nyírség	$\boldsymbol{\sigma_3}$	0.72	**2.72**	**clear compression**
S Alföld	$\boldsymbol{\sigma_3}$	0.70	**2.70**	**clear compression**

It is visible in *Tables 1* and *2* that the σ_1-principal axes are always nearly horizontal, which implies the lack of normal faults. In most of the regions, the σ_2 axes point the nearest to the vertical position; these areas are most frequently subject to strike slip

faulting. The R' values indicate the presence of normal faults, too (for example in Kisalföld); furthermore, they appear as transtension (strike-slip with normal faulting components) in *Table 2*. This can be traced back to the inaccuracy and insufficient amount of utilized focal mechanisms.

In three regions – the southern part of Alföld and Bánság, Nyírség and the Dunántúl Highlands – the most frequent tectonic regime is the thrust fault, so the plunges of the σ_3 axes are the largest. This indicates the supposed tectonic style too, based on the determined R' values: clear compression and radial compression are typical in these regions.

As to directions of maximum horizontal compression, the orientation of the stress field is nearly N-S on the west side of the Pannonian Basin (Mur–Mürz–Zilina zone, NW Croatia), NE-SW in the middle of Hungary (Jászság, the region of Nógrád–Gömör) and E-W in the eastern areas (Nyírség, S Alföld). Because this variety is getting larger in the northern and southern areas, the stress field diverges more and more to the east. *Figure 10* illustrates the calculated Sh_{max} directions along with the stress trajectories [6].

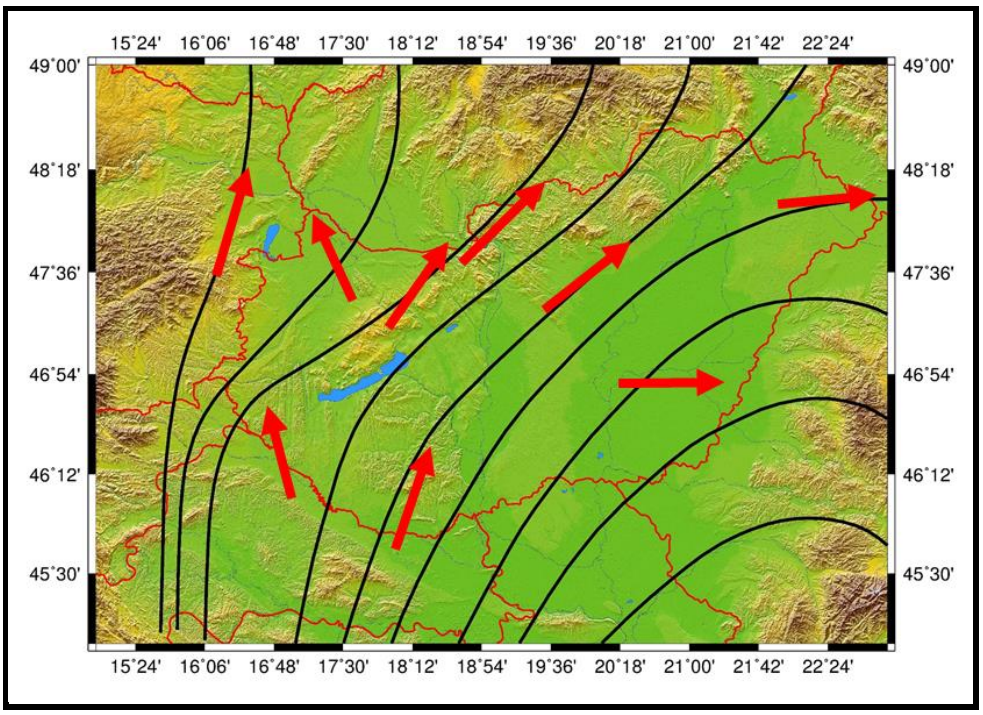

Figure 11
Established directions of maximum horizontal compression (red arrows) and stress trajectories [6]

On the whole, my results are similar to others previously published. As an extension to existing interpretations, stress inversions could also be carried out where data coverage was poor or the existing information was inaccurate (*Figure 12*, blue arrows). These results could also be regarded as fairly reliable. There are two areas, however, where the determined direction of maximum horizontal compression includes an angle differing from the previously published directions (*Figure 12*, red arrows): the region of Kisalföld and the southern part of Alföld and Bánság, respectively. In the region of Kisalföld, the principal direction of stress field is near N-S (here, normal faults can be present too: see *Table 2*) and the dominant orientation in S Alföld is about E-W.

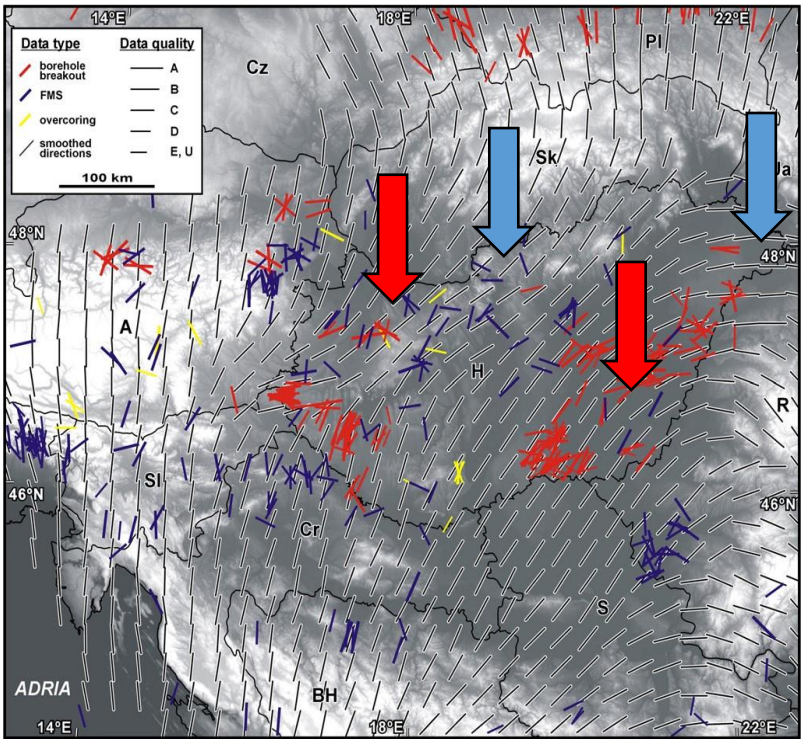

Figure 12
Smoothed stress pattern [6] shown together with my present estimations (blue and red arrows)

6. CONCLUSIONS

In Hungary and the surrounding regions, medium seismicity and diffuse earthquake-distribution is typical, while the seismic network is not very extensive. For these reasons, a relatively small number of focal mechanism solutions are known and these data sometimes have low reliability (because of inaccurate classification). In addition, I had to take into account the variable orientations of stress trajectories; thus, I

was unable to carry out calculations for the entire country as a whole, which implied division into smaller regions. Based on the presented estimations and R' values, it could be shown that the recent stress field rotates from N-S to E-W from west to east, and that the dominant tectonic regimes are thrust faults and strike-slips. These results support the existence of compressional and strike-slip stress fields and, in a broader view, a fan-like stress field. Thus, it is conceivable that the results of my estimations are in good agreement with previously published data on the stress field in Hungary. However, there are some disagreements due to the accuracy of used focal mechanisms or the tectonic background [4], [5], [6], [30], [31].

ACKNOWLEDGEMENTS

I am grateful for the help and support of my supervisor, Zoltán Wéber, senior seismologist at MTA CSFK GGI Kövesligethy Radó Seismological Observatory, in Budapest. Moreover, I thank other researchers at the Kövesligethy Radó Seismological Observatory and at GeoRisk Ltd. for giving me the possibility to utilize all seismological data.

LIST OF SYMBOLS

Symbol	Description	Unit
R	the shape ratio	
R'	derived from shape ratio	
σ_{11}, σ_1	the greatest principal stress axis (PSA)	°
σ_{22}, σ_2	the second greatest PSA	°
σ_{33}, σ_3	the smallest PSA	°
$\bar{\bar{\Sigma}}$	the stress tensor	
$\bar{m}$	the model vector	
$\bar{d}$	the data vector	
$\bar{\bar{G}}$	kernel matrix	
n	normal vector	
t_i	stress vector	
τ	shear stress	Pa
τ_c	critical shear stress	Pa
C	cohesive force	N
μ	friction coefficient	
p	pore pressure	Pa
I	fault instability	
Θ	fracture angle	° or rad
σ_n	normal compressional stress	Pa

REFERENCES

[1] VAVRYCUK, V.: Iterative joint inversion for stress and fault orientations from focal mechanisms. *Geophysical Journal International*, 2014, 199, 69–77.

[2] GERNER P.–BADA, G.–DÖVÉNYI, P.–MÜLLER, B.–ONCESCU, M. C.–CLOETINGH, S.–HORVÁTH, F.: *Recent tectonic stress and crustal deformation in and around the Pannonian Basin: data and models, in The Mediterranean Basins: Tertiary Extension within the Alpine Orogen*. (Eds.: DURAND, B.–JOLIVET, L.–HORVÁTH, F.–SERANNE, M.) Geological Society, London, Special Publications, 1999, 156, 269–294.

[3] GRENERCZY, Gy.–SELLA, G.–STEIN, S.–KENYERES, A.: Tectonic implications of the GPS velocity field in the northern Adriatic region. *Geophysical Research Letters*, 2005, 32 (16).

[4] BADA G.–DÖVÉNYI P.–HORVÁTH F.–SZAFIÁN P.–WINDHOFFER G.: Jelenkori feszültségtér a Pannon-medencében és alpi-dinári-kárpáti környezetében. *Földtani Közlöny*, 2007. 137/3, 327–359.

[5] BADA, G.–GRENERCZY, G.–TÓTH, L.–HORVÁTH, F.–STEIN, S.–CLOETINGH, S.–WINDHOFFER, G.–FODOR, L.–PINTER, N.–FEJES, I.: Motion of Adria and ongoing inversion of the Pannonian basin: Seismicity, GPS velocities, and stress transfer. In STEIN, S.–MAZZOTTI, S. (ed.): *Continental Intraplate Earthquake Science, Hazard, and Policy Issues: Geological Society of America Special Paper 425.* 2007, 243–262, Doi:10.1130/2007-2425(16).

[6] BADA, G.–HORVÁTH, F.–DÖVÉNYI, P.–SZAFIÁN, P.–WINDHOFFER, G.–CLOETINGH, S.: Present-day stress field and tectonic inversion in the Pannonian Basin. *Global and Planetary Change*, 2007, 58/1, 165–180.

[7] World Stress Map Project: www.world-stress-map.org, 2008.

[8] TÓTH L.–MÓNUS P.–ZSÍROS T.–KISZELY M.–CZIFRA T.: *Magyarországi földrengések évkönyve 2002. – Hungarian Earthquake Bulletin 2002*. GeoRisk-MTA GGKI, Budapest, 2003. 104.

[9] TÓTH L.–MÓNUS P.–ZSÍROS T.–KISZELY M.–CZIFRA T.: *Magyarországi földrengések évkönyve 2003.* GeoRisk-MTA GGKI, Budapest, 2004.

[10] TÓTH L.–MÓNUS P.–ZSÍROS T.–KISZELY M.–CZIFRA T.: *Magyarországi földrengések évkönyve 2004*. GeoRisk-MTA GGKI, Budapest, 2005.

[11] TÓTH L.–MÓNUS P.–ZSÍROS T.–KISZELY M.–CZIFRA T.: *Magyarországi földrengések évkönyve 2005*. GeoRisk-MTA GGKI, Budapest, 2006.

[12] TÓTH L.–MÓNUS P.–ZSÍROS T.–BUS Z.–KISZELY M.–CZIFRA T.: *Magyarországi földrengések évkönyve 2006*. GeoRisk-MTA GGKI, Budapest, 2007.

[13] TÓTH L.–MÓNUS P.–ZSÍROS T.–BUS Z.–KISZELY M.–CZIFRA T.: *Magyarországi földrengések évkönyve 2007*. GeoRisk-MTA GGKI, Budapest, 2008.

[14] TÓTH L.–MÓNUS P.–ZSÍROS T.–BUS Z.–KISZELY M.–CZIFRA T.: *Magyarországi földrengések évkönyve 2008*. GeoRisk-MTA GGKI, Budapest, 2009.

[15] TÓTH L.–MÓNUS P.–ZSÍROS T.–BUS Z.–KISZELY M.–CZIFRA T.: *Magyarországi földrengések évkönyve 2009*. GeoRisk-MTA GGKI, Budapest, 2010.

[16] TÓTH L.–MÓNUS P.–ZSÍROS T.–BUS Z.–KISZELY M.–CZIFRA T.: *Magyarországi földrengések évkönyve 2010*. GeoRisk-MTA GGKI, Budapest, 2011.

[17] TÓTH L.–MÓNUS P.–KISZELY, M.: *Magyarországi földrengések évkönyve-Hungarian Earthquake Bulletin-2013*. GeoRisk, Budapest, 2014.

[18] TÓTH L.–MÓNUS P.–KISZELY M.: *Magyarországi földrengések évkönyve – Hungarian Earthquake Bulletin* 2015. GeoRisk, Budapest, Doi:10.7914/SN/HM, 2015.

[19] GRÁCZER, Z.–CZIFRA, T.–KISZELY, M.–MÓNUS, P.–ZSÍROS, T.: *Magyar Nemzeti Szeizmológiai Bulletin 2011*. Kövesligethy Radó Szeizmológiai Obszervatórium, MTA CSFK GGI, Budapest, 2012.

[20] GRÁCZER Z. (szerk.)–CZIFRA T.–GYŐRI E.–KISZELY M.–MÓNUS P.–SÜLE B.–SZANYI Gy.–TÓTH L.–VARGA P.–WESZTERGOM V.–WÉBER Z.–ZSÍROS T.: *Magyar Nemzeti Szeizmológiai Bulletin 2012*. Kövesligethy Radó Szeizmológiai Obszervatórium, MTA CSFK GGI, Budapest, 2013.

[21] WÉBER, Z.–SÜLE, B.: Source Properties of the 29 January 2011 ML 4.5 Oroszlány (Hungary) Mainshock and Its Aftershocks. *Bulletin of the Seismological Society of America,* February 2014, Vol. 104, No. 1, 113–127, Doi: 10.1785/0120130152

[22] WÉBER, Z.: Probabilistic waveform inversion for 22 earthquake moment tensors in Hungary: new constraints on the tectonic stress pattern inside the Pannonian basin. *Geophysical Journal International*, 2016, 204, 236–249.

[23] WÉBER, Z.: Source parameters for the 2013-2015 earthquake sequence in Nógrád county, Hungary. *Journal of Seismology*, 2016, 20(3), 987–999. Doi: 10.1007/s10950-016-9576-6

[24] European Mediterrean Seismological Centre: http://www.emsc-csem.org/

[25] WÉBER Z.: Full waveform inversion of Hungarian earthquake data for hypocenter coordinates and focal mechanisms (in Hungarian). *Magyar Geofizika,* 2006, 47, 173–177.

[26] GEPHART, J. W.: FMSI: A FORTRAN program for inverting fault/slickenside and earthquake focal mechanism data to obtain the regional stress tensor. *Computers & Geosciences,* 1990, Vol. 16, No. 7, 953–989.

[27] ANGELIER, J.: Inversion of earthquake focal mechanisms to obtain the seismotectonic stress IV – a new method free of choice among nodal planes. *Geophysical Journal International*, 2002, 150, 588–609.

[28] MICHAEL, A.J.: Determination of stress from slip data: Faults and folds. *Journal of Geophysical Research,* 1984, Vol. 89, No. B13, 11517–11526.

[29] HARDEBECK, J. L.–MICHAEL, A. J.: Damped regional-scale stress inversions: Methodology and examples for southern California and the Coalinga aftershock sequence. *Journal of Geophysical Research*, 2006, Vol. 111, No. B11310. Doi: 10.1029/2005JB004144.

[30] HORVÁTH F.: A Pannon-medence geodinamikája – Eszmetörténeti tanulmány és geofizikai szintézis. *DSc Thesis, Hungarian Academy of Sciences*, 2007.

[31] HORVÁTH F.–BADA G.–WINDHOFFER G.–CSONTOS L.–DÖVÉNYI P.–FODOR L.–GRENERCZY Gy.–SÍKHEGYI F.–SZAFIÁN P.–SZÉKELY B.–TIMÁR G.–TÓTH L.–TÓTH T.: *A Pannon-medence jelenkori geodinamikájának atlasza: Euro-konform térképsorozat és magyarázó*. (OTKA nyilvántartási szám: T034928), 2005.

Geosciences and Engineering, Vol. 5, No. 8 (2016), pp.85–97.

STATISTICAL EVALUATION OF LATEROLOG RESISTIVITY MEASUREMENTS

ATTILA GALSA[1]–DEZSŐ DRAHOS[1]

[1]*Department of Geophysics and Space Sciences, Eötvös Loránd University*
gali@pangea.elte.hu

Abstract: A statistical inversion method was developed to determine the model parameters describing the borehole environment from laterolog measurements. Five types of laterolog tools were applied to estimate the optimal values of the model parameters, their dispersion, the correlation coefficients and the data error. For three-parameter inversion the model parameters were the resistivity of the far zone *(R_t)*, the resistivity of the invaded, flushed zone *(R_{x0})* and the relative diameter of the invasion *(D)*, while for two-parameter inversion R_{x0} was fixed and considered to be equal to the apparent resistivity from microlaterolog measurement. Based on the inversion of synthetic data it was established that (1) estimated data error approximates well to the additive error, (2) the higher the additive error, the larger the dispersion of the model parameters, (3) dispersions of the model parameters from two-parameters inversion are lower than from three. Two-parameter inversion was applied to calculate R_t and D in a borehole drilled for hydrocarbon (HC) exploration. Lower estimated dispersion of D and R_t was found in the HC-saturated bed, where the contrast of the resistivity of the far zone and the invaded zone is more pronounced.

Keywords: *well log inversion, laterologs, quality-controlled inversion, statistical evaluation, diameter of invasion*

1. INTRODUCTION

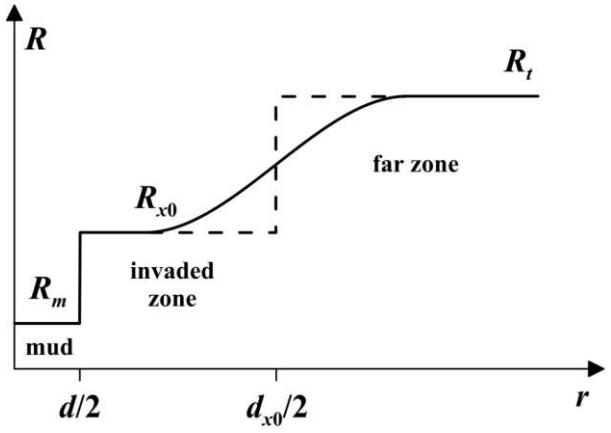

Figure 1. *Radial resistivity variation from the borehole axis (solid line) and the applied three-layered resistivity model (dashed line)*

The main goal of the quantitative evaluation of the electric well logging is to determine the resistivity of the drilled strata sequence. It is especially important in hydrocarbon (HC) exploration, since the HC saturation can be calculated from the rock resistivity.

The role of focused electric probes (laterologs or focused induction probes) has gradually increased over the last decades, as they are less distorted by the finite thickness of beds than normal electric probes reserving the same radial sensitivity (e.g. [1], [2], [3]). The logs measured by focused electric tools are evaluated using so-called butterfly diagrams or their digital analogues. Thus, the resistivity of the undisturbed far zone R_t and the diameter of the invaded zone d_{x0} can be computed. For the evaluation of several laterolog measurements FÖLSZ [4] presented the solution of the direct problem. The evaluation and the effectiveness of the combination of different laterolog measurements were published in BARLAI [5], [6]. *Figure 1* delineates the three-layered resistivity model applied in this study, where R_m, R_{x0} and d denote the resistivity of the mud, the invaded zone and the borehole diameter, respectively, while r is the radial distance from the borehole axis.

Table 1

Two-type evaluation (ML–PLH and OL–PLH) of the laterolog combination

							ML–PLH		OL–PLH	
Layer	***H*** **[m]**	***R*(MLL)** **[Ωm]**	***R*(OL)** **[Ωm]**	***R*(PLR)** **[Ωm]**	***R*(PLH)** **[Ωm]**	***R*(ML)** **[Ωm]**	R_t **[Ωm]**	***D***	R_t **[Ωm]**	***D***
1	3.2	6.8	145	13.5	135	140	530	4.3	170	2.2
2	1.5	6.2	22	4.5	18	12.5	40	4	30	3.5
3	1.7	6.5	27.5	6	23	16.4	52	4.5	38	3.2
4	5.5	6.2	32.5	7	30.2	17.7	60	3.5	42	3.2
5	1.1	3.2	17.5	4.9	20.1	12.8	43	3.6	20	1
6	1.9	6.2	24	6	27.5	15	52	3.7	26	1
7	1.3	7	60	6.8	45	27.5	100	3.5	80	3.1
8	6.5	6.5	50	8	45	20	75	2.9	63	2.5
9	2.2	11.2	100	12	80	115	450	4.6	170	2.8
10	10.2	8.2	55	9.5	45	22.5	80	3.5	71.5	3
11	2.2	6.5	40	7.5	45	17.5	60	3	45	2
12	9.6	5.5	7.5	3.5	9	6	18.8	3	–	–

Based on the evaluation routine it was observed that different diagrams give different results for R_t and the relative diameter of the invaded zone, $D = d_{x0}/d$. This is well illustrated in *Table 1*, adopted from the report entitled 'Evaluation of resistivity using Perkov curve system' [7]. A sketch of the laterolog tools (MLL – microlaterolog, OL – optimal laterolog, PLR – short pseudolaterolog, PLH – long pseudolaterolog, ML – deep laterolog) is drawn in *Figure 2*. The evaluation was accomplished based on ML–PLH and OL–PLH combinations *(Figure 3)* and the results are shown in the last columns of *Table 1* for *12* layers with different thickness of *H*. The difference between the results from ML–PLH and OL–PLH diagrams for the true resistivity *(R_t)* and the relative diameter of the invasion *(D)* may

exceed 15–20% for all layers but in some cases it is much higher (e.g. layers 1 and 9). The OL–PLH combination does not give solution for layer 12 because the point lies outside of the butterfly diagram. Obviously, if a third combination were constructed, the solution would differ from the results of the former two.

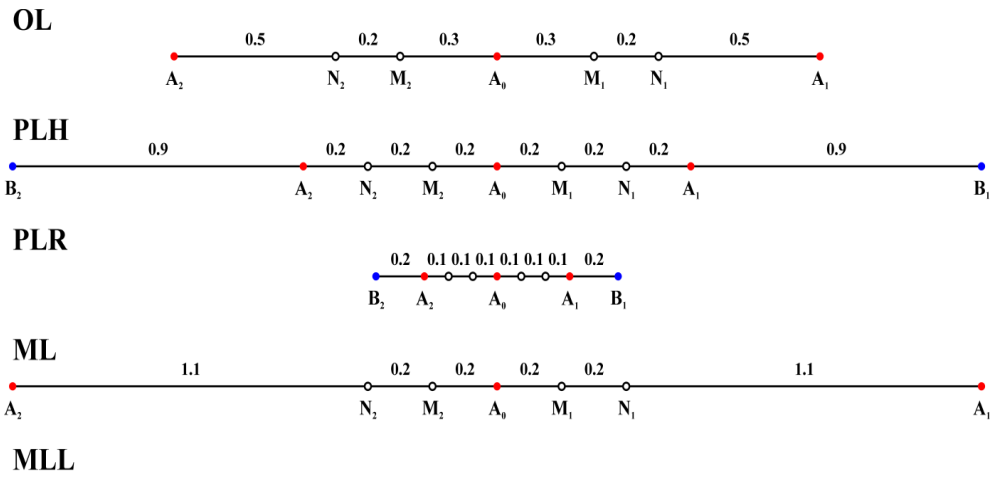

Figure 2. *The sketch of studied laterolog types. OL: optimal laterolog, PLH: long pseudolaterolog, PLR: short pseudolaterolog, ML: deep laterolog, MLL: microlaterolog. A and B are current, M and N are monitor electrodes. Distances are given in meters*

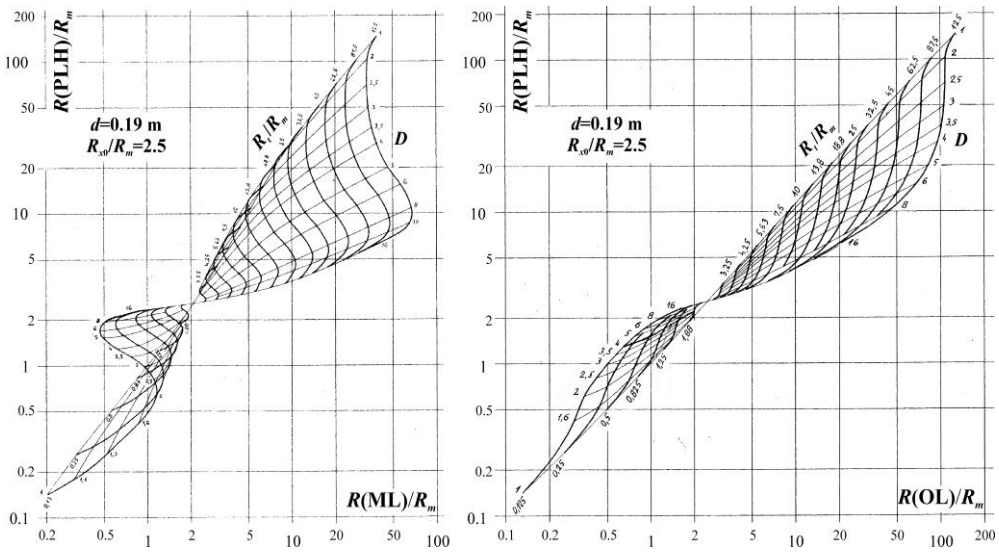

Figure 3

Butterfly diagrams for ML–PLH (left) and OL–PLH (right) laterolog combinations

Discrepancies can be deduced from the difference of the idealized theoretical model and the reality, since the model only approximates the complexity of the reality. Several probes reflect dissimilarly the model–reality deviation. The question is raised: Which evaluation is correct? How to get an optimal solution based on different results? How reliable/accurate is the solution? This paper offers a method to compute an optimal solution that corresponds to the different laterolog measurements.

2. METHOD

The essence of the statistical evaluation method is that a probabilistic variable is prescribed for the difference between the solution of the idealized model and the measured data. The statistical evaluation method gives the estimation of the expectation value and the dispersion providing information about the accuracy of the results obtained. This is called quality-controlled inversion.

The statistical evaluation of measured data is well known and frequently applied e.g. in experimental physics. In geophysics the works of HALFIN [8] and GOLCMAN [9] are fundamental. In Hungary statistical evaluation has been successfully used in geoelectric exploration [10], [11]. It is worth mentioning the work of INMAN [12] and of GLENN and WARD [13], [14] who applied the method in geoelectric surveys. In well logging Schlumberger [15], Dresser Atlas [16], and [17] applied statistical evaluation methods.

2.1. Mathematical model

If the geological and physical model described exactly the reality, and the measurement were errorless, the measured data, d_i and theoretical solution of the direct problem, $g_i(\mathbf{m})$ would be equal for all measurements i in model parameter field $\mathbf{m}$. However, the measurement always entails errors and the model only approximates the real environment; thus, the values d_i and $g_i(\mathbf{m})$ will be never identical. The measured and calculated values differ,

$$\begin{aligned} d_1 &= g_1(\mathbf{m}) + \varepsilon_1, \\ d_2 &= g_2(\mathbf{m}) + \varepsilon_2, \\ &\ldots \\ d_n &= g_n(\mathbf{m}) + \varepsilon_n, \end{aligned} \tag{1}$$

where d_i and $g_i(\mathbf{m})$ denote the measured and calculated laterolog resistivities for the tool of $i=1\ldots n$ and ϵ_i is the error, which is assumed to be random.

It is supposed that ϵ_i errors have n-dimensional normal distribution, the probability density function of which is

$$\mathrm{prd}(\boldsymbol{\varepsilon}) = \frac{1}{\sqrt{(2\pi)^n \det \mathbf{C}}} \exp\left[-\frac{1}{2} \boldsymbol{\varepsilon}^T \mathbf{C}^{-1} \boldsymbol{\varepsilon} \right], \tag{2}$$

where **C** is the covariance matrix of the data [18] and *T* superscript denotes the transpose of the error vector. Logarithmic scale was used for the inversion of the electric measurements, logarithmically scaled resistivities are assumed equally weighted and uncorrelated. To solve the inversion problem the maximum likelihood estimation method was applied,

$$L(\mathbf{m}) = \frac{1}{\sqrt{(2\pi\sigma_0^2)^n}} \exp\left[-\frac{1}{2} \sum_{i=1}^{n} \varepsilon_i^2 \frac{1}{\sigma_0^2} \right] = \max . \tag{3}$$

The task is to determine the parameter vector **m**, which reproduces the measured data with the highest probability.

Due to the normal distribution of the errors, the maximum likelihood estimation method is equivalent to the least squares method. This means that for determining the estimation of **m**, the condition

$$\lambda(\mathbf{m}) = \sum_{i=1}^{n} \left[\log R_i^{(M)} - \log R_i^{(C)} \right]^2 = \min \tag{4}$$

must be satisfied, where $R_i^{(M)}$ and $R_i^{(C)}$ denote the measured and calculated resistivity for the measurement of *i*. Whereas λ is not a linear function of the model parameters **m**, the minimization can be achieved by approximative algorithms, e.g. conjugate gradient, simulated annealing, or simplex method [19]. By minimizing λ in *Equation (4)* the estimation of **m** is determined, and from the value of λ the variance of the measurements (σ_0^2) and the covariance matrix of the model parameters can be calculated, including the estimation of the standard deviation, $\sigma(m_i)$ and the correlation coefficients $\rho(m_i,m_j)$. Since the inversion was done on a logarithmic scale, the estimated model parameters are

$$\mathbf{m}^T = (\log R_t, \log R_{x0}, D) . \tag{5}$$

The resistivity of mud (R_m) and the borehole diameter (d) are supposed to be known from other logs so they are fixed during the inversion. The resistivity of the invaded zone can be obtained from microlaterolog measurements; thus, R_{x0} can be fixed as well. During the inversion with two free parameters the model vector of

$$\mathbf{m}^T = (\log R_t, D) \tag{6}$$

is estimated. There are numerous methods to solve the laterolog direct problem (e.g. [20], [21], [22]). However, during the minimization the direct problem solution must be done many times, so the linear filter theory developed by GHOSH [23] was applied to accelerate computation.

3. APPLICATION OF THE DEVELOPED INVERSION METHOD

3.1. Inversion of simulated measurements

Two models of shallow (D = 3) and deep (D = 6) invasion with additive error of 5, 10 and 20% were investigated. All four macrolaterologs (OL, PLH, PLR and ML) were taken into account to estimate the model parameters and their dispersions and correlation coefficients by inversion. Three-parameter inversion was used to evaluate R_t, R_{x0} and D, then two-parameter inversion was applied to estimate R_t and D, while R_{x0} was supposed known. Results are listed in *Table 2*. Exact model parameters are indicated in the row entitled Model for shallow (left, D = 3) and deep (right, D = 6) invasion. The next row describes the values of the additive errors. Then the estimated model parameters follow with their estimated deviations converted to linear scale. (e.g. R_t = 498 Ωm and 1.10 represent the error interval of (453 Ωm, 548 Ωm) for R_t). It is followed by correlation coefficients (ρ) between the estimated model parameters and the estimation of the data error (σ_0). The inversion procedure is reapplied when R_{x0} is fixed (two-parameter inversion).

Table 2

Estimation of the model parameters (R_t, D, R_{x0}) with their dispersion, correlation coefficients (ρ) and the estimated data error (σ_0) for three and two parameters

INVERSION FOR THREE PARAMETERS												
Model	**D = 3, R_{x0} = 20 Ωm, R_t = 500 Ωm**						**D = 6, R_{x0} = 20 Ωm, R_t = 500 Ωm**					
Error	**5%**		**10%**		**20%**		**5%**		**10%**		**20%**	
R_t [Ωm]	498	1.10	504	1.10	449	1.15	501	1.05	492	1.15	471	1.25
D	3.0	1.05	3.1	1.16	2,9	1.25	6.1	1.05	6.0	1.21	5.5	1.28
R_{x0} [Ωm]	20	1.26	18	1.71	19	2.33	20	1.07	21	1.29	14	1.56
ρ(R_t, D)	0.57		0.56		0.58		0.85		0.84		0.76	
ρ (R_t, R_{x0})	0.34		0.35		0.35		0.62		0.61		0.53	
ρ (R_{x0}, D)	0.86		0.87		0.86		0.82		0.82		0.80	
σ_0 [%]	5		14		21		4		12		21	
INVERSION FOR TWO PARAMETERS												
R_t [Ωm]	501	1.04	505	1.07	473	1.12	494	1.04	511	1.12	445	1.19
D	3.1	1.03	3.1	1.07	2.9	1.11	5.9	1.04	6.3	1.10	5.9	1.13
ρ (R_t, D)	0.55		0.55		0.60		0.73		0.77		0.72	
σ_0 [%]	5		10		17		4		10		17	

Conclusions from *Table 2* are as follows:

- Calculated σ_0 approximates well to the additive error.
- Increasing error enhances the dispersion of the model parameters, but does not affect substantially the correlation coefficients.
- Dispersions of the model parameters are lower for two-parameter inversion than for three.
- For three-parameter inversion $\rho(R_{x0}, D)$ is high, indicating that the increased resistivity in the invaded zone is compromised with the farther 'far zone'.
- For deep invasion $(D = 6)$ $\rho(R_t, D)$ is higher (than for shallow invasion), because higher far zone resistivity is balanced by the farther 'far zone'.
- For deep invasion the dispersion of R_{x0} is lower, since the current flows longer in the invaded zone.

These observations agree well with the qualitative expectations.

3.2. Evaluation of layer data

Two-parameter inversion (R_t, D) was performed based on the data obtained from butterfly diagrams *(Figure 3)*. The results from the quality-controlled inversion and the butterfly diagrams are compared in *Table 3*. Results for R_t and D are shown in the first and second part of the table, respectively. Columns OL–PLH and ML–PLH repeat the results from the butterfly diagrams listed in *Table 1*. The middle number in the column entitled "Estimation" gives the estimation of the parameter, the numbers before and after it represent the parameter interval calculated by dispersion. The table reveals that the results from OL–PLH evaluation are close to the lower end of the range of deviation, while the results from ML–PLH evaluation are close to the higher end. In general, the evaluation from OL–PLH and ML–PLH diagrams harmonize with the interval of the deviation of the estimated parameters.

Table 3

Comparison of the results obtained from two-parameter inversion and butterfly diagrams (OL–PLH and ML–PLH)

Layer	OL–PLH R_t [Ωm]	Estimation R_t [Ωm]			ML–PLH R_t [Ωm]	OL–PLH D	Estimation D			ML–PLH D
1	170	188	304	492	530	2.2	1.7	2.5	3.7	4.3
2	30	24	33	46	40	3.5	2.3	3.5	5.3	4.0
3	38	35	44	55	52	3.2	2.4	3.2	4.3	4.5
4	42	40	51	65	60	3.2	2.1	2.7	3.5	3.5
5	20	25	32	41	43	1.0	1.9	2.5	3.3	3.6
6	26	30	40	53	52	1.0	1.9	2.8	4.1	3.7
7	80	59	86	125	100	3.1	2.0	2.9	4.2	3.5
8	63	52	68	89	75	2.5	1.7	2.3	3.1	2.9
9	170	158	268	456	450	2.8	2.4	3.8	6.0	4.6
10	71.5	59	76	98	80	3.0	1.9	2.5	3.3	3.5
11	45	41	56	78	60	2.0	1.5	2.2	3.2	3.0
12	–	5	14	39	18.8	–	1.1	8.1	59	3.0

3.3. Evaluation of point-wise data of logs

Two-parameter resistivity evaluation was made based on laterolog logs bored for HC exploration, that is R_t and D were estimated. Independently of the variability of the bedding logs were digitized at depth points located at 0.5 m intervals, an infinitely thick layer model was fitted for the inversion of each depth point. Paying attention to our experience from the evaluation of synthetic data (Section 3.1) and real layer data (Section 3.2) it was observed that the expected parameter accuracy is at least 10–20%. Thus the inversion method was simplified. The substance of the method is that the interpreter – based on his/her geological knowledge – determines intervals of [R_t(min), R_t(max)], [D(min, D(max)] and [R_{x0}(min), R_{x0}(max)]. Within these intervals a logarithmically equidistant mesh is generated for the resistivities and the theoretical values of log$R^{(C)}(R_{t,k}, D_l, R_{x0,m})$ are computed at each grid point (relative distance between grids is 7%). On the supposition that the values of R(MLL) correspond to R_{x0}, the plane $R_{x0,m0}$ nearest to the value of R(MLL) is selected. On this plane

$$\lambda_{k,l,m0} = \sum_{i=1}^{n} \left[\log R_{i,k,l,m0}^{(M)} - \log R_{i,k,l,m0}^{(C)} \right]^2 \tag{7}$$

is calculated on the grid points. From these values the minimum is chosen; the model parameters belonging to this minimum provide the solution. The dispersion of the parameter is approximated by the difference quotient calculated from the adjacent grid points. The method of the minimum finding is sketched in *Figure 4*.

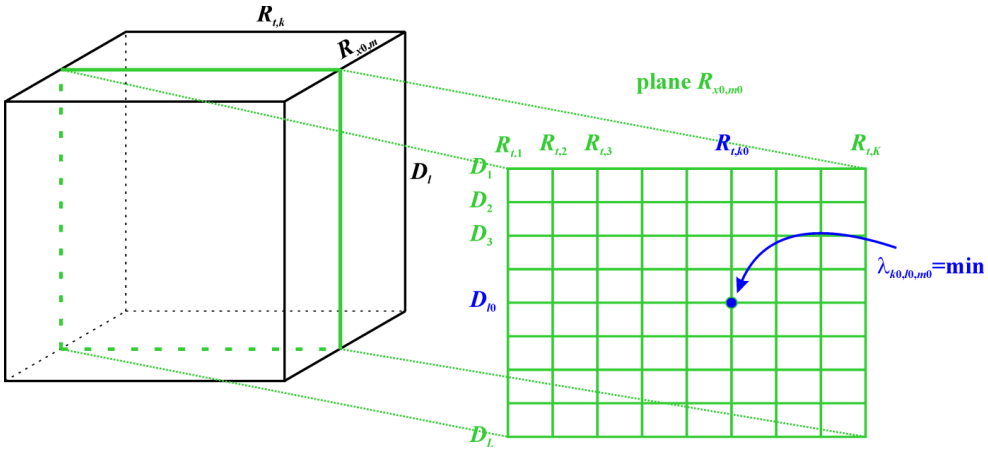

Figure 4. *Sketch of minimum estimation. The minimum value of λ on the plane of $R_{x0,m0}$ gives the solution of the inversion for R_t and D*

The result of the evaluation of the laterolog logs is shown in *Figure 5*. The log can be divided into three sections:

1 2260–2282 m shale (marl);
2 2282–2304 m sandstone saturated with HC;
3 2304–2318 m sandstone saturated with water.

From the calculated parameters (D and R_t) and their deviations [$\sigma(D)$ and $\sigma(R_t)$], error intervals were determined and visualized in the logs with the form of D-$\sigma(D)$, D, $D + \sigma(D)$ as well as $R_t - \sigma(R_t)$, R_t, $R_t + \sigma(R_t)$.

In the shale zone, large error is observed for D owing to the fact that the diameter of the invasion is meaningless in a homogeneous environment. In the HC-saturated section the error interval is thin, demonstrating that the estimation for D is well defined. The reason for that is the high resistivity ratio of R_t and R_{x0}. In the

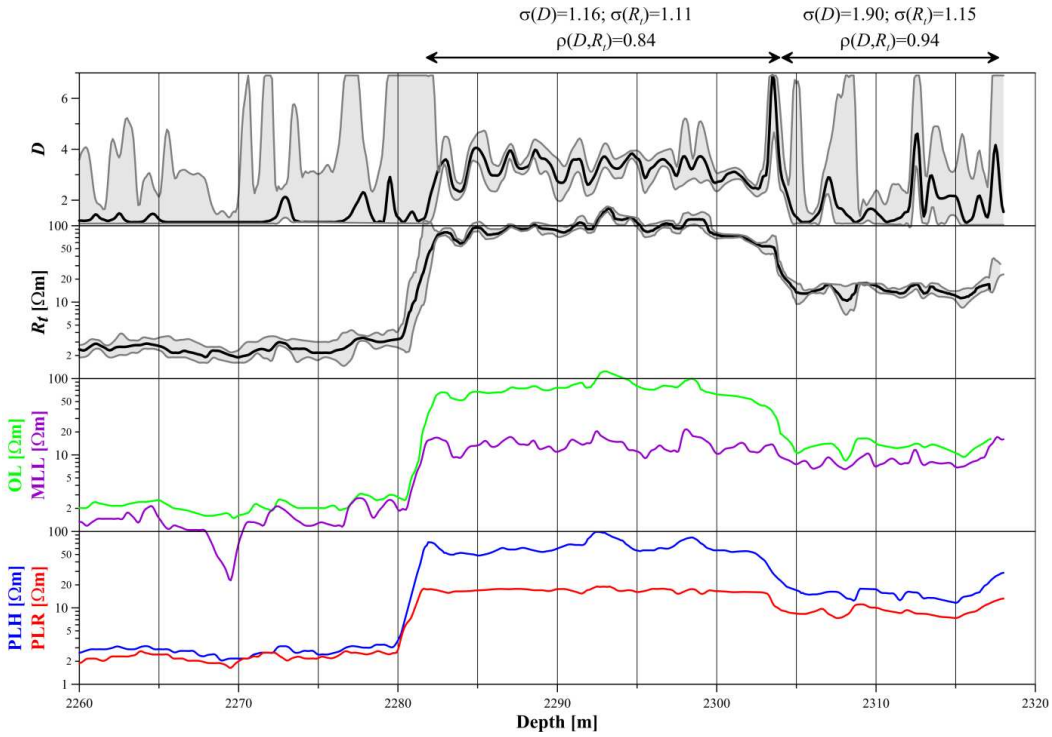

Figure 5. *Measured laterolog logs (below: PLH, PLR, OL, MLL) and result logs (D, R_t) from two-parameter statistical inversion method. Result logs are characterized by their error intervals (shaded). Average standard deviation of the result logs and the correlation coefficient of the model parameters for the HC-saturated (2282–2304 m) and water-saturated (2304–2318 m) zones are indicated.*

water-saturated layer this contrast is much lower; thus, the deviation of D is higher again. In the HC-saturated bed the average deviations are $\sigma(D) = 1.16$ and $\sigma(R_t) = 1.11$, the correlation coefficient is $\rho(D, R_t) = 0.84$. In the water-saturated zone they are $\sigma(D) = 1.90$ and $\sigma(R_t) = 1.15$ and $\rho(D, R_t) = 0.94$. The estimations for $\sigma(R_t)$ are lower, and especially small in the bed with HC content.

4. DISCUSSION AND CONCLUSIONS

The main goal of the study was to investigate the statistical inversion evaluation of laterolog measurements. The advantage of the statistical method is that it provides quality-controlled results, and it points out the uncertainty of the solution due to the always existing errors. In order to apply the evaluation method the direct problem must be solved many times, thus the linear filter theory developed by GHOSH [23] was applied.

First, synthetic erroneous data were inverted and the error propagation was investigated. It was established that at least 15–20% deviation is probable for real data. Based on the inversion results the estimation of R_{x0} is very uncertain for three-parameter evaluation. Therefore, the use of the two-parameter inversion is suggested in practice, where the resistivity of the invaded zone is fixed, R_{x0}=R(MLL).

Inversion method of laterolog logs was developed for industrial application, where the minimum finding was simplified. Here the resistivity values were calculated at discrete grid points in a three-dimensional model domain (R_t, D, R_{x0}). The minimum between the calculated and measured resistivities *[Equation (7)]* determined the parameter estimations and their deviations. Inverted model parameters were more accurate when the resistivity contrast between the invaded and far zone was large, that is, in the HC-saturated zone.

We note that the deviation of the resistivity of the far zone $\sigma(R_t)$ propagates to the water saturation via e.g. the simplified Archie formula,

$$S_w = \sqrt{\frac{R_0}{R_t}}, \tag{8}$$

where S_w and R_0 denote the water saturation of the far zone and the resistivity of the water saturated rock, respectively. Following the rule of the error propagation the relative deviation of the water saturation is

$$\left(\frac{\sigma_{Sw}}{S_w}\right)^2 = \left(\frac{1}{2}\frac{\sigma_{R0}}{R_0}\right)^2 + \left(\frac{1}{2}\frac{\sigma_{Rt}}{R_t}\right)^2, \tag{9}$$

which means that only half of the relative deviation of R_t contributes to the relative deviation of S_w.

ACKNOWLEDGEMENTS

Authors thank Dr. Gábor Kovács for technical assistance.

LIST OF SYMBOLS

Symbol	Description	Unit
C	covariance matrix of the data	Ωm^{-2}
d	borehole diameter	m
D	relative diameter of the invaded zone, d_{x0}/d	–
d_i	measured resistivity of the i-th laterolog log	Ωm
d_{x0}	diameter of the invaded zone	m
g_i	calculated resistivity of the i-th laterolog log	Ωm
H	layer thickness	m
L	maximum likelihood function	–
m	model parameter vector	–
n	number of laterolog tools	–
R	apparent resistivity	Ωm
$R^{(C)}$	calculated apparent resistivity	Ωm
$R^{(M)}$	measured apparent resistivity	Ωm
R_0	resistivity of the rock saturated by water	Ωm
R_m	resistivity of mud	Ωm
R_t	resistivity of the far zone	Ωm
R_{x0}	resistivity of the invaded zone	Ωm
S_w	water saturation	–
ϵ_i	error of the i-th laterolog log	Ωm
λ	objective function to be minimized	–
$\rho(m_i,m_j)$	correlation coefficient between parameter m_i and m_j	–
$\sigma(m_i)$	dispersion of the model parameter m_i	–
σ_0	estimated data error	%

REFERENCES

[1] SCHLUMBERGER: *Log interpretation principles/applications*. Schlumberger Wireline & Testing, SMP-7017, Sugar Land, Texas, 1991.

[2] DENG, S.–LI, Z.: Response of dual laterolog and fast correction for layer thickness and shoulder bed in horizontal wells. *Petroleum Exploration and Development*, 2009, 36(6), 725–729.

[3] SABOORIAN-JOOYBARI, H.–DEJAMB, M.–CHEN, Z.–POURAFSHARY, P.: Comprehensive evaluation of fracture parameters by dual laterolog data. *Journal of Applied Geophysics*, 2016, 131, 214–221.

[4] FÖLSZ A.: *A laterolog kiértékelési görbeseregek számításának újabb eredményei*. OKGT report, 1965.

[5] BARLAI Z.: Az ellenállásszelvényezés kombinatív alkalmazása a szénhidrogén-kutatásban. *Magyar Geofizika*, 1967. VIII. 100–110.

[6] BARLAI Z.: *Fúrólyukban végzett geofizikai vizsgálatok speciális problémái*. I. kötet, NIM Műszaki Dokumentációs és Ford. Iroda, Budapest, 1971.

[7] BESSENYEI I.–MARKÓ L.: *A fajlagos ellenállás meghatározásának vizsgálata a Perkov-görbeseregek alapján.* OKGT report, Budapest, 1974.

[8] HALFIN, L.A.,: Informacionnaja tyeorija intyerpretacii geofizicseszkih isszledovannyij. Dokl. *An. SzSzSzR*, 122(6), 1958.

[9] GOLCMAN, F.M.: *Sztatiszticseszkie modelü interpretacii*. Izd. Nauka, Moszkva, 1971.

[10] CSEREPES L.: Statisztikus paraméter-meghatározási algoritmusok alkalmazása geoelektromos mérésekre. *Magyar Geofizika*, 1978. XIX. 15–22.

[11] SALÁT P.: *Geoelektromos mérések újraértékelése*. Research report, KFH 3958/KG–951/1975, Budapest, 1979.

[12] INMAN, J.R.: Resistivity inversion with ridge-regression. *Geophysics*, 1975, 40, 798–817.

[13] GLENN, W. E.–WARD, S. H.: Statistical evaluation of electrical sounding methods. Part I: Experiment design. *Geophysics*, 1976, 41(6), 1207–1221.

[14] GLENN, W.E.–WARD, S.H.: Statistical evaluation of electrical sounding methods. Part II: Applied electromagnetic depth sounding. *Geophysics*, 1976, 41(6), 1222–1235.

[15] MAYER, C.–SIBBIT, A.: *GLOBAL, a new approach to computer-processed log interpretation.* SPE-AIME , Ann. Fall Mtg., 1980, Dallas, Texas, SPE 9341.

[16] ALBERTY, M.–HASHMY, K.H.: Application of ULTRA to log Analysis. *SPWLA Transactions*, 1984, 1–17.

[17] DOBRÓKA, M.–SZABÓ, N. P.–TÓTH, J.–VASS, P.: Interval inversion approach for an improved interpretation of well logs. *Geophysics*, 2016, 81(2), D163–D175.

[18] MENKE, W.: *Geophysical Data Analysis: Discrete Inverse Theory*. Third edition, Elsevier, 2012.

[19] PRESS, W. H. –TEUKOLSKY, S. A.–VETTERLING, W. T.–FLANNERY, B. P.: *Numerical recipes*. Third edition, Cambridge University Press, 2007.

[20] COZZOLINO, K.–DA CONCEIÇÃO DA SILVA, J.: Synthetic focusing and simulation of dual laterolog tool in axisymmetric subsurface models. *Journal of Applied Geophysics*, 2007, 61, 102–110.

[21] DENG, S.–MO, X.–LU, C.–ZHANG, Y.–LIU, L.: Numerical simulation of the dual laterolog response to fractures and caves in fractured-cavernous formation. *Petroleum Exploration and Development*, 2012, 39(6), 751–757.

[22] JARZYNA, J. A.–CICHY, A.–DRAHOS, D.–GALSA, A.–BAŁA, M.J.–OSSOWSKI, A.: New methods for modeling laterolog resistivity corrections. *Acta Geophysica*, 2016, 64(2), 417–442.

[23] GHOSH, D. P.: The application of linear filter theory to the direct interpretation of geoelectrical resistivity sounding measurements. *Geophysical Prospecting*, 1971, 19(2), 192–217.

Geosciences and Engineering, Vol. 5, No. 8 (2016), pp.98–110.

AN INTRODUCTION TO ROBUST TOMOGRAPHY METHODS

UTKU KALE–MIHÁLY DOBRÓKA*

Department of Geophysics, University of Miskolc, Hungary
**dobroka@uni-miskolc.hu*

Abstract: Seismic tomography is an important imaging tool in mining, and oil & gas exploration. The aim of this article is to test outlier sensitivity for a medium sized tomography example applied to synthetic data. Throughout the test, tomographic algorithms were applied to demonstrate the powers of algorithms. One of the major problems in seismic tomography is to find an effective tomography algorithm that gives acceptable result just not only for Gaussian data but also in the case of non-Gaussian distribution of noise, for example when a small portion of data contains large error. The purpose of test was to reduce the sensitivity to outliers by using most common algorithms in use today: Conjugated Gradients (CG) and Simultaneous Iterative Reconstruction Technique (SIRT). We also modified the CG and SIRT by utilizing Cauchy-Steiner Weights to define a robust tomography algorithm.

Keywords*: Robust tomography, Conjugated Gradients (CG), Simultaneous Iterative Reconstruction Technique (SIRT), Model Distance, Data Distance, Cauchy-Steiner Weights, ART Correction, Gaussian and non-Gaussian data.*

1. INTRODUCTION

In this paper, robust tomography images are produced by utilizing some generalized tomography algorithms. Seismic tomography method plays an important role to solve some geophysical problems in mining as well as in oil and gas exploration. In practice, measured data always contain noise that can cause considerable distortion on the tomographic reconstruction. Non-Gaussian noise distribution (including outliers) is extremely dangerous in distorting the tomographic reconstruction. In order to reduce the outlier sensitivity, we have applied a generalized version of some tomography algorithms. Traditionally, least squares problems in tomography have been solved by row action methods such as Algebraic Reconstruction Technique (ART) or Simultaneous Iterative Reconstruction Technique (SIRT). It was also proved [1] that Conjugate Gradient (CG) method can also be used in even large-scale tomographic least squares inversion. It is well known in inverse problem theory that the traditional least square methods give optimal results only if the data noises follow a Gaussian distribution. However, in practice Gaussian distribution seldom occurs. This means that the traditional CG SIRT approaches will not give robust results for non-Gaussian data. Cauchy inversion is frequently used in geophysical inversion as a robust optimization technique [2]. In this paper we mod-

ified the traditional algorithms by utilizing the Cauchy-Steiner Weights in order to make the CG and SIRT methods more robust.

2. EXPERIMENTAL

Using synthetic travel-time data the generalized (robust) tomography algorithms are tested in a medium-sized tomography example. In experiments, the tomography model was defined by a rectangular test area of the size 100 × 100 cells *(Figure 1)*. In the model three anomalies with velocities 5 km/s (marked red) are located in a homogeneous background of 4 km/s velocity (blue). Sources and receivers were positioned along the x and y axes in an arrangement fulfilling the requirement of full tomographic ray coverage, so the theoretical (noiseless) travel time data were computed along 60,000 ray traces. In the model, there are 60,000 known equations and 10,000 unknown parameters, so this is an overdetermined inverse problem.

Two datasets were generated for the tests. In Dataset I the ideal (noiseless) dataset was contaminated with Gaussian noise (the size of the noise is 1% of the theoretical travel-time). Thus, Dataset I follows the Gaussian distribution. In order to simulate non-Gaussian noise, we produce another dataset that is highly distorted. The second dataset was created from the first one by adding 20% extra noise to a randomly selected 20% portion of the data. That means 80% of Dataset II contains 1% noise parameters and 20% of Dataset II contains outliers (20% extra noise).

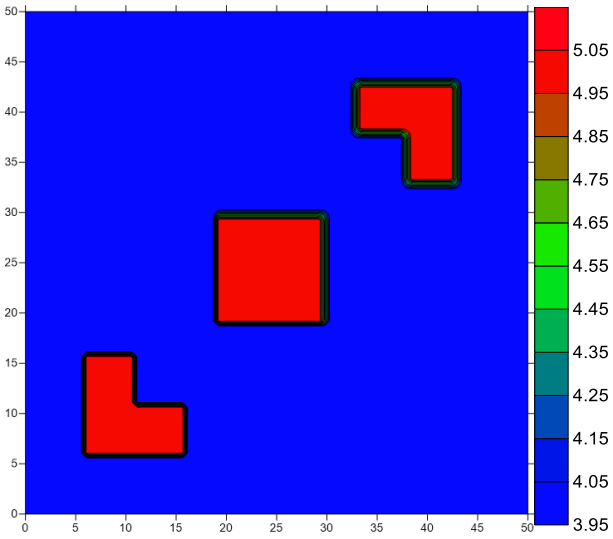

Figure 1. *The model used for numerical tests*

Dataset II is very noisy, and thus not a typical dataset in practice. However, to study the tomographic algorithms it is better to use a challenging dataset to demon-

strate the power of the method. (In all of our figures the velocity is measured in km/s units, the distances are assumed in units of cell's size along x and y directions, respectively.)

3. MODEL DISTANCE AND DATA DISTANCE

In order to characterize the accuracy of the reconstruction the (relative) model distance

$$D_{mod} = \sqrt{\frac{1}{M}\sum_{j=1}^{M}\left(\frac{x_j - x_j^{(0)}}{x_j^{(0)}}\right)^2} \tag{1}$$

was used. Here x_j and $x_j^{(0)}$ denotes the slownesses (reciprocal values of the propagation velocities) in the *j*-th cell of the reconstructed image and the model, respectively, and *M* is the number of cells. Similarly, the relative data distance was also calculated:

$$D_{dat} = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(\frac{t_k^{(input)} - t_k^{(reconstruaed)}}{t_k^{(input)}}\right)^2}\ , \tag{2}$$

where *N* is the number of travel time data.

4. CAUCHY-STEINER WEIGHTS AND THE METHOD OF ITERATIVELY REWEIGHTED LEAST SQUARES

Robust inversion can be achieved in various ways. To solve the least-absolute-deviation problem, the method of Iteratively Reweighted Least Squares (IRLS) was proposed by Scales [3]. Combining the two methods, the IRLS algorithm can be applied using Cauchy weights, defined as

$$W_{kk} = \frac{\sigma^2}{\sigma^2 + r_k^2}\ , \tag{3}$$

where σ^2 is the scale parameter and r_k is the *k*-th residual. The scale parameter σ^2 of the Cauchy distribution should be known a priori because the data residuals change from iteration to iteration. A condition for the lower bound of the scale parameter was derived by Amundsen [2].

In the framework of Steiner's Most Frequent Value (MFV) method [4] the scale parameter can be determined in an internal iteration. In the *(j+1)*-th step of this procedure the ε_{j+1}^2 (Steiner's scale factor) can be calculated, once ε_j^2 is known, as

$$\varepsilon_{j+1}^2 = 3\frac{\sum_{k=1}^{N}\frac{r_k^2}{\left(\varepsilon_j^2 + r_k^2\right)^2}}{\sum_{k=1}^{N}\left(\frac{1}{\varepsilon_j^2 + r_k^2}\right)^2}, \tag{4}$$

where the starting value ε_0 in the *0*-th step is given as

$$\varepsilon_0 \le \frac{\sqrt{3}}{2}\left(r_{max} - r_{min}\right). \tag{5}$$

It can be seen that the above procedure derives the scale parameter from the data set (deviation between measured and calculated data). The stop criterion can be easily defined by experience (for example by a fixed number of iterations). After this the Cauchy-Steiner weights can be calculated by inserting the Steiner's scale parameter (given in the last step of the above internal iterations) into the Cauchy formula (3), which gives the form

$$w_k = \frac{\varepsilon^2}{\varepsilon^2 + r_k^2}. \tag{6}$$

In a Cauchy-Steiner weighted inverse problem the objective function

$$E_w = (\vec{r}, \underline{\underline{W}}\vec{r}) \tag{7}$$

is minimized using the Iteratively Reweighted Least Squares method. In the framework of this algorithm a *0*-th order solution is derived using the (non-weighted) LSQ method and the weights are calculated as

$$w_k^{(0)} = \frac{\varepsilon^2}{\varepsilon^2 + (r_k^{(0)})^2} \tag{8}$$

with

$$r_k^{(0)} = t_k^{measured} - t_k^{(0)}, \tag{9}$$

where the $t_k^{(0)}$ travel times are calculated on the slowness field given by solving the LSQ problem. In the first iteration the objective function

$$E_w^{(1)} = \sum_{k=1}^{N} w_k^{(0)} (r_k^{(1)})^2 \tag{10}$$

is minimized, resulting in the linear set of normal equations

$$\underline{\underline{D}}^T \underline{\underline{W}}^{(0)} \underline{\underline{D}} \, \delta \vec{s} = \underline{\underline{D}}^T \underline{\underline{W}}^{(0)} \delta \vec{t} \tag{11}$$

of the weighted Least Squares method where the $\underline{\underline{W}}^{(0)}$ weighting matrix is of the diagonal form

$$W_{kk}^{(0)} = w_k^{(0)}. \tag{12}$$

Here $\underline{\underline{D}}$ is the distance matrix with the D_{kj} elements giving the length of the ray section in the *j*-th cell belonging to the *k*-th ray, $\vec{s}$ is the slowness vector, $\vec{t}$ is the travel time vector. This procedure is repeated, giving the typical *j*-th iteration step

$$\underline{\underline{D}}^T \underline{\underline{W}}^{(j-1)} \underline{\underline{D}} \, \delta \vec{s} = \underline{\underline{D}}^T \underline{\underline{W}}^{(j-1)} \delta \vec{t} \tag{13}$$

with the $\underline{\underline{W}}^{(j-1)}$ weighting matrix

$$W_{kk}^{(j-1)} = w_k^{(j-1)}. \tag{14}$$

(In these steps the normal equation is linear, because the weights are always calculated in a previous step. Here we note that each step of these iterations contains an internal loop for the determination of the Steiner's scale parameter.) This iteration is repeated until a proper stop criterion is met.

5. CONJUGATE GRADIENT METHOD

In order to solve the normal equations of the type

$$\underline{\underline{D}}^T \underline{\underline{D}} \, \vec{x} = \underline{\underline{D}}^T \vec{b} \, , \tag{15}$$

Scales [1] developed a tomographically very efficient variant of the Conjugate Gradient method. In order to solve the normal equations of the weighted least squares

method, this is modified as follows. Let $\vec{x}_o$ be an initial estimate and compute the vectors

$$\vec{s}_o = \vec{b} - \underline{\underline{D}}\,\vec{x}_o, \qquad \vec{r}_o = \vec{p}_o = \underline{\underline{D}}^T \underline{\underline{W}}\left(\vec{b} - \underline{\underline{D}}\,\vec{x}_o\right), \qquad \vec{q}_o = \underline{\underline{D}}\vec{p}_o \tag{16}$$

and start the iteration

$$\alpha_{k+1} = \frac{\left(\vec{r}_k, \vec{r}_k\right)}{\left(\vec{q}_k, \vec{q}_k\right)} \tag{17}$$

$$\vec{x}_{k+1} = x_k + \alpha_{k+1}\vec{p}_k \tag{18}$$

$$\vec{s}_{k+1} = \vec{s}_k - \alpha_{k+1}\vec{q}_k \tag{19}$$

$$\vec{r}_{k+1} = \underline{\underline{D}}^T \underline{\underline{W}}\,\vec{s}_{k+1} \tag{20}$$

$$\beta_{k+1} = \frac{\left(\vec{r}_{k+1}, \vec{r}_{k+1}\right)}{\left(\vec{r}_k, \vec{r}_k\right)} \tag{21}$$

$$\vec{p}_{k+1} = \vec{r}_{k+1} + \beta_{k+1}\vec{p}_k \tag{22}$$

$$\vec{q}_{k+1} = \underline{\underline{D}}\vec{p}_{k+1} \tag{23}$$

where $k = 0,1,2...$ refers to the iteration number. This procedure differs from the ordinary Conjugate Gradient algorithm only at two points, where beside the transpose of the $\underline{\underline{D}}$ matrix, also the $\underline{\underline{W}}$ weight matrix appears.

5.1. Conjugate Gradients with Gaussian dataset

At first the original (non-weighted) CG algorithm was applied for reconstruction of Dataset I. The resulting velocity distribution is shown in *Figure 2*.

The model distance is around 6%, which is quite an acceptable result. The image shows a relatively good reconstruction, as the CG algorithm minimizes the L_2 norm and solves the normal equation of the LSQ method, although the dataset contains 1% Gaussian noise.

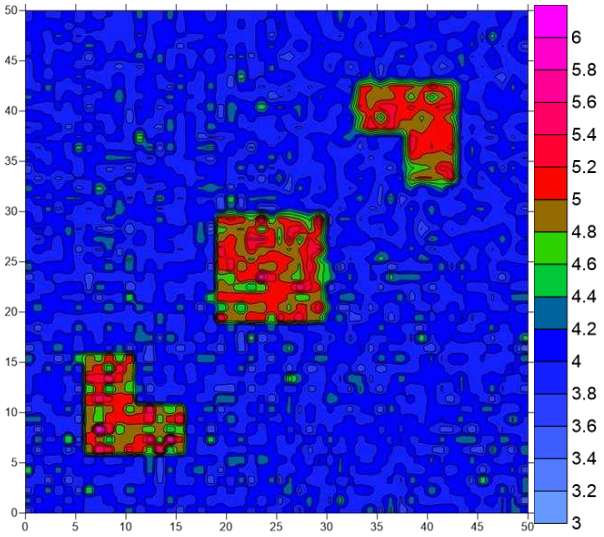

Figure 2. *Tomographic CGRAD inversion of Dataset I (data distance: 0.00948; model distance: 0.0579)*

5.2. Weighted Conjugate Gradients (W-CGRAD) with Gaussian data

If the ordinary Weighted Conjugate Gradients (W-CGRAD) algorithm is applied to Gaussian data (Dataset I), the velocity distribution looks as presented in *Figure 3*.

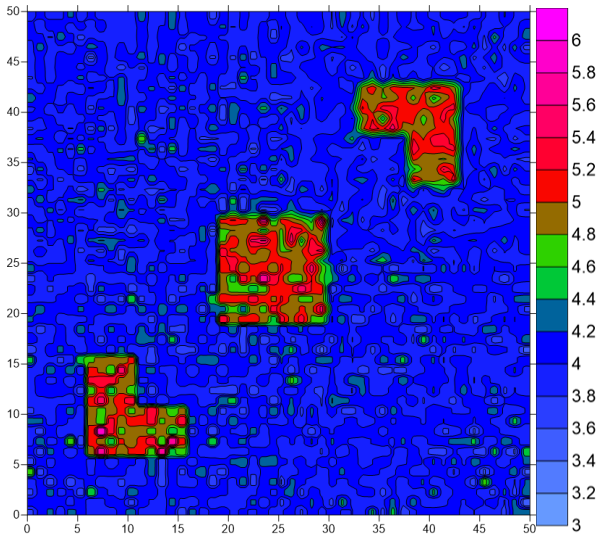

Figure 3. *Tomographic W-CG inversion of Dataset 1 (model distance: 0.0641)*

The Conjugate Gradients result in *Figure 2* is slightly better, because the W-CGRAD method solves the Gaussian least squares problem and the dataset follows a Gaussian distribution (the optimal method is the LSQ without weighting).

5.3. Conjugate Gradients with non-Gaussian dataset

If the ordinary Conjugate Gradients (CGRAD) algorithm is applied to non-Gaussian data (Dataset II), the velocity distribution looks as presented in *Figure 4*.

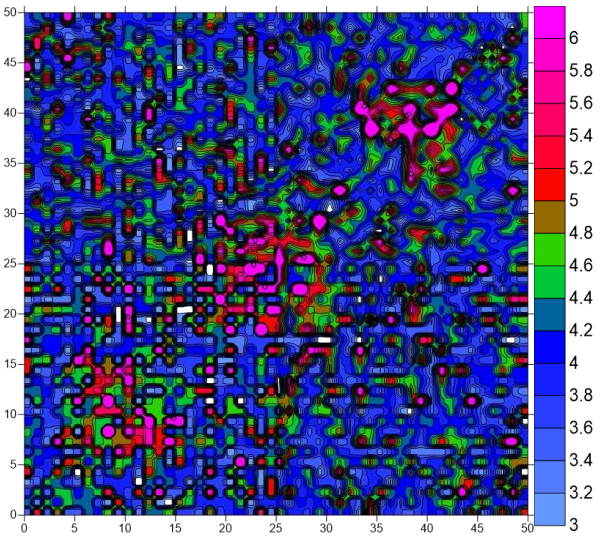

Figure 4. *Ordinary tomographic CGRAD inversion of Dataset II containing outliers (model distance: 0.250)*

In this case the model distance is 25%, which is really high for the tomography model. The image is highly distorted and almost unrecognizable. This result shows that the simple Gaussian Least Square Method is very sensitive to non-Gaussian noises, for example outliers in the datasets.

5.4. Weighted Conjugate Gradients with non-Gaussian dataset

If the ordinary Conjugate Gradients algorithm is modified to solve the normal equations of the Weighted LSQ method with the use of Cauchy-Steiner Weights, a robust tomography method can be defined (W-CGRAD). *Figure 5* shows the reconstructed image where the weighted version of CG algorithm is applied to process Dataset II. It can be seen that the influence of outliers is highly reduced due to the weighting and a highly improved image is obtained. This result shows that the use of Cauchy-Steiner weights is very efficient in reducing the influence of outliers.

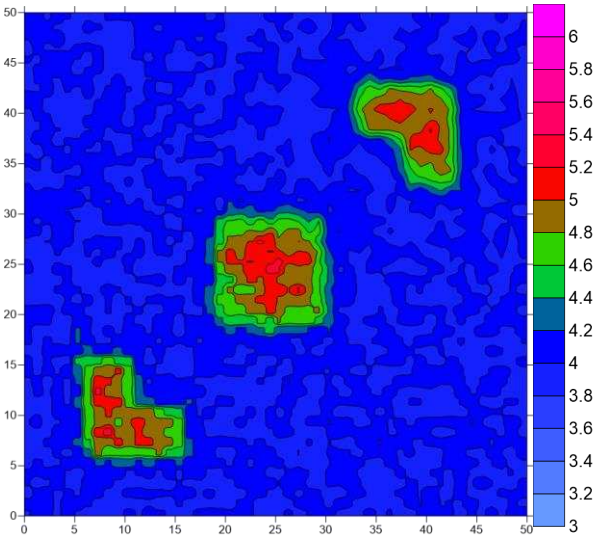

Figure 5. *The tomographic reconstruction of Dataset II (containing outliers) using the weighted W-CGRAD method (Model distance: 0.0871)*

6. SIMULTANEOUS ITERATIVE RECONSTRUCTION TECHNIQUE (SIRT)

Firstly the back projection method was the first to be used for seismic tomography. Later the Least Square problems in tomography came to be solved by Algebraic Reconstruction Technique (ART) and Simultaneous Iterative Reconstruction Technique (SIRT). SIRT remains one of the most frequently used iterative methods in seismic tomography with the improvement formula

$$x_j^{(q+1)} = x_j^{(q)} + \frac{1}{Q_j} \sum_{i=1}^{Q_j} \frac{D_{ij} r_i^{(q)}}{\sum_k D_{ik}^2}. \tag{24}$$

Q_j is the number of traces running through the *j*-th cell, $r_i^{(q)}$ means the difference between the *i*-th measured and calculated data and D_{ij} is the ray section of the *i*-th ray in the *j*-th cell.

As can be seen, the SIRT formula calculates the arithmetic mean of the ART corrections calculated for the given (*j*-th) cell. It is well known that arithmetic mean is efficient only if the data noise follows Gaussian distribution. In our case, we also have non-Gaussian noise distribution. Therefore, we have to use the weighted mean instead of arithmetic one

$$x_j^{(q+1)} = x_j^{(q)} + \frac{1}{\sum_{l=1}^{Q_j} W_{ll}} \sum_{i=1}^{Q_j} W_{ii} \frac{D_{ij} r_i^{(q)}}{\sum_k D_{ik}^2}, \tag{25}$$

where Q_j is the number of traces covered in the *j*-th cell, $r_i^{(q)}$ means the difference between measured and calculated data, D_{ij} is the ray section of the *i*-th ray trace in the *j*-th cell and W_{ii} is the Cauchy-Steiner weight belonging to the *i*-th travel time.

6.1. Simultaneous Iterative Reconstruction Technique (SIRT) with Gaussian dataset

It is well known that the SIRT method is one of the best methods in tomography when the distribution of noises follows Gaussian statistics. This can be proved by using Dataset I in a SIRT reconstruction.

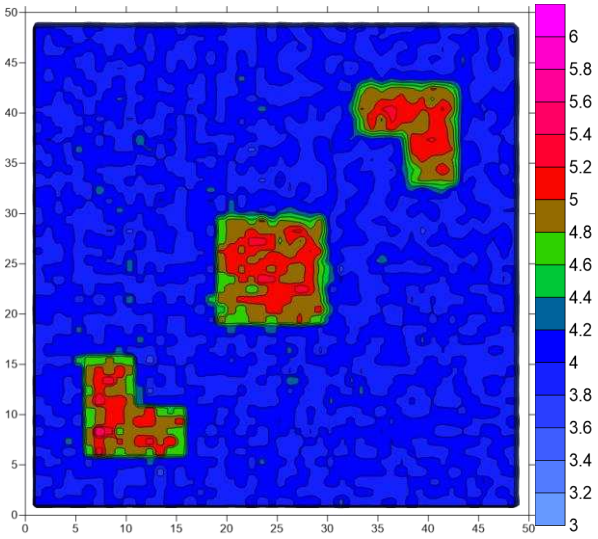

Figure 6. *Tomographic SIRT inversion of tDataset I (data distance: 0.00973; model distance: 0.0216)*

Figure 6 shows that the SIRT method gives nearly three times better model distance compared to the results of CG method.

6.2. Weighted Simultaneous Iterative Reconstruction Technique (W-SIRT) with Gaussian dataset

As shown in *Figure 7*, using Dataset I the W-SIRT method gives similar reconstruction to that given by the traditional SIRT *(Figure 6)*, which seems to be slightly better; however, there is only a negligible difference in the model distances.

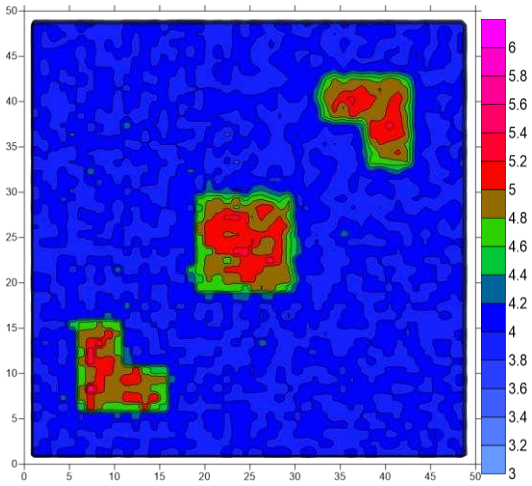

Figure 7. *Tomographic W-SIRT inversion of Dataset I (model distance: 0.0227)*

6.3. Simultaneous Iterative Reconstruction Technique with non-Gaussian dataset

The SIRT method produces a relatively distorted image in the case of Dataset II, containing outliers. This is demonstrated in *Figure 8*.

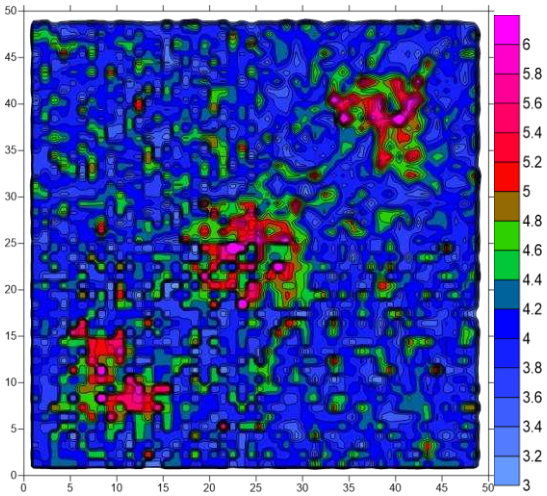

Figure 8. *The tomographic SIRT inversion of Dataset II (model distance: 0.0635)*

As can be seen, the SIRT method is also very sensitive to the non-Gaussian nature of the noise, though it is better than CG with the non-Gaussian dataset (the model

distance for CG with non-Gaussian data was 0.250). In spite of this fact, the reconstruction in *Figure 8* is not acceptable.

6.4. Weighted Simultaneous Iterative Reconstruction Technique (W-SIRT) with non-Gaussian dataset

If we use the W-SIRT method for reconstructing Dataset II, we obtain the velocity distribution shown in *Figure 9*. It can be seen that the algorithm using Cauchy-Steiner weights is highly resistant to outlier data and gives a good reconstruction result.

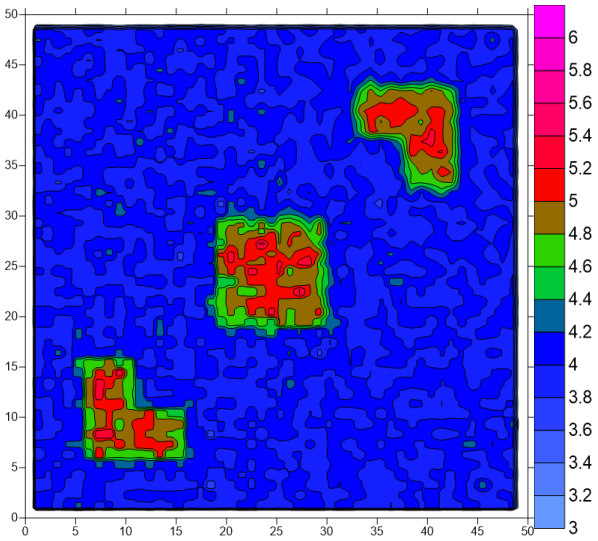

Figure 9. *The tomographic reconstruction of Dataset II by the W-SIRT method (model distance: 0.0242)*

The model distance is equal to 0.0242, which is nearly the same as when the SIRT method was applied on the Gaussian dataset, where the model distance was 0.026. Of course, SIRT with the Gaussian dataset should give a better result than W-SIRT with the non-Gaussian dataset. Investigations with outlier data proved that the W-SIRT algorithm is computationally noise resistant and computationally economic.

7. CONCLUSIONS

The proposed tomography algorithms are tested for a medium-size tomography example using synthetic travel time data. It is proved that, compared to their traditional versions, the outlier sensitivity of the generalized tomography methods is sufficiently reduced. Compared to the traditional CG algorithm the new W-CGRAD algorithm is more robust and resistant to outliers. The SIRT method was also modified by using

Cauchy-Steiner weights and it was proved that, compared to its original version, the W-SIRT method is less sensitive to outlier data.

ACKNOWLEDGEMENTS

The research work of author Utku Kale was supported by the Stipendium Hungaricum Scholarship Program. The research of Mihály Dobróka was supported by the National Research, Development and Innovation Office – NKFIH, Project No. K 109441.

REFERENCES

[1] SCALES, J. A.: Tomographic inversion via the conjugate gradient method. *Geophysics*, 1987, 52, 179–185.

[2] AMUNDSEN, L.: Comparison of the least-squares criterion and the Cauchy criterion in frequency-wavenumber inversion. *Geophysics*, 1991, 56:(12), 2027–2038.

[3] SCALES, J. A.–GERSZTENKORN, A.–TREITEL, S.: Fast l_p solution of large, sparse, linear systems: Application to seismic travel time tomography. *Journal of Computational Physics*, 1988, 75. 314–333.

[4] STEINER, F.: Most frequent value procedures. (A short monograph). *Geophysical Transactions*, 1988, 34 (2–3), 139–260.

Geosciences and Engineering, Vol. 5, No. 8 (2016), pp.111–123.

LABORATORY INDUCED POLARIZATION DATA PROCESSED WITH SERIES EXPANSION INVERSION

ANETT KISS[1]–MIHÁLY DOBRÓKA[1]*–ENDRE TURAI[1]–
JUDIT SOMOGYINÉ MOLNÁR[2,3]

[1]*Department of Geophysics, University of Miskolc, Hungary*
[2]*Institute of Electrical & Electronic Engineering, University of Miskolc*
[3]*MTA-ME Geoengineering Research Group*
**dobroka@uni-miskolc.hu*

Abstract: Induced Polarization (IP) measurements are often used in ore exploration. The decaying potential is characteristic for polarizable rocks after the excitation current is turned off. From the data, the apparent polarizability curve can be deduced. However, the polarization effects are hidden in the time constant spectrum which is connected to the polarizability through an integral expression. This serves as the forward modelling in the inversion procedure. In this paper the efficiency of series expansion based inversion is investigated on laboratory time-domain IP data. Cell-wise constant functions are chosen for basis functions. The unknowns/model parameters of the problem are the series expansion coefficients. After they are estimated within the inversion procedure, the time-constant spectrum containing the polarization effects can be determined. Results prove the applicability of the method to rock samples measured in the laboratory.

Keywords: *induced polarization, time-constant spectrum, series expansion, inversion*

1. INTRODUCTION

Electromagnetic methods in geophysics have a variety of applications, not only in mineral or fluid exploration, but in connection with engineering and environmental problems as well. One of these is the Induced Polarization (IP) method, which utilizes the different polarization effects arising when an artificial source induces current in the given medium, resulting in a measureable potential difference. Schlumberger was the first (in the 1920s) to observe this phenomenon. It was noticed during resistivity measurements that the potential measured between the measuring electrodes does not equal zero immediately after turning off the excitation current; instead, its value follows a decay curve. Despite the early observations, the first mathematical formulation of the phenomenon was described only in 1959 by Seigel [1]. Beside its many successful applications in ore exploration [1, 2, 3, 4], the advantage of the method has been proven in hydro-geophysics [5, 6], archeology [7, 8], and geology including structures or sand-clay layers [9, 10], as well as in the detection and characterization of contaminated areas such as waste dumps [11, 12].

Investigations in connection with ores have been executed not only in the field, but in laboratories as well [13, 14]. Measurements can be done in both frequency and time domains. We shortly describe the essence of the former one, but our present investigations focus on time-domain IP measurements, and thus this method is described in more detail. At the end of the 1970s and beginning of the 1980s intensive development of IP method and measuring system took place at the Department of Geophysics, University of Miskolc. Nowadays these very precise and valuable time-domain data sets [15] measured on 21 rock samples containing different ore minerals are being reprocessed. For this purpose we follow the inversion development strategy of the Department of Geophysics. Several studies have proven the effectiveness of series expansion based inversion in geophysical data processing [16, 17, 18, 19]. It was successfully applied in field IP measurements, as well [20, 21]. Now we show how it can be utilized for processing laboratory TDIP measurements and determining the time constant spectra containing all of the information about polarization effects.

2. INDUCED POLARIZATION METHOD

It was observed during field resistivity measurements that the potential measured between two electrodes does not equal zero immediately after turning off the excitation current. Rocks were losing their polarization only after a given decay time. This polarizability may arise from several polarization effects, such as

- membrane polarization (caused by the negative charges on the surface of clay particles when the diameter of pores is smaller than the diameter of positive ions),
- filtration polarization (because of the different positive and negative ion mobility),
- redox polarization (in rocks containing oxidative or reductive chemical components),
- electrode (metallic) polarization (electrochemical interaction occurs between the ionic solution of the pore space and the electrically conductive particles of the rock).

In ore exploration the Induced Polarization (IP) method is often used, because the metallic polarizability provides a good basis for distinguishing the ore-rich zones from waste rock. It indicates mostly sulfidic ores (pyrites, chalcopyrite, molybdenite, galena) and some oxides (magnetite, ilmenite), as well as graphite. Several successful field applications [1, 2, 3, 4, 13] have confirmed the effectiveness of the method.

IP measurements are a kind of electromagnetic method that can be done both in the frequency and in the time domain. The principle of Induced Polarization measurements in the frequency domain (FDIP) is that the apparent specific resistivity depends on the frequency. Therefore it is measured at a series of frequencies.

Based on the measured complex specific resistivity data the amplitude and the phase angle of specific resistivity can be derived. The resistivity spectrum is most frequently presented in double logarithmic diagrams showing amplitude and phase versus frequency. The spectra and the derived Cole-Cole parameters provide tools for tasks such as distinguishing different ore types (disseminated, vein-textured etc.) [22].

Our investigations focus on Time-Domain Induced Polarization (TDIP) measurements. An example of an electrode array is shown in *Figure 1* for field measurements. The excitation current is led into the rock through electrodes *A* and *B*, while the induced potential difference is measured between the electrodes *M* and *N*.

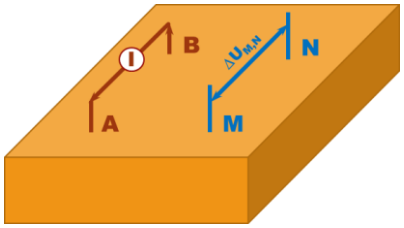

Figure 1. *Example of electrode array during field measurements*

The essence is that the rocks electrically saturate when the excitation current is turned on. After the stabilization of the potential during excitation (U_0), the excitation current is turned off and the decreasing potential measureable between the measuring electrodes ($U(t)$) is registered at given sampling intervals. Based on these values the apparent polarizabilities η_a are calculated

$$\eta_a(t) = \frac{U(t)}{U_0}. \tag{1}$$

3. LABORATORY MEASUREMENTS

Laboratory measurements on rocks in the time domain are very difficult to find in the literature compared to frequency domain (SIP-Spectral Induced Polarization) investigations. Some publications report about results obtained on artificially mixed "rocks". For example Gurin et al. [23] studied mixtures of calibrated sand and ores, while Sumi [24] determined the polarizability of ore minerals in his laboratory model. The principle of laboratory measurements is the same as in the large scale field investigations, but the instrumentation requires more attention – how to grab the sample, avoid electrode polarization, deal with external electric noises, etc. As a result of decades of fruitful method and instrument development at

the Department of Geophysics [15, 25], very precise laboratory measurement data providing apparent polarizability decay curves are available.

The samples are from different geological settings and contain different types of ores. Before the measurements each sample was soaked in distilled water for 24 hours, and then was placed into the sample holder, which was attached to non-polarizable electrodes through permeable walls. The whole sample holder was placed in a metallic cell to protect it from outer electrical noise. The current electrodes were attached to a bipolar current generator [25]. The applied current strength was 10 μA, and first the spontaneous potential was determined. After 15 minutes excitation, the generator was turned off and the potential between the electrodes was measured for 972.8 seconds [25]. The sampling time was determined according to the algorithm

$$t_{k,i} = t_{0,0} a^k (1 + \gamma i), \tag{2}$$

where

$$t_{0,0} = 0.125 sec,\ a = 2,\ \gamma = 0.1,\ i = 0,1,...,9,\ k = 0,1,...,12.$$

4. DATA PROCESSING

In geophysical investigations it was proved that series expansion based inversion is a very effective and valuable tool for data processing. The goal of inversion method development is to process and interpret data measured above complex (laterally and vertically inhomogeneous) geological structures with series expansion discretization in an inversion procedure, where the problem is defined for the series expansion coefficients. The greatest advantage of these methods is that we can obtain appropriate resolution by introducing relatively few series expansion coefficients, so that the problem to be solved leads to a (relatively small) overdetermined inverse problem [16, 17, 18, 19].

Turai et al. [20] applied the series expansion inversion based data processing method to field IP data to solve environmental problems. In this paper the method is used to process laboratory IP data measured on core samples containing ores, to determine the time-constant spectrum and to distinguish different polarization effects.

The forward modeling of the problem was formulated by Turai [26] as

$$\eta_a(t) = \int_0^\infty w(\tau) \exp(-t/\tau) d\tau, \tag{3}$$

where t is the time after turning off the excitation current, τ is the time-constant, and $w(\tau)$ is the time-constant spectrum. The aim is to determine the time-constant spectrum that contains all the polarization effects and can be obtained with the TAU-transformation introduced by Turai [26]

$$w(\tau) = TAU\{\eta_a(t)\}. \tag{4}$$

This can be solved with different methods; for its approximation Turai [27] suggested polynomial interpolations and a Fourier-series solution. In this paper the TAU-transformation is solved as an inverse problem. Since the forward modeling is linear, the inverse problem is linear as well.

First discretization is needed, i.e. the unknown time-constant spectrum characterized with a finite number of parameters should be determined. This operation is completed in the form of series expansion according to a properly chosen basis function system Φ_q

$$w(\tau) = \sum_{q=1}^{Q} B_q \, \Phi_q(\tau), \tag{5}$$

where the series expansion coefficients B_q mean the unknown parameters of the problem and $\Phi_q(\tau)$ indicates the known basis functions. Turai and Dobróka [28] applied series expansion according to Dirac delta functions capable of describing the "lineal" spectrum. For our task we have chosen the cell-wise constant functions for basis functions, which provide a "striped" time-constant spectrum

$$\Phi_q(\tau) = \begin{cases} 1 & \text{if } \tau_{q-1} < \tau < \tau_q \\ 0 & \text{else} \end{cases}, \tag{6}$$

where Q means the number of series expansion parts ($q = 1,2,...,Q$). Substituting *Equation (5)* into the response equation, *Equation (3)* can be rewritten in the form

$$\eta_a(t_k) = \sum_{q=1}^{Q} B_q \int_0^{\infty} \Phi_q(\tau) \exp(-t_k / \tau) d\tau, \tag{7}$$

which provides the (theoretical) apparent polarizability value at t_k ($k = 1,2,...,K$) after the excitation current is turned off, where K is the number of measured data. Let us introduce the notation G_{kq} for the integral in *Equation (7)*. With the use of

cell-wise constant basis functions $\Phi_q(\tau)$ [see *Equation (6)*], the Jacobian matrix obtains the simple form for the interval of interest

$$G_{kq} = \int_{\tau_{q-1}}^{\tau_q} exp(-t_k / \tau) d\tau \,. \tag{8}$$

As a result, the calculated apparent polarizability data can be calculated easily

$$\eta_k^{calculated} = \sum_{q=1}^{Q} B_q \, G_{kq} \,. \tag{9}$$

Introducing the deviation vector as $\vec{e} = \vec{\eta}^{measured} - \vec{\eta}^{calculated}$, the objective function of the inverse problem can be defined as its L2 norm. To obtain more stability during the inversion the damped least squares method is used, which solution for the series expansion coefficients can be written as

$$\vec{B} = (\underline{\underline{G}}^T \underline{\underline{G}} + \varepsilon^2 \underline{\underline{I}})^{-1} \underline{\underline{G}}^T \vec{\eta}^{measured}, \tag{10}$$

where ε is the damping factor, $\underline{\underline{I}}$ is the unity matrix. The elements of the model parameter vector in the last iteration step are considered the solution, which can be substituted into *Equation (5)* together with the known basis functions providing the time-constant spectrum.

5. APPLICATION

For this paper data sets measured on three cylindrical rock samples (No. ***1***, ***14*** and ***16***) were chosen to study the applicability of our series expansion based inversion method in the time-constant spectrum determination. All samples originate from Recsk, Hungary. In general, the metasomatic skarn ore body was formed at the contact of an upper Miocene diorite intrusion and Triassic limestone at 1 km depth. Chalcopyrite in the exoskarn, sphalerite farther from the contact, and porphyric Cu mineralization in the intrusion are characteristic [29]. Unfortunately the research report [15] contains a very poor description of the measured samples:

- Sample ***1*** – Fe-Cu skarn – deep level: 900 m
- Sample ***14*** – chalcopyrite (skarn)
- Sample ***16*** – metasomatic polymetallic Zn-Pb-Cu-Fe ore (deep level).

The potential difference was measured at 130 discrete sampling points [for reference times see *Equation (2)*] after turning off the excitation current, resulting in

130 apparent polarizability data for each sample. They are plotted with blue dots in *Figures 2a*, *3a* and *4a* for samples ***1***, ***14*** and ***16***, respectively. It can be clearly seen that these rock samples were very strongly polarized, since the polarizability values calculated from the measured potential differences after the turning off the excitation current exceeded 0.80 (80 %) at the beginning of the decay curve. As the samples contain ore minerals (Fe, Cu, Zn, Pb), this phenomenon can be explained by metallic polarization. The most remarkable potential drop happened in the first 200 s, after which the decay curve became less steep. *Table 1* shows the polarizability values at the first (0.125 s) and the last (972.8 s) sampling times.

Table 1

Measured polarizability values at the first and last sampling times

Sample	**η [–]** ***(t = 0.125 s)***	**η [–]** ***(t = 972.8 s)***
1	0.97892	0.16859
14	0.96318	0.7465
16	0.82470	0.6341

During the inversion process the range of time-constants was divided into 90 logarithmically equidistant intervals, i. e. the number of cell-wise constant basis functions and the M number of model parameters (series expansion coefficients) was equal to 90 during the discretization. $N=130$ measured apparent polarizability data were inverted, providing an overdetermination ratio (N/M) of 1.44. To make the inversion stable, the damped least squares method was chosen with a starting damping factor $\varepsilon = 0.01$, which was multiplied by a 0.9 factor in each iteration step. Experiments showed that 60 iterations provide acceptable results; further calculations brought about no change in the model parameters. The apparent polarizability data calculated with inversion are plotted with red curves in *Figures 2a*, *3a* and *4a* together with the measured ones. The small distances between measured and calculated data can be seen on the graphs. To characterize the accuracy of data fit of the inversion estimates we also calculated the root mean squares (RMS) by applying the formula

$$RMS = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(\frac{\eta_k^{measured} - \eta_k^{calculated}}{\eta_k^{measured}}\right)^2} \cdot 100\,[\%], \tag{11}$$

where $\eta_k^{measured}$ is the k-th measured data and $\eta_k^{calculated}$ is the k-th calculated data. As can be seen in *Table 2*, the RMS values are below 1%, characterizing high quality measurement data sets and an accurate inversion procedure.

Table 2
Distances in data space

Sample	RMS [%]
1	0.60
14	0.68
16	0.71

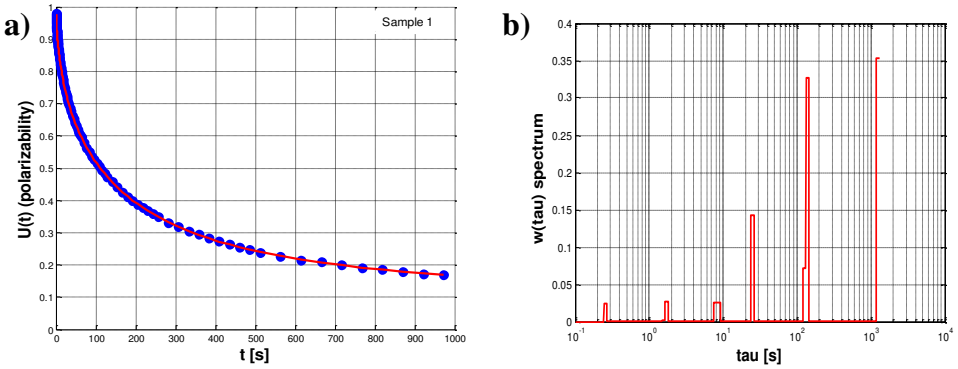

Figure 2
(a) *Decay curve of the measured and calculated apparent polarizabilities.* ***(b)*** *Time constant spectrum of Sample* ***1*** *determined by series expansion based inversion*

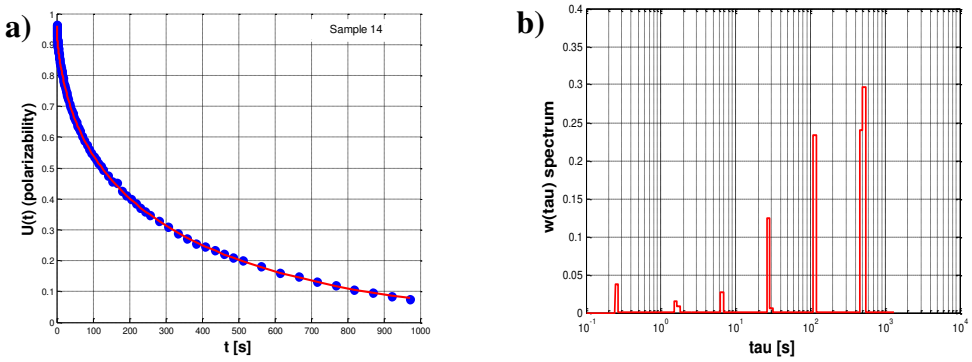

Figure 3
(a) *Decay curve of the measured and calculated apparent polarizabilities.* ***(b)*** *Time constant spectrum of Sample* ***14*** *determined by series expansion based inversion*

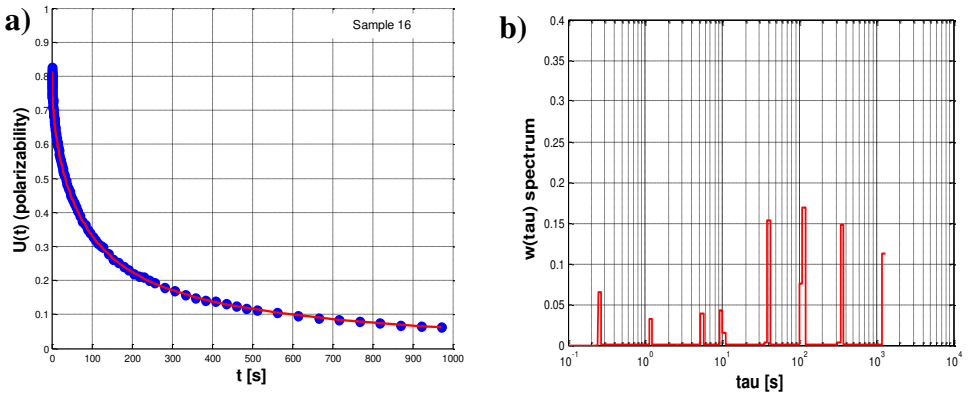

Figure 4

(a) *Decay curve of the measured and calculated apparent polarizabilities.* ***(b)*** *Time constant spectrum of Sample* ***16*** *determined by series expansion based inversion*

Based on the series expansion coefficients estimated with inversion, the time-constant spectra were determined by applying *Equation (5)*. The time constants cover a large interval in the given samples $\tau = [0 - 1300]$ (sec). To make the peaks of the spectra indicating polarization effects more visible, logarithmic scale was chosen for the time-constant axis.

The spectra of the three samples are plotted in *Figures 2b*, *3b* and *4b*. As can be seen, the highest amplitudes appear in the range of higher τ (above 30 or even 100 seconds). For Sample ***1*** six peaks are visible on the spectrum with growing amplitudes at $\tau \approx 0.27\,s$, $\tau \approx 1.8\,s$, $\tau \approx 8.3 - 9.1\,s$, $\tau \approx 26.11\,s$, a small amplitude at $\tau \approx 131.98\,s$ is followed by the peak at $\tau \approx 145.18\,s$ with high amplitude. The last peak appears at $\tau \approx 1300\,s$.

In the case of Sample ***14***, the spectrum looks slightly different. The very high time-constant ($\tau > 1000\,s$) is missing. Peaks appear at $\tau \approx 0.27\,s$, $\tau \approx 1.6 - 1.8\,s$, $\tau \approx 6.8\,s$, $\tau \approx 28\,s$ with a very small amplitude "shoulder" at $\tau \approx 31\,s$, $\tau \approx 120\,s$ followed by the highest peak at $\tau \approx 500 - 550\,s$ time constants.

The time constant spectrum of Sample ***16*** contains 8 peaks at $\tau \approx 0.27\,s$, $\tau \approx 1.2\,s$, $\tau \approx 5.6\,s$, $\tau \approx 10 - 11\,s$, $\tau \approx 42\,s$, $\tau \approx 100 - 120\,s$, $\tau \approx 377\,s$, $\tau \approx 1300\,s$.

It can be seen from the results that the time-constant spectrum can be determined with appropriate resolution if apparent polarizability data are processed with series expansion based inversion. Thus, different polarization effects can be distinguished. From the graphs it can be concluded that the spectra seem to have nearly lineal characteristic. A more detailed geological description of the samples and a more comprehensive study of the spectrum are required to link the peaks in the time-constant spectrum to geological features.

6. CONCLUSIONS

Time-domain Induced Polarization measurements are often used during ore exploration, and several studies have confirmed its successful field application. In this paper we showed that laboratory TDIP data can be successfully processed with series expansion based inversion. After the determination of series expansion coefficients in the framework of inversion, the time-constant spectrum was calculated, where more polarization effects were distinguished. Further research is needed to investigate the relation of spectrum peaks appearing at different time constants and lithological composition of rocks. For this a more detailed geological description of the samples is needed as well.

ACKNOWLEDGEMENT

The research of Anett Kiss was supported by the ÚNKP-16-3. New National Excellence Program of the Ministry of Human Capacities. The research of Mihály Dobróka and Endre Turai was supported by the National Research, Development and Innovation Office – NKFIH, project No. K 109441. The research of Judit Somogyiné Molnár was carried out in the framework of the Terplán Zénó Program.

LIST OF SYMBOLS

Symbol	Description	Unit
a	constant	–
$\vec{B}$	vector of series expansion coefficients (model parameters)	–
$\vec{e}$	deviation vector	–
$\underline{\underline{G}}$	Jacobian matrix	–
$\underline{\underline{G}}^T$	transposed Jacobian matrix	–
i	index	–
$\underline{\underline{I}}$	unity matrix	–
k	index	–
K	number of measured data	–
M	number of model parameters	–
N	number of data	–
q	index	–
Q	number of series expansion coefficients	–
RMS	Root Mean Square	%
t	sampling time	s
$TAU\{\ \}$	operator of TAU-transformation	–

Symbol	Description	Unit
$U(t)$	potential at time t after the excitation current is turned off	V
U_0	potential during excitation	V
$w(\tau)$	time constant spectrum	–
ε	damping factor	–
$\Phi(\tau)$	basis function (known)	–
γ	index	–
η_a	apparent polarizability	–
$\vec{\eta}^{measured}$	vector of measured polarizability data	–
τ	time constant	s

REFERENCES

[1] SEIGEL, H. O.: Mathematical formulation and type curves for induced polarization. *Geophysics*, 24:(3), 1959, pp. 547–565.

[2] WAIT, J. R.: *Overvoltage Research and Geophysical Applications*. London, 1959, Pergamon Press.

[3] KELLER, G. W.–FRISCHKNECHT, F. C.: *Electrical Methods in Geophysical Prospecting*. Oxford, 1966, Pergamon Press.

[4] TAVAKOLI, S.–BAUER, T. E.–RASMUSSEN, T. M.–WEIHED, P.–ELMING, S. A.: Deep massive sulphide exploration using 2D and 3D geoelectrical and induced polarization data in Skellefte mining district, northern Sweden. *Geophysical Prospecting*, 64, 2016, pp. 1602–1619.

[5] JOHNSON, T. C.–VERSTEEG, R. J.–WARD, A.–DAY-LEWIS, F. D.–REVIL, A.: Improved hydrogeophysical characterization and monitoring through parallel modeling and inversion of time-domain resistivity and induced-polarization data. *Geophysics*, 75:(4), 2010, pp. WA27-WA41.

[6] TITOV, K.–TARASOV, A.–ILYIN, Y.–SELEZNEV, N.–BOYD, A.: Relationships between induced polarization relaxation time and hydraulic properties of sandstone. *Geophysical Journal International*, 180:(3), 2010, pp. 1095–1106.

[7] ASPINALL, A.–LYNAM, J.: Induced polarization as a technique for archaeological surveying. *Prospezioni Archeologiche*, 3, 1968, pp. 91–93.

[8] ABU ZEID, N.–CORRADINI, E.–BIGNARDI, S.–MORANDI, N.–NIZZO, V.–SANTARATO, G.: Unusual Geophysical Techniques in Archaeology – HVSR and Induced Polarization, A Case History. *Near Surface Geoscience 2016 Conference*, Barcelona, Paper We 22 A13, 2016.

[9] ILICETO, V.–SANTARATO, G.–VERONESE, S.: An approach to the identification of fine sediments by Induced Polarization laboratory measurements. *Geophysical Prospecting*, 30, 1982, pp. 331–347.

[10] GAZOTY, A.–FIANDACA, G.–PEDERSEN, J.–AUKEN, E.–CHRISTIANSEN, A. V.–PEDERSEN, J. K.: Application of time domain induced polarization to the mapping of lithotypes in a landfill site. *Hydrology and Earth System Sciences*, 16, 2012, pp. 1793–1804.

[11] VIEZZOLI, A.–CULL, J. P.–MASSIE, D.: Mapping fly-ash water pond leakage with TEM and IP data at Loy Yang coal-mine (Australia). *Near Surface Geophysics*, 4:(5), 2006, pp. 305–311.

[12] TURAI E.–BARACZA K.–SZILVÁSI M.–TÓTH M.–SZŰCS P.–KOVÁCS B.–FÖLDESSY J.–MADARÁSZ T.: Környezetszennyezések vizsgálata geoelektromos geofizikai módszerekkel. In: NÉMETH Norbert (szerk.): *Természeti erőforrásaink az észak-magyarországi térségben: Az előadások összefoglalói*. Sárospatak, 2016. Magyarhoni Földtani Társulat, pp. 11–12.

[13] VANHALA, H.–PELTONIEMI, M.: Spectral IP studies of Finnish ore prospects. *Geophysics*, 57:(12), 1992, pp. 1545–1555.

[14] CHEN, R.–HE, X.–YAO, H.–ZHAO, X.–SHEN, R.: A complex impedance measurement system for high impedance sample. *29th Annual Symposium on the Application of Geophysics to Engineering and Environmental Problems (SAGEEP)*, Denver, Colorado, USA, 2016.

[15] Research Report: *Jelentés a kőzetek és ásványok gerjesztett potenciál kőzetfizikai paraméterének méréséről.* Nehézipari Műszaki Egyetem, Geofizikai Tanszék, 1984.

[16] VASS P.–DOBRÓKA M.: Sorfejtéses inverzió I. Fourier transzformáció, mint inverz feladat. *Magyar Geofizika*, 2010. 50:(4), pp. 141–152.

[17] SZABÓ, N. P.: Shale volume estimation based on the factor analysis of well-logging data. *Acta Geophysica*,59:(5), 2011, pp. 935–953.

[18] DOBRÓKA, M.–VÖLGYESI, L.: Sorfejtéses inverzió IV. A nehézségi erőtér potenciál-függvényének inverziós előállítása. *Magyar Geofizika*, 2010. 51:(3), pp. 143–149.

[19] SZABÓ, N. P.: Hydraulic conductivity explored by factor analysis of borehole geophysical data. *Hydrogeology Journal*, 23:(5), 2015, pp. 869–882.

[20] TURAI, E.–DOBRÓKA, M.–HERCZEG, Á.: Sorfejtéses inverzió III. Gerjesztett polarizációs adatok inverziós feldolgozása. *Magyar Geofizika*, 2010. 51:(2), pp. 88–98.

[21] TURAI E.: Data Processing Method Developments using TAU-transformation of time domain IP data: II. Interpretation results of field measured data. *Acta Geodaetica et Geophysica Hungarica*, 46:(4), 2011, pp. 391–400.

[22] VANHALA, H.: Laboratory and field studies of environmental and exploration applications of the spectral induced-polarization (SIP) method. Report. *Geologian tutkimuskeskus*, Espoo, Finland, 1997.

[23] GURIN G.–TARASOV A.–ILYIN Y.–TITOV K.: Time domain spectral induced polarization of disseminated electronic conductors: Laboratory data analysis through the Debye decomposition approach. *Journal of Applied Geophysics*, 98, 2013, pp. 44–53.

[24] SUMI F.: The Induced Polarization Method in Ore Investigation. *20th Meeting of the European Association of Exploration Geophysicists*, Hague, 1961.

[25] Research report: *A gerjesztett polarizáció (GP) dinamikus kőzetfizikai paraméterének meghatározására szolgáló műszer-módszer továbbfejlesztése*. Nehézipari Műszaki Egyetem, Geofizikai Tanszék, 1979.

[26] TURAI, E.: GP time-domain görbék Tau-transzformációja. *Magyar Geofizika*, 1981. 22:(1), pp. 29–36.

[27] TURAI, E.: TAU-transformation of time-domain IP curves. *Annales Univ. Sci. Budapestinensis De R. Eötvös Nominatae, Sectio Geophysica et Meteorologica*, Tomus I–II, 1985.

[28] TURAI, E.–DOBRÓKA, M.: Data processing method developments using TAU-transformation of time-domain IP data – I. Theoretical basis. *Acta Geodaetica et Geophysica Hungarica*, 43:(3), 2011, pp. 283–290.

[29] FÖLDESSY J., SZEBÉNYI G.: The mineralizations of the Recsk Deeps and Lahóca – short geological overview. In Recsk And Lahóca Geology Of The Paleogene Ore Complex, *Geosciences - Publications of the University of Miskolc, Series A, Mining*, 73, 2008, pp.85–98.

Geosciences and Engineering, Vol. 5, No. 8 (2016), pp.124–134.

HEAVY METAL SOURCE IDENTIFICATION AND ANALYSIS USING MULTIVARIATE STATISTICAL METHODS IN SOILS FROM AKUSE AREA, SOUTH-EASTERN GHANA

DANIEL ODURO BOATEY NUAMAH

Department of Geophysics, University of Miskolc
danielobnuamah54@gmail.com

Abstract: Multivariate statistical methods (principal component analysis, cluster analysis, and correlation analysis) have been applied to coastal soils of Akuse, Southeastern Ghana to determine heavy metal sources (Cu, Zn, Ni, Pb, Cr, Co and Ba). Thirty-four Composite soil samples were taken on a grid of 1km × 1km from the study area and analyzed using X-ray fluorescence analytical protocol. The study showed that: (i) Ni, Pb, Co and Ba had anthropogenic sources (ii) Zn and Cu were associated with parent materials and therefore had natural sources whilst (iii) Cr showed varied sources of both natural and anthropogenic. From the statistical mean estimation, Chromium (Cr) showed the highest mean concentration (mean = 442.8 ± 68.658 ppm). It was followed by Barium (Ba) (mean = 261 ± 137.88 ppm) and Nickel (mean = 83.94 ± 59.915 ppm). Elements such as copper, zinc and Cobalt showed average mean values ranging from (70.45 ppm – 47.47 ppm). Significant positive correlations were found between all metals except Cr and Cu (with r = –0.62). Element pairs such as Cu-Ni, Zn-Cr, Ni-Pb, Ni-Ba, Ni-Co and Cu-Zn showed strongly positive correlation coefficient "r", indicating their association in the study area. The principal component analyses (PCA) yielded 3 components which accounted for 84.73% of the total variance in the data set, indicating that the remaining 16.38% were not explained by these components. The first cluster (Ni, Pb, Co, Ba) accounts for 45.75% of the total variance, the second (Cu, Zn) for 22.01%, the third contains only Cr and suggests both natural and anthropogenic origins. The distribution of heavy metals in the soils had been greatly affected by soil formation, atmospheric deposition and human activities.

Keywords: *multivariate statistics, cluster analysis, principal component analysis, correlation, communalities*

1. INTRODUCTION

The characteristic of soil contaminated by heavy metals is commonly influenced by total heavy metal contents. The study area falls within the Dangme West District of the Greater Accra region in Southeastern Ghana. Although some studies have already extended to the investigation of heavy metal fractions within the Akuse area, it is still far from enough. These studies provided inadequate information about the bioavailability and toxicity of heavy metals. This study focuses on heavy metal sources, which is critical for the monitoring and assessment of soil contamination. There are two main sources of heavy metals in the soil: (i) natural background, which

represents heavy metal concentration derived from parent rocks; (ii) anthropogenic contamination resulting from application of agrochemicals, addition of organic amendments, animal manure, mineral fertilizer, and sewage sludge. Generally, there are more heavy metals in soils originated from anthropogenic sources than natural sources. Due to human settlement, the area of farmland in Akuse has been rapidly increasing. Meanwhile, agro-environmental pollution caused by chemical fertilizers and pesticides, waste discharges such as animal wastes, sewage irrigation, and sludge application are so serious. Therefore, the identification of heavy metal sources in the present work would offer essential information on the monitoring and assessment process of agricultural soils in the Akuse area. The aims of the study were: (1) to determine average local concentrations of some heavy metals (Cu, Zn, Ni, Pb, Cr, Co and Ba); (2) to define their natural and/or anthropogenic sources; (3) to identify their local or exotic sources causing contamination in top soils. A multivariate statistic approach using principal component analysis, cluster analysis and correlation analysis were adopted to assist the interpretation of geochemical data and to distinguish different sources of heavy metals.

2. REGIONAL GEOLOGIC SETTINGS

The study area *(Figure 1)* form part of the southeastern segment of the Trans-Saharan belt exposed in southeastern Ghana and adjoining parts of Togo and Benin known as Dahomeyide orogenic settings [1, 2, 3] precisely the Shai Hills area. The principal tectonic elements of the Dahomeyide orogen are: (1) the deformed edge of the West African Craton with its cover rocks consisting of craton verging nappes and thrust sheets bounded by ductile shear zones; (2) the suture zone representing the eastern boundary of the autochthonous West African Craton; (3) exotic rocks that form the granitoid gneiss complexes east of the suture zone *(Figure 1)*. Metamorphism is generally in the amphibolite facies. Gneisses with garnet, pyroxene and scapolite occur among more ordinary quartzo-feldspathic, biotite and hornblende-bearing varieties. Eclogites (high-pressure garnet-pyroxene rocks chemically equivalent to basalt) have been recorded from among large masses of mafic gneisses that include amphibolites and pyroxenites and contain much garnet. In southeastern Ghana, the High Pressure mafic granulites have been referred to as Shai Hills gneiss [1], and are tectonically juxtaposed with the alkaline gneiss complex in the suture zone. Rocks of the Shai Hills Gneiss unit are folded into west and southwest verging nappes and crop out in inselbergs of the Accra Plains, which include the northsouth trending Shai Hills. Most of the isolated hills are asymmetrical with steep, west-facing scarp slopes such as Krobo Hills and the prominent Osu Yongwa. Several quarries in Krobo, Shai Hills and in the low hills south of Shai Hills provide access to fresh outcrops for detailed sampling. In these quarries, the distinctive features of the rocks exposed are the prominent modal layering and extensive veining. The layering is discontinuous and consists of alternating garnet-rich and hornblende-rich zones, which give the rock a streaky appearance and are interpreted to be shear induced. The veins occur in all sizes and orientated to the tectonic layering in these rocks.

2.1. Geology of study Area

Strongly metamorphosed ancient sediments occur within the study area with Dahomeyan gneiss and schist's occupying most of the plains cover. Basic gneiss forms a number of large inselbergs (isolated rocky hills) in the north and center of the area. Small rock outcrops are also common in the north close to the inselbergs but are rare in south and southeast. The eastern belt of acidic gneiss consists mainly of the grained metamorphosed rocks rather richer in minerals than the rocks in the western part and with many fewer quartz veins. The soil type is largely clayey and fine grain in texture. There are no known mineral deposition of commercial and economic value in the area, except for clays of various types occurring in different places used for pottery and bricks.

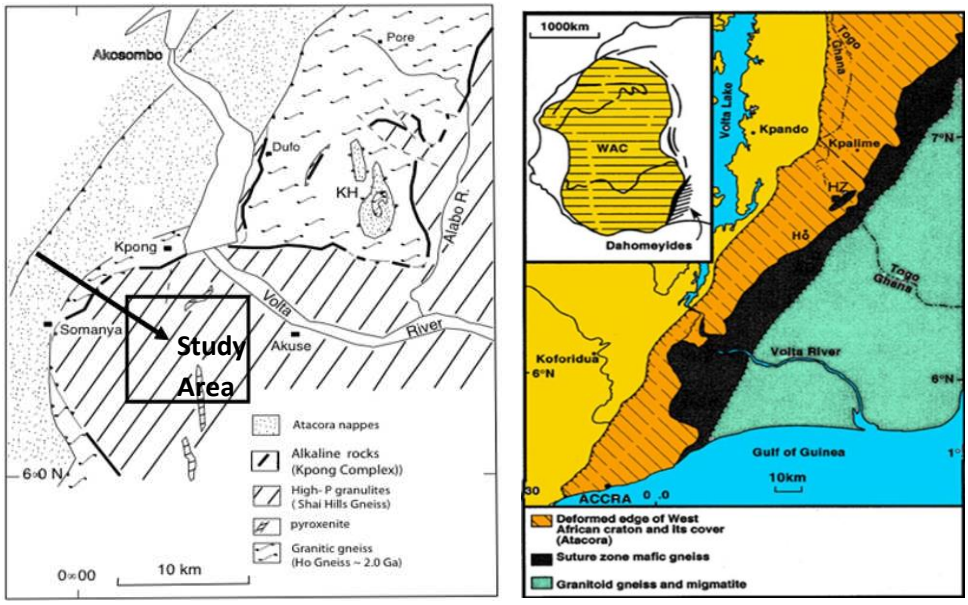

Figure 1. *Location maps showing the study area and the Dahomeyide Tectonic Settings, South-eastern Ghana (Attoh et al., 1997)*

3. MATERIALS AND METHODS

3.1. Sample Collection and analysis

The sampling protocol employed in this work is soil sampling. This was selected based on the time and resources available, the size of the program, the indicators to be measured and data sources to collect information from. A total of 34 samples were taken by six groups. All samples were taken from precise and accurate sample points except E4 which deviated as a result of inaccessible point of location. Using the earth chisel, the samples were taken at a depth range of 50–90 cm.This was intended to

remove the top soil. The samples were dried, crushed and sieved using a 108 and 106 micrometer sieves. They were shaken continuously for three minutes producing finer samples. Five grams (5 g) of each sample were placed in a paper bag, labeled and taken to the laboratory for analyses. Pellets were prepared from the samples and analyzed with X-Ray Fluorescence spectrometer (XRF). This system provided a complete elemental analyses and qualitative characterization. In all, 34 soil samples were analyzed.

3.2. Statistical analysis

Multivariate statistical analysis was performed with the aid of Microsoft excel and IBM statistical software SPSS version 16.0. Pearson's correlation coefficient between the variables were calculated in the forms of matrix and used as a measure of similarity and relationship between the six elements. Cluster analysis has most frequently been employed as a classification tool and has been used by some researchers as a means of representing the structure of data via the construction of dendrogram [4, 5, 6] or overlapping clusters [7]. Cluster analysis was also performed in order to evaluate the metal pollution and distinguish the sources. Between groups average linkage method was tested during the classifying procedure. Factor analysis, using principal component solution was also carried out on the data. During factor extraction, the shared variance of a variable is partitioned from its unique variance and error variance to reveal the underlying factor structure, thus, only shared variance appears in the solution [8]. Principal component analysis does not discriminate between shared and unique variance. When the factors are uncorrelated and communalities are moderate, it can produce inflated values of variance accounted for by the components [9, 10]. Since factor analysis only analyzes shared variance, factor analysis should yield the same solution while also avoiding the inflation of estimates of variance accounted for [8]. Principal component analyses (PCA) is designed to transform original variable into new uncorrelated variables called components. It was used to explain the variance observed in the data, and to understand the relationship between the different elements. Thus, Factor analysis is a multivariate analytical technique, which derives a subset of uncorrelated variables called factors that explain the variance observed in the original dataset [11, 12, 13]. It was used to unearth the latent structure of a set of variables and attempted to identify few factors that were responsible for the correlation among a large number of observed variables. Its application was to describe, if possible, the covariance relationship among many variables in terms of a few underlying, but unobservable, random quantities called factors.

4. RESULTS AND DISCUSSION

4.1. Correlation between total heavy metal contents

Table 1 shows the total metal content in the soil calculated in mean values with their corresponding standard deviations. Chromium (Cr) showed the highest mean concentration (mean = 442.8 ± 68.658 ppm) followed by Barium (Ba) (mean = 261 ±

137.88 ppm). *Table 2* shows the Pearson's correlation coefficient of heavy metals in the samples. The Pearson's correlation coefficient measures the strength of a linear relationship between any two variables on a scale of –1 (perfect inverse relation) through 0 (no relation) to +1 (perfect sympathetic relation). Moderate positive correlations were found between all metals except Cr and Cu (with r = –0.62). There exists poorer correlation between metals such as Cu-Pb, Cu-Ba, Zn-Ba and Cr-Pb showing their low association in the field. This implies that the pollution sources of these metals has no significant effect on their chemical properties and hence are not closely related in the field of survey. Relatively strong positively correlated metals included Cu-Zn, Cr-Ni, Co-Ni, Ba-Ni, Ba-Co and Pb-Ni *(Figure 3)*. Others such as Zn-Ni and Co-Cu showed weaker positive correlations. The above relations are further showed with scattered diagrams below.

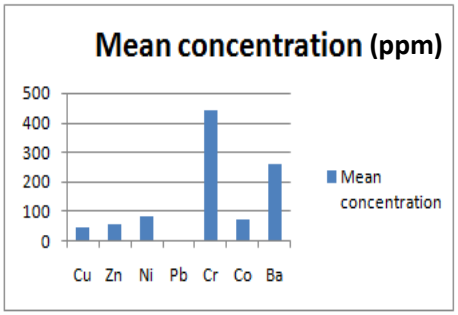

Figure 2. Plot showing mean concentrations of heavy metals

Table 1
Descriptive Statistics (ppm)

	Mean (ppm)	Std. Deviation (ppm)	N
Cu	47.47	18.221	34
Zn	57.26	13.274	34
Ni	83.94	59.915	34
Pb	2.14	1.028	34
Cr	442.85	405.658	34
Co	70.45	38.296	33
Ba	261.13	137.880	34

Table 2
Correlation Matrix

Elements	Cu	Zn	Ni	Pb	Cr	Co	Ba
Cu	1.000						
Zn	0.697	1.000					
Ni	0.188	0.404	1.000				
Pb	0.093	0.106	0.515	1.000			
Cr	-0.062	0.216	0.672	0.011	1.000		
Co	0.435	0.327	0.646	0.504	0.196	1.000	
Ba	0.049	0.057	0.664	0.510	0.182	0.712	1.000

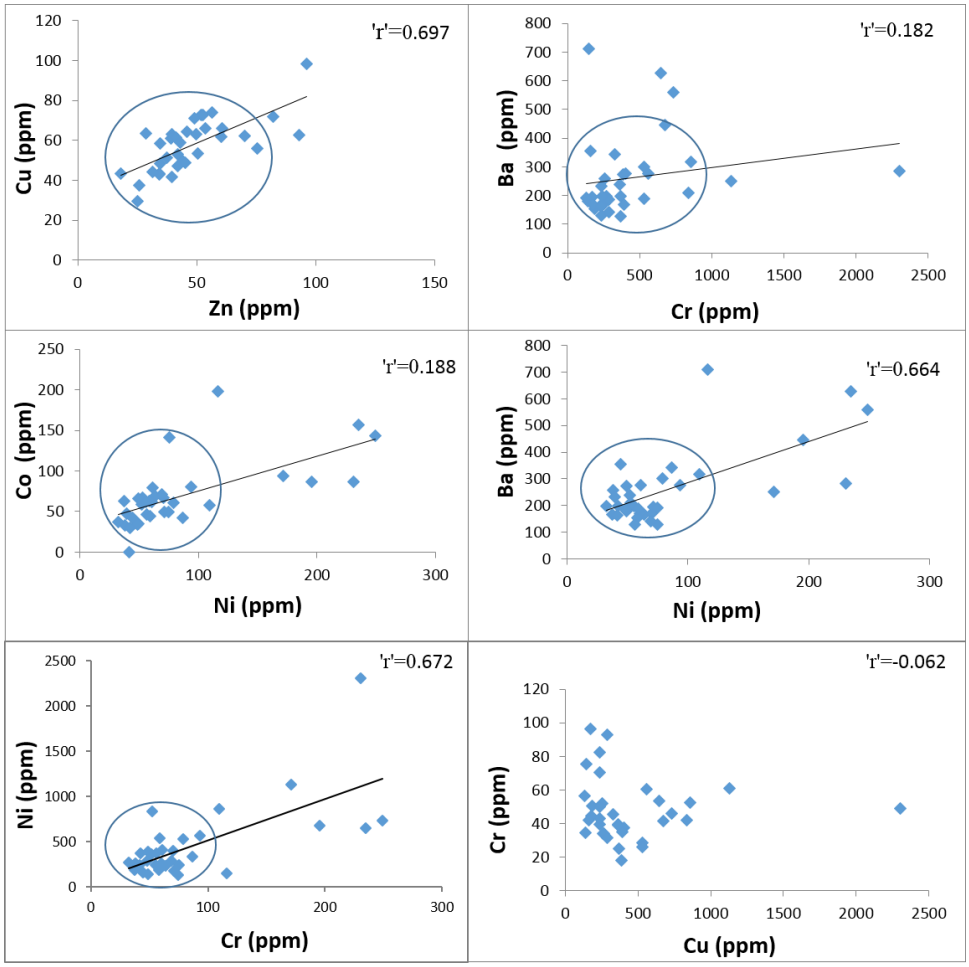

Figure 3. *Correlation plots of heavy metal in the study area*

4.2. Principal Component Analysis

Table 3 shows the Total variance explained by principal components in cumulative percentages whilst *Table 4* shows the principal component coefficient for heavy metals in the Akuse area extracted in three components. The PCA yielded 3 components which explained 84.73% of the total variance in the data set, indicating the remaining 16.38% were not explained by these axes. The first component accounts for 45.75% of the total variance and contains Ni, Pb, Co and Ba. Component 1 showed significant positive load for all the heavy metals. Ni showed the highest load in this factor (load value 0.901). This was followed by Co, Ba, Pb, Zn, Cr and Cu respectively. Although Zn was not classified in this component, it showed significant loading in the component. The metals Ni, Pb, Co and Ba were also significantly correlated *(Table 2)*, indicating that they are from the same source.

Table 3
Total variance explained by principal components

Com-ponent	Initial Eigen values			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.203	45.754	45.754	3.203	45.754	45.754
2	1.541	22.010	67.764	1.541	22.010	67.764
3	1.187	16.963	84.726	1.187	16.963	84.726
4	0.544	7.775	92.501			
5	0.283	4.041	96.542			
6	0.165	2.361	98.903			
7	0.077	1.097	100.000			

Table 4
Principal component coefficients

	Component		
	1	2	3
Cu	0.444	0.817	-0.188
Zn	0.528	0.740	0.176
Ni	0.901	-0.184	0.308
Pb	0.629	-0.294	-0.434
Cr	0.456	-0.176	0.846
Co	0.859	-0.006	-0.256
Ba	0.757	-0.419	-0.238

This factor suggested an anthropogenic source since Ba, Ni and Co showed a higher mean *(Table 1)* value exceeding permissible background values. The second component accounts for 22.01% of the variance and contains Cu and Zn. The close association of Cu and Zn shows that they are from possibly naturally related source. The two heavy metals were also highly correlated with r = 0.697 *(Table 2)* showing their close association in the area. The remaining metal exhibited significant negative loads in this component showing their disassociation with Cu and Zn in the study area. This accession is further explained by lower correlation values showed by Cu and Zn with the other heavy metals *(Table 2)*. The third component accounts for 16.963% of the variance and contains only Cr (load value 0.846). This factor suggests a source of

mixed origins of natural and anthropogenic as chromium (Cr) showed the highest mean value *(Table 1).*

4.3 Cluster Analysis

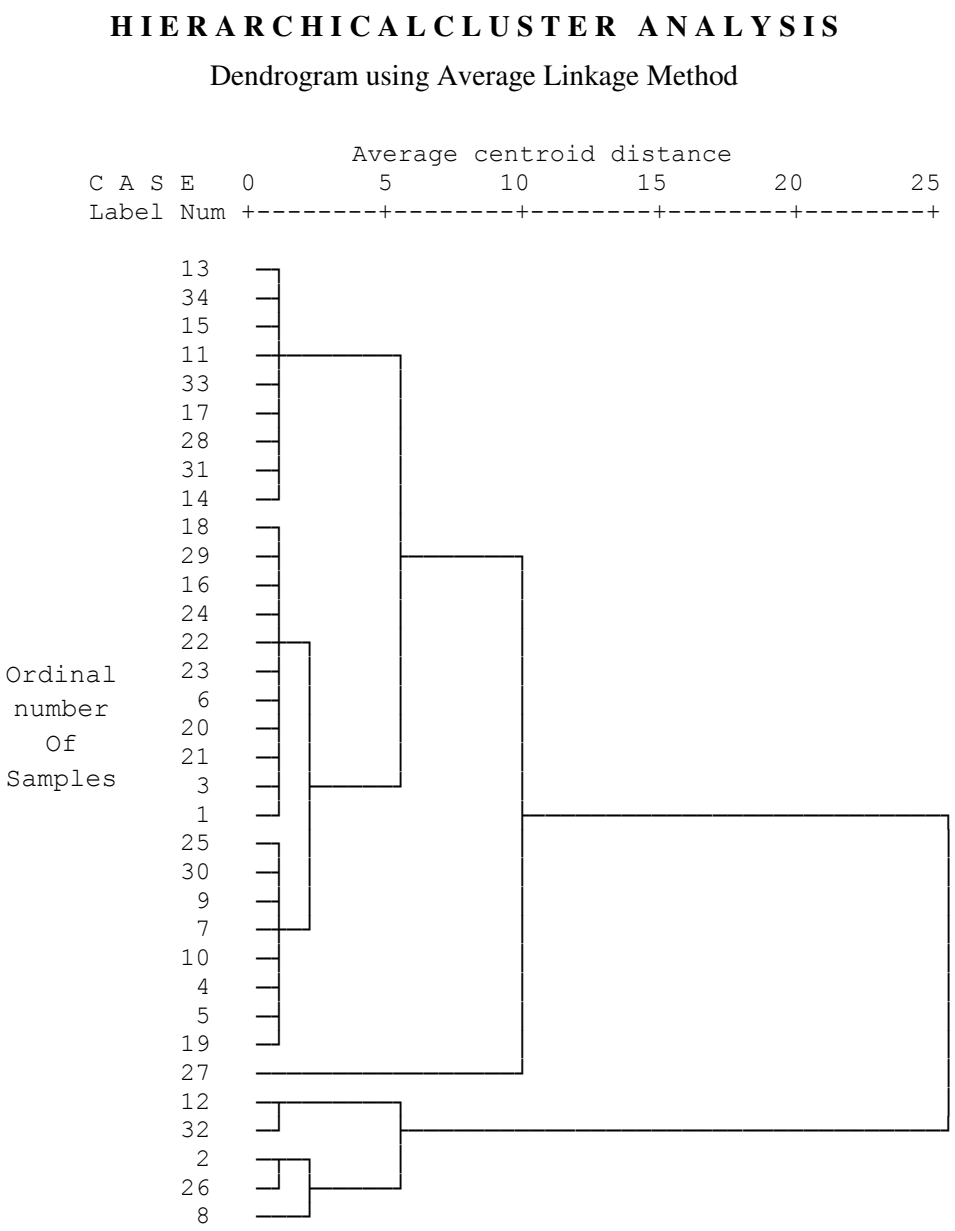

Figure 4.

Connections and relations for interdependence among the different metals in the Akuse area were also established by means of cluster analysis. Hierarchical cluster analysis created a hierarchy of clusters which were represented in a tree structure called dendrogram. The dendrogram of the hierarchical cluster analysis of metal contents is illustrated below *(Figure 4)*. Dissimilarities between clusters of objects are defined in several ways including the maximum dissimilarity (complete linkage), minimum dissimilarity (single linkage) and average dissimilarity (average linkage) [14]. Group-average linkage (agglomerative) clustering as applied in this study evaluates cluster quality based on all similarities between data including pairs from the same cluster but self-similarities are not included in the average. Group average linkage satisfies the reducibility principle which ensures that the procedure cannot result in inversion in the dendrogram representing the hierarchy [15]. It is also known to minimize the distortion imposed on the inter-object similarity matrix when a hierarchic classification is generated [16].

The results revealed four clusters of elements, the first cluster included elements that had previously been interpreted as anthropogenic (Ni, Pb, Co and Ba) while the second clusters contained the natural source elements (Cu and Zn). Cr concentrations were represented by cluster 3. Cluster 1 ranges from 13-14 on the dendrogram, Cluster 2 ranges from 1-18, Cluster 3 ranges from 19–25 with the rest scattered in cluster 4. These group distributions also corresponded to the locations of samples. Cluster 4 contained local abnormal points and could be neglected in analysis. Based on earlier discussions, the results suggested again that Ni, Pb, Co and Ba were mainly controlled by anthropogenic sources, while, Cu and Zn had natural sources. Cluster analysis gave similar results to correlation and factor analyses, enabling the identification of the two sources of heavy metals. Therefore, this analysis re-affirmed that the elements studied come from two different sources in the soils except Cr which has a mixed source.

5. CONCLUSION

Multivariate statistical methods applied in this study proved useful in the characterization of heavy metal sources in soils from Akuse area. Relatively homogenous groups of metals were identified, and varied sources or origins of metals could be distinguished based on their elemental association, level of accumulation and chemical properties. Similar geochemical properties such as ionic size and radius accounted for the positive correlation observed among heavy metals Ni, Cr, Co, Zn and Cu which belong to the transition trace elements group. Among the factors affecting the distribution of heavy metals in the soil are the abundance of the element in the parent rock, the nature of the weathering in the study area and human activities. Chromium and barium showed excessively high concentrations in the study area. The sequences of geo-accumulation of heavy metal in a decreasing order are Cr, Ba, Ni, Co, Zn, Cu and Pb.

The study showed that: (i) Ni, Pb, Co and Ba had anthropogenic sources (probably from over use of chemical fertilizers and pesticides, industrial and discharges, animal wastes, sewage irrigation, etc.), (ii) Zn and Cu were associated with parent materials and therefore had natural sources, thus, the weathering of parent materials and subsequent pedogenesis due to the alluvial deposits. These elements are largely associated with high pressure acidic gneisses complexes scattered across the north and western parts of the study area. (iii) Cr showed varied sources of both natural and anthropogenic. The effect of heavy metals in the soils was greatly affected by soil formation, atmospheric deposition and human activities. These findings provided essential information on the possible sources of heavy metals without assessing their extent of pollution in the study area but would still go a long way to contribute to the monitoring and assessment process of agricultural soils in the area.

ACKNOWLEDGEMENT

I want to thank the entire lecturers and staff of the Earth Science Department (University of Ghana), who helped in the provision of geological equipment and assistance during the field work which was exceedingly courteous and helpful. Also, my sincere gratitude to the entire staff of the Geological Survey Department (Accra branch) for their help during the preparation and analysis of samples.

I am greatly indebted to Dr. Frank K. Nyame (University of Ghana) and Dr. Boamah (Geological Survey Department) for their technical advice. I would like to express my sincere gratitude to the academic staff, Department of Geophysics (University of Miskolc) for their assistance and encouragement to make this paper a success.

REFERENCE

[1] ATTOH, K.–DALLMEYER, R.D.–AFFATON P. (1997)
Chronological nappe assemblage in the Pan African Dahomeyide orogeny, West Africa. Evidence from 40Ar/39Ar mineral ages. *Precambrian Research*, Vol. 82, pp 82.

[2] AFFATON, P.–RAHMAN, M. A.–TROMPETTE, R.–SOUGY, J., (1991)
The Dahomeyide orogeny; Tectonothermal Evolution and Relationships with the Volta Basin. Springer Berlin Heidelberg.

[3] CASTAING, C.–FEYBESSE, J.L–THIEBLEMONT, D.–TRIBOULET C.–CHEVREMONT, P. (1994)
Palaeogeographical Reconstructions of the Pan, African/brasilianoorogen: Closure of an Oceanic Domain or Intracontinental Convergence between Major blocks? *Precambrian Research*, 67, 327–344.

[4] BERTIN, J. (1967)
Sémiologie Graphique. Paris: Gauthier-Villars.

[5] Hartigan, J. A. (1967),
Representation of Similarity Matrices by Trees, *Journal of the American Statistical Association*, 62, 1140–58.

[6] SZABÓ, N. P.–DOBRÓKA, M.–KAVANDA, R. (2013)
Cluster Analysis Assisted Float-Encoded Genetic Algorithm for a More Automated Characterization of Hydrocarbon Reservoirs. *Intelligent Control and Automation*, 2013, 4, 362–370.

[7] SHEPARD R. N.–P. ARABIE (1979)
Additive Clustering: Representation of Similarities as Combinations of Discrete Over-Lapping Properties, *Psychological Review*, 86, 87–123.

[8] COSTELLO, A. B.–OSBORNE, J. W. (2009)
Best practices in exploratory factor analysis; four recommendations for getting the most from your analysis. *The Pan-African Management Review*, 2009, Vol. 12, No. 2, 131–146.

[9] GORSUCH, R. L. (1997)
Exploratory factor analysis: its role in item analysis. Journal of Personality Assessment, 68:532-560.

[10] McArdle, J.J (1990). Principles versus principals of structural factor analyses. *Multivariate Behavioral Research*, 25, 81–87.

[11] ANAZAWA–OHMORI, (2005)
The hydrochemistry of surface waters in andesitic volcanic area, Norikura volcano, central Japan. *Chemosphere*, Volume 59, Issue 5, April 2005, pp. 605–615.

[12] BROWN, (1998)
Applied Multivariate Statistics in Geohydrology and Related Sciences Springer. New York, (1998)

[13] YIDANA, S. M–OPHORI, D.–BANOENG-YAKUBO, B. (2008)
A multivariate statistical analysis of surface water chemistry – the Ankobra Basin, Ghana. *Environ. Manag*. 86, 80–87.

[14] CHRISTOPHER D. MANNING–PRABHAKAR RAGHAVAN–HINRICH SCHÜTZE *Introduction to Information Retrieval*. Cambridge University Press, 2008.

[15] F. MURTAGH (1983)
A survey of recent advances in hierarchical clustering algorithms. *The Computer Journal*, 26, 354–35

[16] A. EL HAMDOUCHI–P. WILLET (1989)
Comparison of hierarchic agglomerate clustering methods for document retrieval. *The Computer Journal*, Vol. 32, No. 3.

Geosciences and Engineering, Vol. 5, No. 8 (2016), pp.135–146.

DEVELOPMENT OF NEW COMPLEX SOFTWARE FOR INVESTIGATING ACOUSTIC VELOCITIES UNDER PRESSURE

JUDIT SOMOGYINÉ MOLNÁR

Institute of Electrical & Electronic Engineering, University of Miskolc
MTA-ME Geoengineering Research Group
gfmj@uni-miskolc.hu

Abstract: Studying the pressure dependence of acoustic wave velocities in the laboratory provides important information to interpret seismic measurements in relation to petrophysical parameters. For this purpose, accurate measurement data are required, especially P and S wave velocity. In this study new software was developed in LabVIEW to measure longitudinal and transverse wave velocity under pressure on rock samples. A geophysical inversion method was used for data processing in which quantitative petrophysical model equations explaining the pressure dependence of wave velocities served as response equations. The inversion algorithm was realized in MATLAB, which was implemented in the LabVIEW measuring and data acquisition program, resulting in a complex software.

Keywords: *acoustic measurements, LabVIEW software, seismic velocity, pressure*

1. INTRODUCTION

Extrapolation of seismic/acoustic properties measured on rock samples in laboratory at a given pressure state plays a significant role in the interpretation of seismic data. Based on the velocities, estimates of pressure dependent porosity – a key characteristic in well logging – can be derived [1], [2]. It has been observed that pressure has a greater influence on velocities in the beginning phase of loading and later lessens, and that the velocities tend to a limit value. The most frequently used mechanisms for explaining this process are based on the closure of microcracks [3] or pores [4] in rocks under pressure. For an analytical description of the nonlinear velocity vs. pressure relationship, exponential functions are most commonly used [5], [6].

Rock properties can be described by material characteristic parameters. The dynamic elastic moduli, e.g. compression *(K)*, shear *(G)*, Young's modulus *(E)*, or Lamé coefficients *(μ* and *λ)* are important and often specified quantities and can be expressed by the acoustic propagation velocities [7]

$$K = \rho\left(\alpha^2 - \frac{4}{3}\beta^2\right), \tag{1}$$

$$G = \mu = \beta^2\rho, \tag{2}$$

$$\lambda = \alpha^2 \rho - 2\mu \,, \tag{3}$$

$$E = \beta^2 \rho \frac{3\alpha^2 - 4\beta^2}{\alpha^2 - \beta^2} \,. \tag{4}$$

The pressure dependence of the density of a rock is assumed negligible in comparison of velocities (i.e. the density ρ is considered to be a constant value). It can be seen that for the description of the pressure dependence of the elastic parameters, pressure-dependent velocity data measured in laboratory are required. In this paper new software is presented that allows us to measure velocities accurately, which is important for the precise deduction of the pressure dependent elastic parameters.

2. EQUIPMENT FOR LOADING THE SAMPLES

Measurements on different cylindrical sandstone samples under uniaxial stresses were performed. The digitally controlled test system *(Figure 1)* includes a pressure cell, a load frame (max. 300 kN) and a pressure generator (max. 80 MPa).

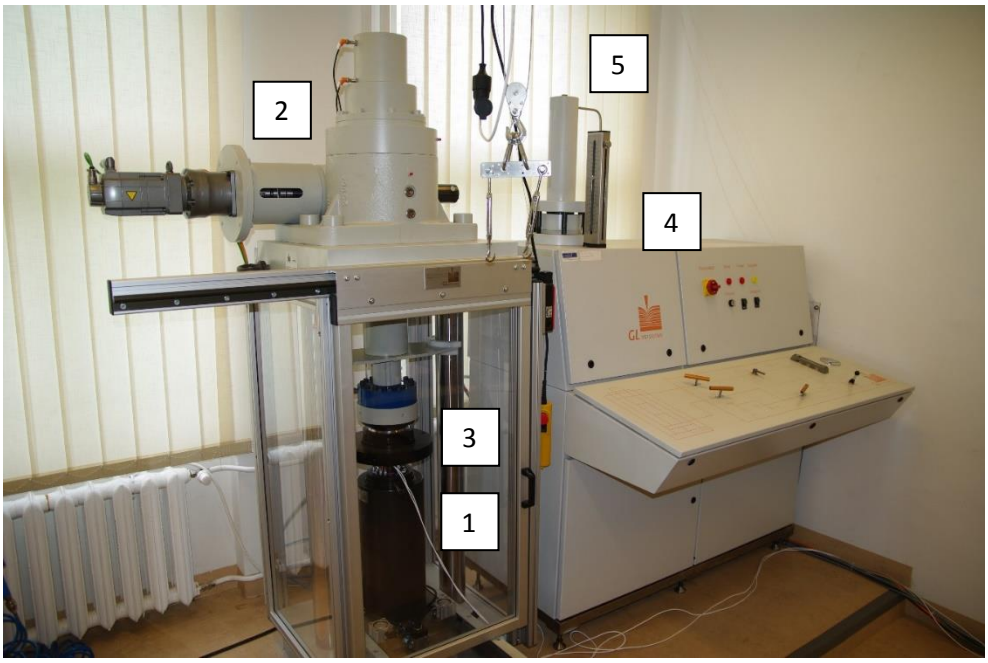

Figure 1
Equipment: (1) pressure cell; (2) load frame; (3) pressure plate; (4) pressure generator; (5) piston

The pressure cell enables triaxial loading ($\sigma_1 \neq \sigma_2 = \sigma_3$) of 35 mm diameter cylindrical, maximum 100 mm height specimens. The pressure stamps include piezoelectric crystals for acoustic measurements. Transducers in the stamps have 1 MHz eigenfrequency and are sealed against confining pressure.

3. LABVIEW SOFTWARE TO MEASURE P/S WAVE VELOCITIES UNDER PRESSURE

Acoustic longitudinal (P) and transverse (S) wave velocities are measured in the laboratory, primarily by using the pulse transmission technique [8]. The arrival times of the pulse can be detected by the software, i.e. if the length of the sample is known, it is possible to measure P and S wave velocities. This is the so-called pulse transmission technique. Note that to calculate the correct velocity one has to measure first the arrival time of acoustic wave propagates only in the pressure stamps. By pushing the stamps together one can measure the time belonging to the first arrival, the so-called "t deads" (P: 17120 ns, S: 31380 ns).

The new measuring & data acquisition program was designed in LabVIEW. For the development of the software a National Instruments PCI-6251 card connected to an NI BNC-2110 adapter as well as LabVIEW installed on a PC. The adapter has 15 BNC connectors and a 30-pin terminal block. Since the pressure cell has an ED4-200 RT-type electrical lead-through and I would like to connect the analog signal to the adapter by BNC connector, an FFA.0S LEMO-BNC shielded twisted pair cable was used for signal transmission. A Tektronix AM502 differential amplifier was used to increase the amplitude of the signal.

The measuring equipment provides the possibility to measure the propagation times of both P and S waves. The software should be suitable for measuring, visualizing (real-time) ultrasonic waves, supporting data export in .txt, and also evaluating the measured data by geophysical inversion method. Since it can happen that the parameters (such as the channels to be measured) should be changed while the program is running, but elementary functions were applied when writing the program, rather than the built-in DAQ Assistant. In the task framework for the core of the program the following parameters were defined: channels, timing, triggering and reading *(Figure 2)*.

It can be seen that analog voltage signal measurement was set at the created channel. The specific physical channel can be selected via a drop-down menu on the front panel (e.g. AI0-channel). The measurement range is ±5 V and at the Input terminal configuration common ground point measurement was set (referenced single ended – RSE). In RSE mode one can measure the signal relative to the ground point. The measurement starts by clicking on the button called "Mérés indítása". The timing of the detected samples from the physical channel can be adjusted by the AI Sample Clock, for which a 1 MHz sampling rate and finite sampling (8192 sample) were set. The AI Start Trigger Signal function block (hereinafter FB) ensures the start of sampling. As can be seen in *Figure 2*, rising edge trigger mode was chosen

with a 2 V trigger level, i.e. the sampling starts if the detected analog signal exceeds 2 V.

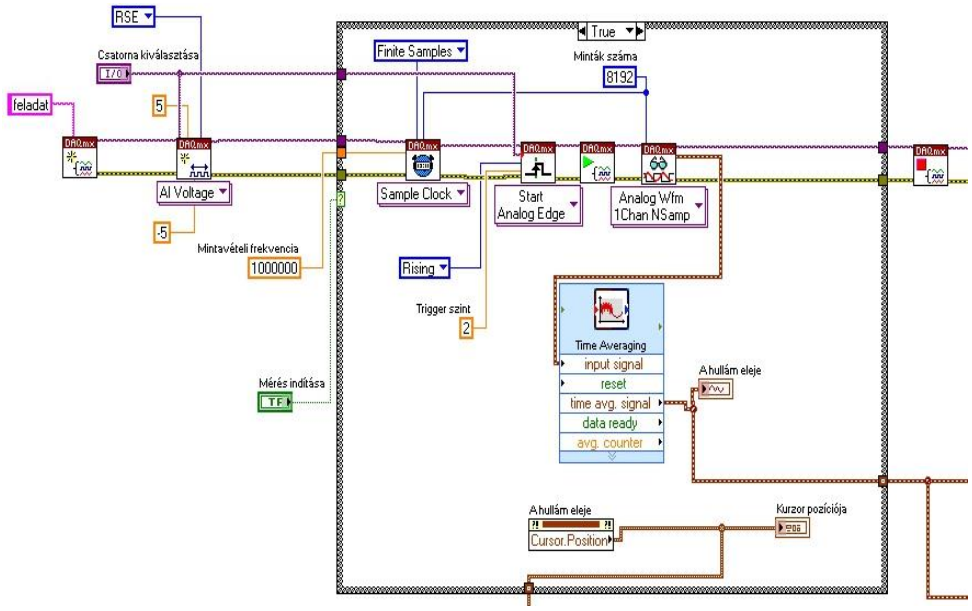

Figure 2
Core of the LabVIEW software

The DAQmx Read FB is responsible for reading the detected voltage values (a total of 8192). The values were read out in waveform format, because then the time axis data are automatically obtained from the timing data of sampling, so the wave can be visualized immediately. A 256-fold summary was applied in the software (Time Averaging FB), which means a running average of 256 consecutive measurements. With this method the amplitudes of external noises in the displayed wave were significantly reduced; consequently, the first arrivals can be determined accurately. After stacking the signal it can be easily displayed in a waveform graph (lower graph in *Figure 3*). Since the first arrival is expected in the first/second quarter of the signal (of course this depends on the length and material of the sample) a second graph was created (the upper graph in *Figure 3*).

The actual value of the signal can be obtained by adding a cursor to the upper graph, which is represented by a yellow cross in *Figure 3*. The X–Y coordinates belonging to the easily picked first arrival can be seen in the indicator called “Kurzor pozíciója”. Since the propagation velocity is calculated from the arrival time and the sample length, the operator can specify the length of the sample in the front panel (the actual velocity is displayed in the upper right corner in *Figure 3*).

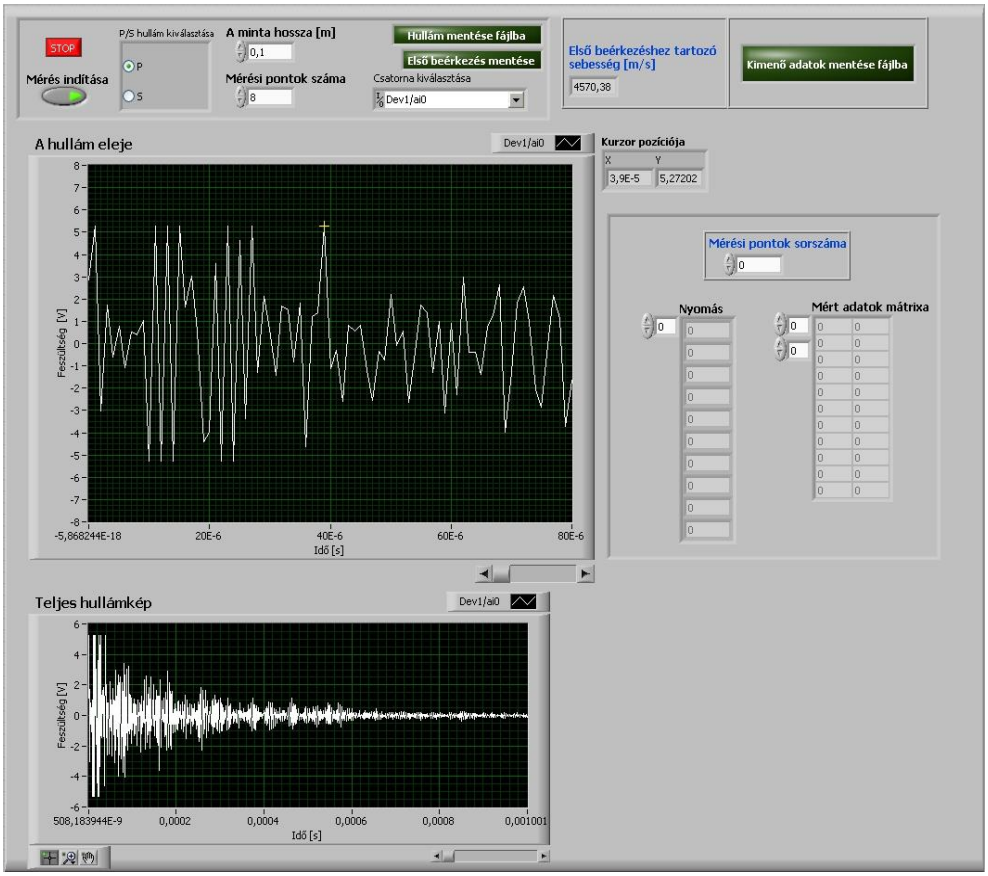

Figure 3
The user interface of the developed LabVIEW software

The so-called "t deads" should be taken into correction at velocity calculation. Since "t deads" are different for the P and S waves, with the help of a radio button one can select what kind of wave is to be measured. The block diagram belonging to this program section can be seen in *Figure 4*.

The voltage-time values related to the picked first arrival (activated when the button called "Első beérkezés mentése" is pressed) as well as the entire signal (the button called "Hullám mentés fájlba" activates Write to speadsheet file.vi) are enabled to save due to the possible subsequent data processing (e.g. the modification of the first arrival). These two saving methods can be seen in *Figure 4*. Since a .txt file can be imported easily into the often used Reflex software for seismic data processing, it was advisable to save these data in .txt format. It is important that in the pop-up window the operator should enter the file extension (.txt) beside the name of the file, as indicated in the documentation of the button *(Figure 4)*.

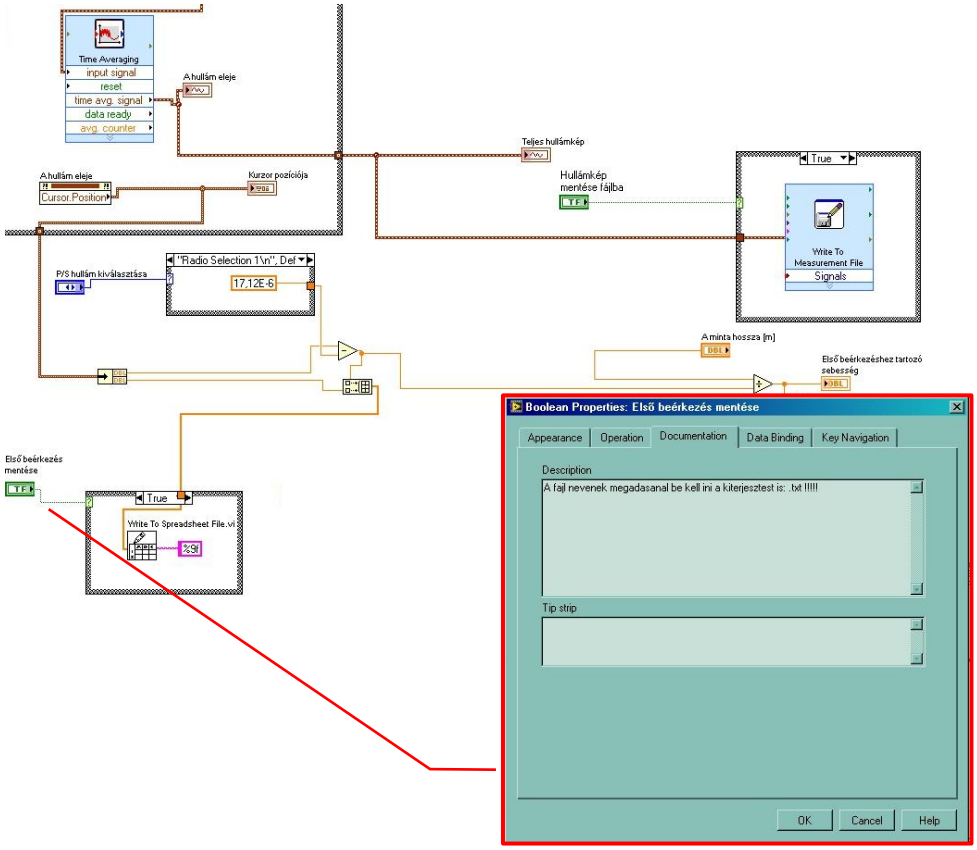

Figure 4
Program section of wave selection and saving

The next step is to load the velocities belong to the first arrivals at a given pressure into a matrix, i.e. the data acquisition. We can see in *Figure 5* that a vector called "Mért adatok vektora" was created by the Initialize array FB. Its number of rows is determined by the control "Mérési pontok száma" (on how many pressure steps we want to measure). This vector is filled with the velocity data belong to the first arrival on each pressure step by the Insert into array FB. The location of each velocity value (row index) in the vector is indicated by the control called "Mérési pontok sorszáma". As we increase this number the velocity value is inserted into the proper row *(Figure 5)*.

A case structure ensures that the program section for number shifting is executed repeatedly (its input is a condition inquiry). In order to avoid the overwriting of each line with the actual velocity value, a *for* cycle was necessary (N in *Figure 5* is equal to the value of the control called "Mérési pontok sorszáma"). In the cycle the Index array FB removes the velocity data due to the running index from the indicator called

"Mért adatok vektora". Therefore the output of the *for* cycle will be a new vector that already contains all the measured velocities *(Figure 5)*.

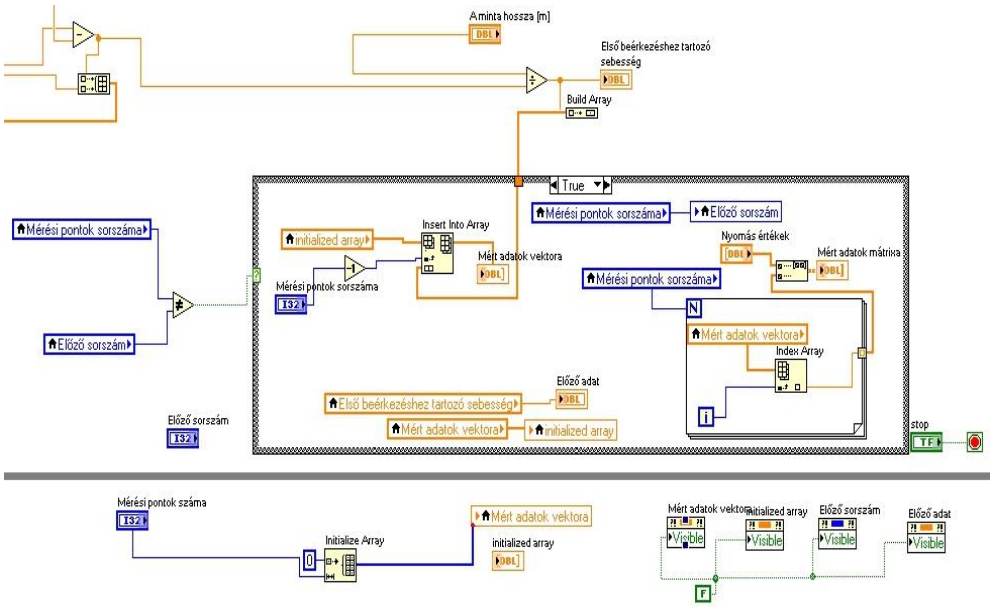

Figure 5
Program section for data acquisition

We can see in *Figure 3* that the operator can input the pressure data manually on the user interface (i.e. the column vector called "Nyomás értékek" will be filled). Thus a matrix called "Mért adatok mátrixa" is created whose first column contains these pressure data and the second the velocity values. Obviously the measurement stops when pressing the Stop button (the program exits from the while loop). Saving the measured pressure-velocity data into a matrix in LabVIEW (by the button called "Kimenő adatok mentése fájlba") enables immediate data processing. In the next section, the data processing by inversion method will be described.

4. DATA PROCESSING BY INVERSION METHOD

In [7] petrophysical models were developed which provide the relationship between P and S wave velocities as well as rock pressure, thus creating the possibility to determine the velocity values at any pressure by the model equations (α: P wave, β: S wave)

$$\alpha = \alpha_0 + \Delta\alpha_0 \, [\, 1\text{-}exp(-\lambda_V \sigma) \,] \,, \qquad \beta = \beta_0 + \Delta\beta_0 \, [\, 1\text{-}exp(-\lambda_V \sigma) \,] \,. \tag{5}$$

These equations provide a theoretical connection between the P and S wave velocities and rock pressure in case of loading. In these equations λ_V is the logarithmic stress sensitivity of the propagation velocities [9], α_0 and β_0 are the propagation velocities at a stress-free state, which can be measured in the laboratory, $\Delta\alpha_0$ and $\Delta\beta_0$ can be considered the velocity drops (relative to the fully compacted state where the pore volume equals zero) caused by the presence of pores at zero pressure [10]. During development BIRCH's qualitative consideration [4] was followed, stating that the main factor determining the pressure dependence of propagation velocity is the closure of pores, i.e. decreasing of pore volume (parameter *V* as the unit pore volume of a rock). Due to increasing pressure from the unloaded state, first the large pores close in the rock sample, and after the slower compression process of smaller pores, nearly all pores are closed. Since the model is based on the pore volume, or rather the change in pore volume, it is suitable to describe the pressure dependence of longitudinal and transverse wave velocities, respectively.

In the terminology of geophysical inversion, *Equation (5)* represent the solution of the direct problem. Parameters α_0, $\Delta\alpha_0$, β_0, $\Delta\beta_0$, λ_V appearing in these equations are not known, but can be determined indirectly by inversion processing of the measurement data. In the inversion program written in MATLAB the least squares method was used [11]. P/S wave velocity data belong to the first arrivals at each pressure measured by the new software provided the input data. The output data of the data processing were the estimated model parameters and the quantities characterizing the goodness of the inversion procedure. Due to the faster data interpretation the measured and calculated data were also displayed graphically.

5. TESTING THE NEW SOFTWARE

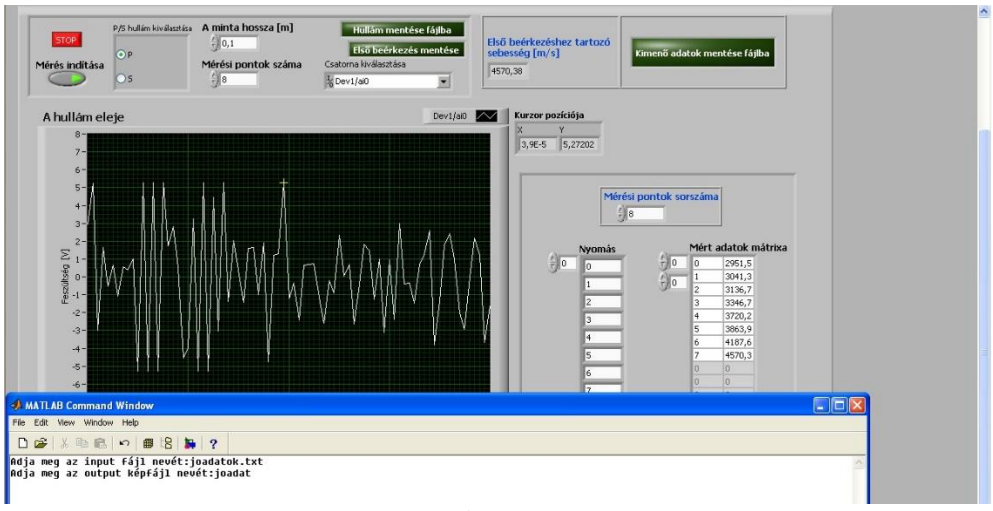

Figure 6
MATLAB under LabVIEW

In this paper one test result is presented. Longitudinal wave velocity was measured on a sandstone sample. The uniaxial loading of the sample was carried out with the measurement equipment detailed in Section 3 (during measurement ramp function was used at 0.05 kN/s velocity). To avoid the failure of the samples it was loaded only up to 1/3 of the uniaxial strength. A 256-fold stacking was applied to increase the signal/noise ratio. During testing the longitudinal wave velocity data measured on the sample together with pressure values were saved in a .txt file.

For data processing the inversion algorithm was executed in the MATLAB Command Window, which opens automatically after LabVIEW starts. This can be seen in Figure 6. This function is only available under the Windows operation system and works in MATLAB version 6.5 and above.

The inversion results (calculated model parameters with estimation errors, RMS and mean spread) can be seen in *Figure 7*.

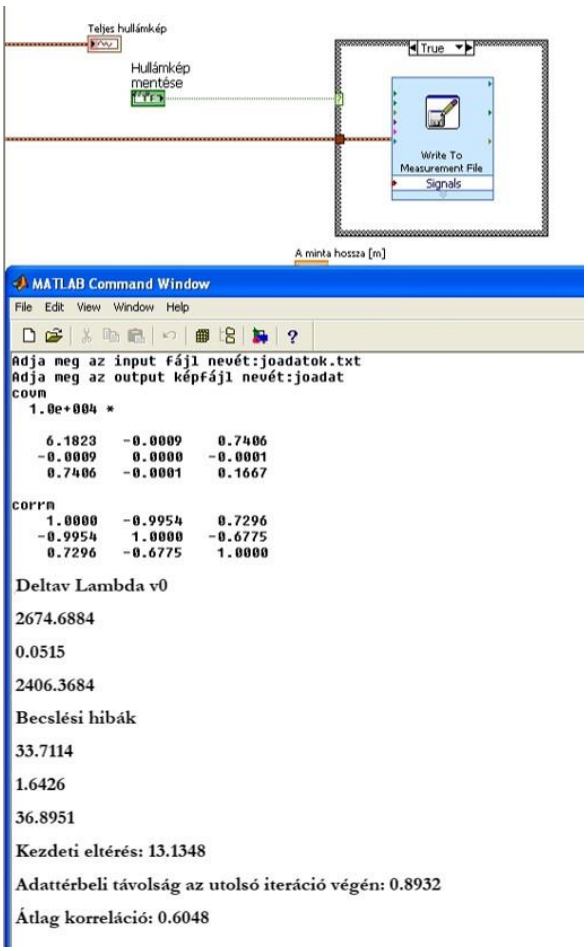

Figure 7
Inversion results

Using the model *Equations (5)* the velocity values at any pressure can be calculated, which can be seen in *Figure 8*. Dots mean the measured values while the solid lines represent the calculated values (one can see that the calculated graph is exponential, which is in conformity with the literature). It can be seen that in the lower pressure range, the increase in velocity with increasing pressure is steep and nonlinear. This is due to the closure of pore volume, which significantly affects the elastic properties of rock and thereby the velocities. In the higher pressure range, the increase in velocity (with increasing pressure) becomes moderate as the closeable pore volume lessens.

Figure 8 shows that the calculated curves are in good accordance with the measured data, proving that the petrophysical model applies well in practice and that the new software as well as the MATLAB algorithm under LabVIEW worked properly.

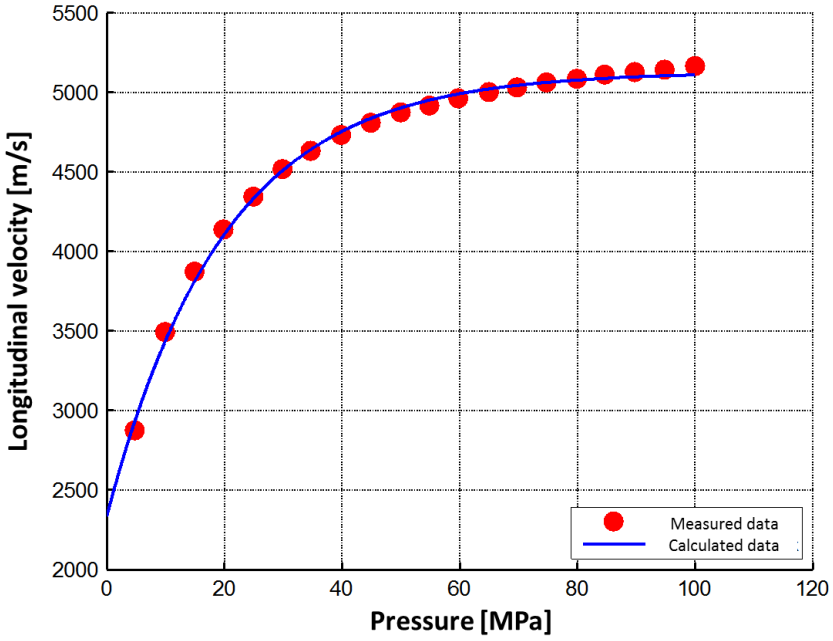

Figure 8
Measured and calculated P wave velocity vs. pressure

6. CONCLUSIONS

In this paper new LabVIEW software was presented to measure longitudinal and transverse wave velocity on rock samples under uniaxial pressure. Measured data were processed in a geophysical inversion procedure (MATLAB algorithm for least squares method) based on the developed rock physical model, which describes the

pressure dependence of acoustic velocities. The decreasing pore volume [4] explanation for the pressure-velocity relationship was used in development of the petrophysical model. To confirm the reliability of the acoustic software, laboratory measurements were made on cylindrical rock samples and data were processed with the inversion software. One example of successful application was presented in the paper. As a result of the developed software, it is possible to use one computer instead of three for a complex measurement.

ACKNOWLEDGEMENTS

The research of Judit Somogyiné Molnár was carried out in the framework of the Terplán Zénó Program.

LIST OF SYMBOLS

Symbol	Description	Unit
E	Young modulus	N/m^2
G	shear modulus	N/m^2
K	compression modulus	N/m^2
V	unit pore volume of a rock	m^3
α	longitudinal (P) wave velocity	m/s
α_0	longitudinal (P) wave velocity at stress-free state	m/s
β	transverse (S) wave velocity	m/s
β_0	transverse (S) wave velocity at stress-free state	m/s
$\Delta\alpha_0$	longitudinal (P) wave velocity-drop at zero pressure	m/s
$\Delta\beta_0$	transverse (S) wave velocity-drop at zero pressure	m/s
λ	first Lamé coefficient	N/m^2
λ_V	new material quality dependent petrophysical parameter	1/MPa
μ	second Lamé coefficient	N/m^2
ρ	density	kg/m^3
σ	stress	MPa

REFERENCES

[1] SZABÓ, N. P.: Shale volume estimation based on the factor analysis of well-logging data. *Acta Geophysica*,59:(5), pp. 935–953.

[2] SZABÓ, N. P.: Hydraulic conductivity explored by factor analysis of borehole geophysical data. *Hydrogeology Journal*, 2015, 23:(5), pp. 869–882.

[3] WALSH, J. B., BRACE, W. F.: A fracture criterion for brittle anisotropic rock. *Journal of Geophysical Research*, 1964, 69, 3449–3456.

[4] BIRCH, F.: The velocity of compression waves in rocks to 10 kilobars, Part 1. *Journal of Geophysics Research*, 1960, 65, 1083–1102.

[5] WEPFER, W. W.–CHRISTENSEN, N. I.: A seismic velocity-confining pressure relation, with applications. *International Journal of Rock Mechanics and Mining Science*, 1991, 28, 451–456.

[6] WANG, Q.–JI, S. C.–SALISBURY, M. H.–XIA, M. B.–PAN, B.–XU, Z. Q.: Shear wave properties and Poisson's ratios of ultrahigh-pressure metamorphic rocks from the Dabie-Sulu orogenic belt: Implications for the crustal composition. *Journal of Geophysical Research,* 2005, 110, B08208.

[7] KISS A.–SOMOGYI MOLNÁR J.: Rugalmassági paraméterek nyomásfüggésének vizsgálata kőszén mintákon. *Bányászati és Kohászati Lapok*, 2016. 149(1), 16–21.

[8] TOKSÖZ, M. N.–JOHNSTON, D. H.–TIMUR, A.: Attenuation of seismic waves in dry and saturated rocks, I. Laboratory measurements. *Geophysics*, 1979, 44(4), 681–690.

[9] SOMOGYI MOLNÁR, J.–KISS, A.–DOBRÓKA, M.: Petrophysical models to describe the pressure dependence of acoustic wave propagation characteristics. *Acta Geodaetica et Geophysica*, 2014, 50(3), 339–352.

[10] JI, S.–WANG, Q.–MARCOTTE, D.–SALISBURY, M. H.–XU, Z.: P wave velocities, anisotropy and hysteresis in ultrahigh-pressure metamorphic rocks as a function of confining pressure. *Journal of Geophysical Research*, 2007, 112, B09204.

[11] MENKE, W.: *Geophysical Data Analysis: Discrete Inverse Theory*. Academic Press, Inc., London., 1984.

Geosciences and Engineering, Vol. 5, No. 8 (2016), pp.147–164.

DATA PROCESSING OF QDAEDALUS MEASUREMENTS

GYULA TÓTH–LAJOS VÖLGYESI

Department of Geodesy and Surveying, Budapest University of Technology and Economics
Budapest, Hungary
toth.gyula@epito.bme.hu, volgyesi@eik.bme.hu

Abstract: QDaedalus is an automated, computer-controlled astrogeodetic measurement system that has been developed especially for the determination of the local vertical deflections. We measured angles by the total station and recorded position of stars on the CCD. Then, by joint inversion of measurements we obtained the direction of the local zenith. We found it is possible to get zenith direction accurate to 0.1 seconds of arc based on 20–30 minutes of measurements. Two iterative robust inversion procedures have been tested: (1) the so-called least-squares Danish method and (2) most frequent value procedure with Cauchy-Steiner weights. Our evaluation has shown that Method 2 achieves better performance based on more than 70 series of measurements at the same station.

Keywords: *Astrogeodetic measurements, deflection of the vertical, inversion, Danish method, Cauchy-Steiner weights*

1. INTRODUCTION

In geodesy and in civil engineering it is important to know the gravity field. For example, physical heights related to equipotential surfaces of gravity are required for certain engineering projects. Intrinsically linked to the gravity field are the local horizon and its normal vector, the local zenith direction (direction of the gravity vector or of the plumb line). Zenith direction is usually referenced to an ellipsoidal normal vector of a geodetic reference system. Deviations of the zenith direction from the ellipsoidal normal vector in North-South and East-West directions are the components (ξ, η) of the deflection of the vertical (DOV). These components can be determined by astrogeodetic methods.

High accuracy astrogeodetic DOV determination was a very time-consuming and expensive field procedure in the past. By high accuracy we mean 0.1″ precision of DOV components. Recent advances in instrumentation and computer technology have speeded up DOV determination considerably by digital zenith cameras. These are, however, quite expensive, dedicated field instruments with high resolution tiltmeters and very sensitive CCD sensors [1]. On the other hand, it is relatively straightforward to modify a modern robotic total station to be suitable for astrogeodetic DOV determination. QDaedalus is such a system [2].

Accuracy of DOV determination depends on several factors. Assuming a standard data processing workflow, HIRT [3] showed that in case of digital zenith cam-

eras the main accuracy factor is the number of independent star observations. That is, by increasing the number of observations the accuracy can be increased, too. However, no such analysis exists for astrogeodetic DOV determinations by robotic total stations. A second open question is whether standard least-squares data processing is optimal for achieving the desired accuracy with a minimum number of data (or measurement time). As is known, least-squares procedures are optimal only for Gaussian data distributions. Therefore, optimality of data processing depends critically on the distribution of measurements.

Our present work addresses these issues. We analyzed the distribution and accuracy of QDaedalus measurements and compared two different data processing methodologies in terms of their accuracy. The first one was a standard robust least-squares inversion for the unknown parameters (the Danish method). The second one was a new procedure for robust inversion using Cauchy-Steiner weights.

In Sections 2 and 3 an overview of the QDaedalus system and the measurement process are given. Section 4 discusses the inversion problem of DOV determination and its solution with the two methods mentioned above. In the following section we present numerical tests with the two methods and in Section 6 some conclusions are drawn.

2. THE QDAEDALUS SYSTEM

The QDaedalus system is a computer-controlled, GNSS-assisted automatic measuring system, which can be used mainly for high-precision determination of the local vertical direction (deflection of the vertical) [2]. The main system component is an adapted Leica TCA 1800 robotic total station. The eyepiece of the instrument is replaced with a CCD sensor, hence instead of visual observation the image of the telescope can be viewed on the computer monitor. Focusing is also computer controlled by using an attached servomotor [4].

The GNSS receiver serves two purposes: first, it provides accurate timing of star observations for both the CCD sensor and control computer; second, it measures and sends WGS84 coordinates to the computer through an RS232-USB adapter. Large volumes of CCD image data are transmitted through a high-speed (hot) line into the computer via a PCMCIA or EXPRESS CARD interface. The external computer controls the robotic total station through another dedicated channel via another RS232-USB adapter and the interface unit of the total station *(Figure 1)*.

The QDaedalus software controls the whole system and processes data. In particular it guides the robotic total station, focuses the telescope, receives and processes CCD images and GNSS data, and for astrogeodetic measurements it computes topocentric star coordinates, maintains a measurement database, and in situ calculates deflection of the vertical components.

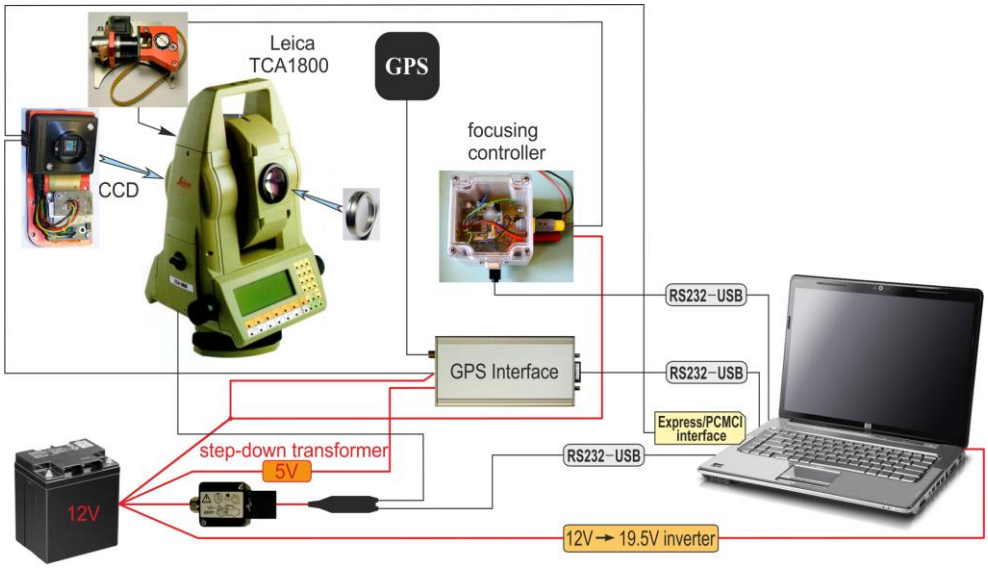

Figure 1
Schematic construction of the QDaedalus system

3. MEASUREMENT OF VERTICAL DEFLECTIONS BY QDAEDALUS SYSTEM

Deflection of the vertical (its components) can be computed from geodetic coordinates of the local vertical and ellipsoidal normal [4]:

$$\xi = \Phi - \varphi, \tag{1}$$

$$\eta = (\Lambda - \lambda)\cos\varphi. \tag{2}$$

GNSS measurements provide geodetic latitude and longitude φ, λ with respect to the WGS84 ellipsoid, whereas astrogeodetic observations provide astronomical latitude and longitude Φ, Λ based on known celestial equatorial coordinates (δ, α) of measured stars [5].

The starting point of astrogeodetic methods is the known celestial equatorial coordinates of observed stars in the quasi-inertial barycentric ICRS system (International Celestial Reference System). Our goal is to determine our position (or local vertical direction) Φ, Λ in the ITRS system (International Terrestrial Reference System), hence we need to transform coordinates of stars δ, α from ICRS into ITRS.

This in practice requires two steps: first we determine local zenith direction with respect to the measured stars and this direction is characterized by the coordi-

nates α_{Zenit} , δ_{Zenit}. In the second step, zenith direction in the celestial equatorial system is transformed into the ITRS system by a coordinate transformation. The relationship between celestial and terrestrial coordinates is shown in *Figure 2*.

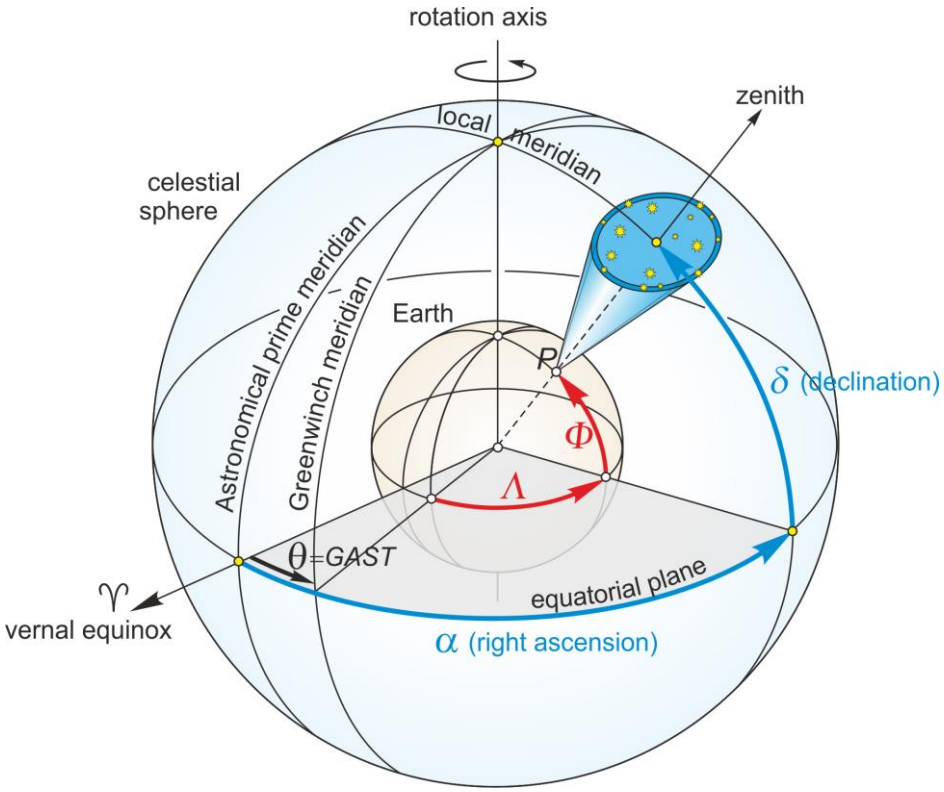

Figure 2
Principle of DOV measurement

Stars very close to zenith are measured with traditional zenith cameras, whereas QDaedalus measures stars along a celestial small circle centered on the zenith. This circle can be visualized as the intersection of the celestial sphere with a circular cone of half apex angle, e.g. 30°, with its axis pointing at zenith. In either case, direction of the local vertical (zenith) is determined from the star measurements.

Due to the software control mentioned, during measurement the pointing axis of the telescope is set to the positions of stars of the FK6 star catalogue computed for the particular observation station and epoch. Stars are selected to have a possibly uniform distribution along the mentioned celestial small circle. Night measurement with QDaedalus is shown in *Figure 3*.

4. INVERSION PROCEDURE

The standing axis of a total station ideally points to the local Zenith direction *(Figure 4)*. Readings *(ℓ, z)* to the pointing axis of the telescope are given in the local horizontal-vertical (Hz-V) system. Position *(x, y)* of the star's image on the CCD sensor can also be determined automatically by image processing. Thereby it is possible to reconstruct (with known calibration constants) the direction to any measured star in the local horizontal-vertical system.

Figure 3
Night measurement by the Qdaedalus system

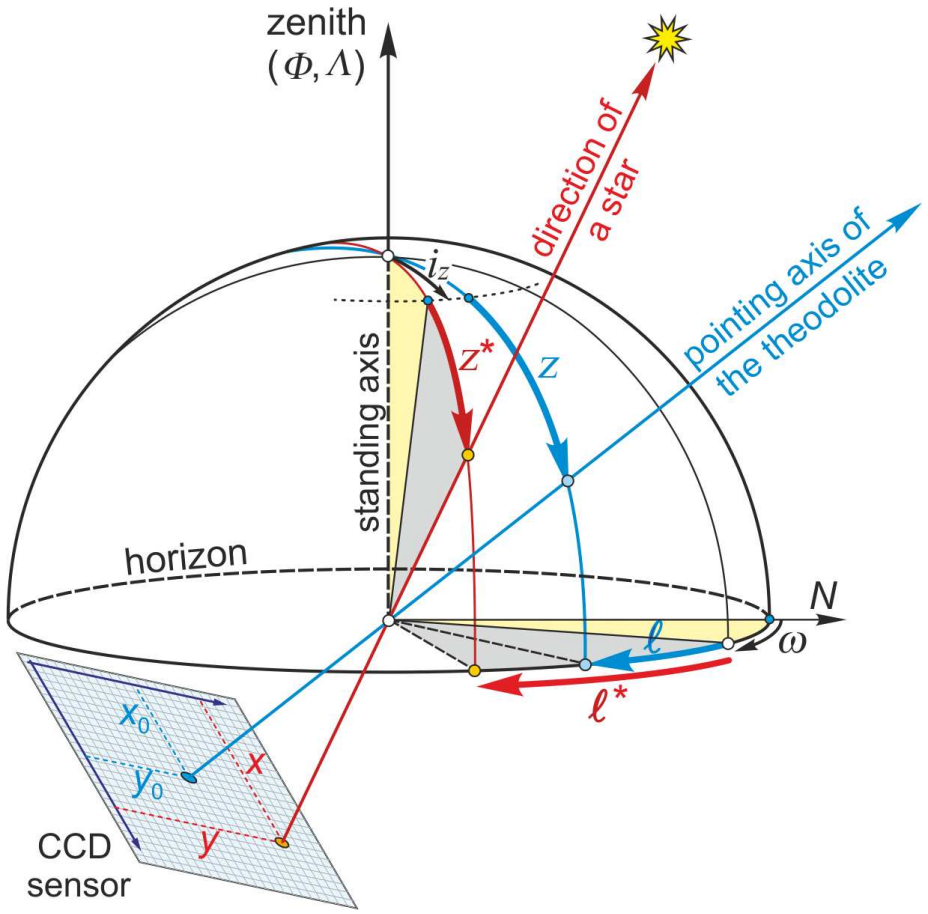

Figure 4
Model of local zenith direction determination

In Section 3 it was mentioned that the position of stars for any measurement epoch can be calculated from barycentric ICRS coordinates in the Earth-fixed ITRS system. By transforming these ITRS coordinates to the local Hz-V system and taking into account astronomical refraction we can compare measured and computed $\left(\ell^*, z^*\right)$ Hz and V directions. These relations are expressed by the following two model equations:

$$\left.\begin{aligned}\left(\ell-\ell^*\right)\sin z^* - a_{11}\left(x-x_0\right) - a_{12}\left(y-y_0\right) = 0\\ z - z^* - a_{21}\left(x-x_0\right) - a_{22}\left(y-y_0\right) = 0\end{aligned}\right\}, \tag{3}$$

where $x_0, y_0, a_{11}, a_{12}, a_{21}, a_{22}$ are calibration constants and

$$\left.\begin{aligned} \ell^* &= f_\ell^*\left(\Phi, \Lambda\right) - \omega \\ z^* &= f_z^*\left(\Phi, \Lambda\right) - i_z - \delta_{refr} \end{aligned}\right\} \tag{4}$$

are functions of the model parameters. Our goal is to determine the model parameter vector

$$\vec{m} = \left\{\Phi, \Lambda, \omega, i_z\right\},$$

where Φ, Λ, ω, i_z, δ_{refr} denote astronomical latitude, longitude, orientation correction, index error of the instrument, and refraction correction, respectively.
We have measurements for each epoch *i* of the telescope axis and position of the star's image on the CCD sensor:

$$\vec{d}_{mi} = \left\{\ell_i, z_i, x_i, y_i\right\}, \quad i = 1, 2, ..., N,$$

where *N* denotes the number of measurement epochs.
Model *Equations (3)* can be written for the *i*-th measurement epoch as

$$\left.\begin{aligned} \left[\ell_i - \ell_i^*\left(\Phi, \Lambda, \omega\right)\right] \sin z_i^*\left(\Phi, \Lambda, i_z\right) - a_{11}\left(x_i - x_0\right) - a_{12}\left(y_i - y_0\right) &= 0 \\ z_i - z_i^*\left(\Phi, \Lambda, i_z\right) - a_{21}\left(x_i - x_0\right) - a_{22}\left(y_i - y_0\right) &= 0 \end{aligned}\right\}. \tag{5}$$

The general form of the above equations is $\vec{g}_i(\vec{d}_{ei}, \vec{m}) = 0$, where $\vec{d}_{ei} = \left\{\ell_i, z_i, x_i, y_i\right\}$ denote theoretical values. The model equations are implicit in both data and parameters. Theoretical values cannot be determined uniquely from model equations since there are 2*N* model equations for 4*N* + 4 unknowns, therefore our problem is underdetermined. We can introduce 4*N* measurement residuals $\vec{e}_i = \vec{d}_{ei} - \vec{d}_{mi}$ instead of theoretical values, and we still have an underdetermined problem. The usual solution to this problem, discussed by e.g. Vanicek and Krakiwsky [6], is to minimize either weighted residuals (length of of $\vec{e}$ in data space)

$$E = \vec{e}^T \underline{\underline{W}}\, \vec{e} = \min \tag{6}$$

or its weighted projection $\vec{e}\,' = \underline{\underline{B}}\, \vec{e}$,

$$E' = \vec{e}'^T \underline{\underline{W}}'\, \vec{e}' = \min \tag{7}$$

in model space under constraint $\vec{g}(\vec{d}_e, \vec{m}) = 0$. We have a constrained minimization problem, which can be solved by the method of Lagrange multipliers. We introduce a vector $\vec{\lambda}$ of Lagrange multipliers or Lagrange correlates from the model space and add the scalar product of Lagrange multipliers and vector of constraints to the minimum condition

$$\Phi\left(\vec{e}, \vec{m}, \vec{\lambda}\right) = \vec{e}^T \underline{\underline{W}}\, \vec{e} + \vec{\lambda}^T \vec{g}\left(\vec{d}_m + \vec{e}, \vec{m}\right) = \min . \tag{8}$$

It has been shown by HADLEY [7] that the addition of the scalar product does not change the location of the extremum. Using the gradient method (linearization) to solve for the extremum yields

$$\left.\begin{aligned} \frac{\partial \Phi}{\partial \vec{e}} &= 2\vec{e}^T \underline{\underline{W}} + \vec{\lambda}^T \frac{\partial \vec{g}}{\partial \vec{d}_e} = 2\vec{e}^T \underline{\underline{W}} + \vec{\lambda}^T \underline{\underline{B}} = 0 \\ \frac{\partial \Phi}{\partial \vec{m}} &= \vec{\lambda}^T \frac{\partial \vec{g}}{\partial \vec{m}} = \vec{\lambda}^T \underline{\underline{G}} = 0 \\ \frac{\partial \Phi}{\partial \vec{\lambda}} &= \vec{g}\left(\vec{d}_m + \vec{e}, \vec{m}\right) \approx \underline{\underline{B}}\, \vec{e} + \underline{\underline{G}}\, \delta\vec{m} + \vec{g}_0 = 0 \end{aligned}\right\} . \tag{9}$$

If we introduce Jacobian matrices of partial derivatives of function: $\vec{g}$

$$\left.\begin{aligned} \frac{\partial \vec{g}}{\partial \vec{d}_e} &= \underline{\underline{B}} \\ \frac{\partial \vec{g}}{\partial \vec{m}} &= \underline{\underline{G}} \end{aligned}\right\} , \tag{10}$$

the following linear system of equations must be solved for determination of minimum

$$\left.\begin{aligned} \underline{\underline{W}}^T \vec{e} + \underline{\underline{B}}^T \vec{\lambda} &= 0 \\ \underline{\underline{G}}^T \vec{\lambda} &= 0 \\ \underline{\underline{B}}\, \vec{e} + \underline{\underline{G}}\, \vec{m} + \vec{g}_0 &= 0 \end{aligned}\right\} , \tag{11}$$

where we substituted $\vec{m}$ for the change of the parameter vector $\delta\vec{m}$.

The inversion solution for the change of the parameter vector is

$$\vec{m} = \left(\underline{\underline{G}}^T \left(\underline{\underline{B}} \underline{\underline{W}}^{-1} \underline{\underline{B}}^T \right)^{-1} \underline{\underline{G}} \right)^{-1} \underline{\underline{G}}^T \left(\underline{\underline{B}} \underline{\underline{W}}^{-1} \underline{\underline{B}}^T \right)^{-1} \vec{g}_0 = \underline{\underline{G}}^{-gh} \vec{g}_0 , \qquad (12)$$

where we introduced on the right-hand side the generalized inverse matrix $\underline{\underline{G}}^{-gh}$.

The result of inversion obviously depends on the chosen weight matrix $\underline{\underline{W}}$. In the next two subsections we discuss two different iterative weighting schemes which aim at robust and resistant solution of the inversion. But first, it seems appropriate to include a brief review on notions of robustness and resistance following STEINER [8].

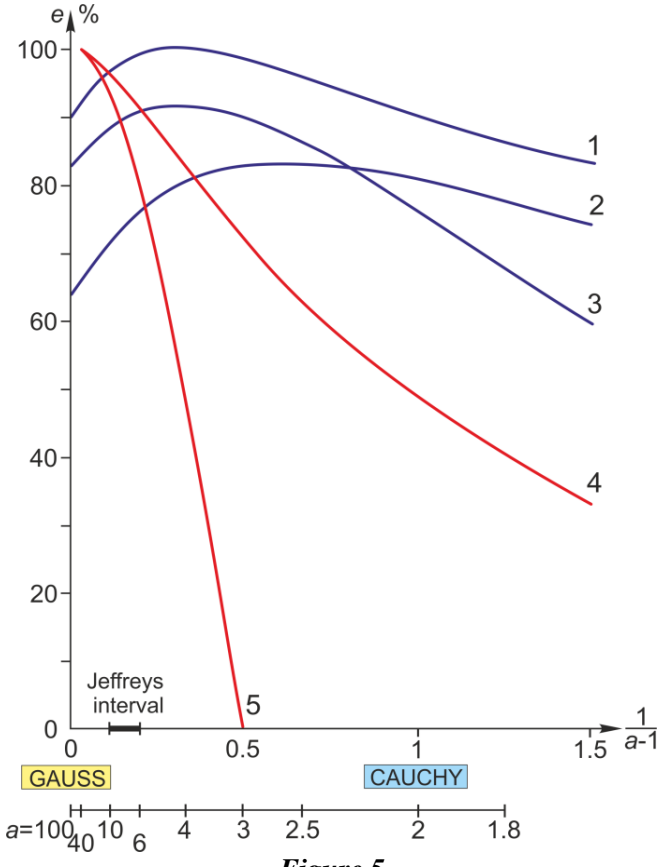

Figure 5

Statistical efficiencies (e %) of some estimation procedures for a large range of probability distribution types [the supermodel $f_a(x)$]. Same as *Figure 2* of STEINER [8]. 1: P_k norm, k = 1.9; 2: median; 3: Danish method, c = 1.5; 4: Danish method, c = 3; 5: L_2 norm

Robustness refers to the high efficiency of estimation for a whole family of distributions (supermodels). Resistance means that our estimation is resistant to outliers. STEINER [8] discussed statistical efficiency of several estimation methods. He obtained for the $f_a(x)$ supermodel the efficiencies visualized in Figure 5. We can see that Method 1, P_k norm (Cauchy-Steiner weights) has higher statistical efficiency for this supermodel than Method 2, the so-called Danish method [9] with parameter $c = 3$. We can see that indeed the Danish method has 100% efficiency for Gaussian distribution, but its efficiency decreases sharply and has approximately only 50% for long-tailed Cauchy distribution. In Section 5 we will test QDaedalus measurement residuals to see if they follow a Gaussian distribution.

4.1. Robust inversion by the Danish method

Resistance to outliers can be ensured by iteratively re-computing weights of residuals according to the so-called Danish method by KRARUP and KUBIK [9]. Weights of the Danish method are defined by the following formulas:

$$\left.\begin{aligned} w_{k+1} &= w_k f\left(e_k\right) \\ f\left(e_k\right) &= \begin{cases} 1, \text{if } \dfrac{\left|e_k\right|\sqrt{w_0}}{\sigma_0} < c \\ \text{else} \exp\left(-\dfrac{\left|e_k\right|\sqrt{w_0}}{c\sigma_0}\right) \end{cases} \end{aligned}\right\}, \tag{13}$$

where k is the iteration number, e_k is residual after step k, σ_0 is square root of the variance factor, w_0 is initial weight and c is constant (usually equal to 3).

4.2. Robust inversion by Cauchy-Steiner weights

Our aim is minimization of residuals $\vec{e'}$ according to *Equation (7)* by using Cauchy-Steiner weights. Projection $\vec{e}' = \underline{\underline{B}}\,\vec{e}$ transfers residuals from measurement space into model space. Constrained minimization of the logarithm of the P_k-norm [10]

$$P_k\left(\vec{x}\right) = \varepsilon\left\{\prod_{i=1}^{n}\left[1+\left(\frac{x_i}{k\varepsilon}\right)^2\right]\right\}^{\frac{1}{2n}}, \tag{14}$$

where ε is the dihesion of the deviations x_i defined by

$$\varepsilon^2 = \frac{3\sum_{i=1}^{n} \frac{x_i^2}{\left(\varepsilon^2 + x_i^2\right)^2}}{\sum_{i=1}^{n} \frac{1}{\left(\varepsilon^2 + x_i^2\right)^2}}, \tag{15}$$

leads by using the method of Lagrange multipliers to the equation

$$\Phi\left(\vec{e}, \vec{m}, \vec{\lambda}\right) = \ln P_k\left(\vec{e}\,'\right) + \vec{\lambda}^T \vec{g}\left(\vec{d}_m + \vec{e}, \vec{m}\right) = \min . \tag{16}$$

Finding the minimum requires vanishing derivatives again using the gradient method, and leads to the equations

$$\left.\begin{aligned} \frac{\partial \Phi}{\partial \vec{e}} &= \frac{\partial \ln P_k\left(\vec{e}\,'\right)}{\partial \vec{e}} + \vec{\lambda}^T \underline{\underline{B}} = \frac{1}{n} \vec{e}^T \underline{\underline{B}}^T \underline{\underline{\underline{W}}}' \underline{\underline{B}} + \vec{\lambda}^T \underline{\underline{B}} = 0 \\ \frac{\partial \Phi}{\partial \vec{m}} &= \vec{\lambda}^T \underline{\underline{G}} = 0 \\ \frac{\partial \Phi}{\partial \vec{\lambda}} &= \underline{\underline{B}} \vec{e} + \underline{\underline{G}} \vec{m} + \vec{g}_0 = 0 \end{aligned}\right\}, \tag{17}$$

where

$$\underline{\underline{\underline{W}}}' = \left\langle \ldots, \frac{\left(k\varepsilon'\right)^2}{\left(k\varepsilon'\right)^2 + e_i'^2}, \ldots \right\rangle \tag{18}$$

is the diagonal matrix of Cauchy-Steiner weights belonging to residuals $\vec{e}'$.
The following linear system of equations must be solved to get the optimal parameter vector

$$\left.\begin{aligned} \underline{\underline{B}}^T \underline{\underline{\underline{W}}}' \underline{\underline{B}} \vec{e} + \underline{\underline{B}}^T \vec{\lambda} &= 0 \\ \underline{\underline{G}}^T \vec{\lambda} &= 0 \\ \underline{\underline{B}} \vec{e} + \underline{\underline{G}} \vec{m} + \vec{g}_0 &= 0 \end{aligned}\right\}. \tag{19}$$

Hence, the inversion solution will be the same as above if we make the substitution $\underline{\underline{B}}^T \underline{\underline{\underline{W}}}' \underline{\underline{B}} \to \underline{\underline{W}}$

$$\vec{m} = \left(\underline{\underline{G}}^T \left(\underline{\underline{B}} \underline{\underline{W}}^{-1} \underline{\underline{B}}^T \right)^{-1} \underline{\underline{G}} \right)^{-1} \underline{\underline{G}}^T \left(\underline{\underline{B}} \underline{\underline{W}}^{-1} \underline{\underline{B}}^T \right)^{-1} \vec{g}_0 = \underline{\underline{G}}^{-gh} \vec{g}_0 \ . \qquad (20)$$

Therefore, we have obtained a robust solution to our inversion problem with Cauchy-Steiner weights. Our experience showed that this formulation yielded good distribution of residuals. On the contrary, when we applied Cauchy-Steiner weights for original residuals $\vec{e}$ (i.e. not projected) we were unable to obtain a good distribution of residuals (certain sets of residuals always became very small at the expense of the others). This may be due to the application of the P_k-norm instead of the usual least-squares norm.

5. INVERSION TEST RESULTS

At the station called Pistahegy, located in the southern part of Budapest, we have more than 70 series of DOV measurements taken on 33 different nights. The average measurement time of each observation is about 50 min. We processed these observations by the two different inversion methods discussed above. The parameter of the P_k-norm was $k = 1$ in all our tests. DOV components in the GRS80 reference system (MORITZ [11]) and their estimated uncertainties (standard deviations) obtained by both methods are shown in *Figure 6*.

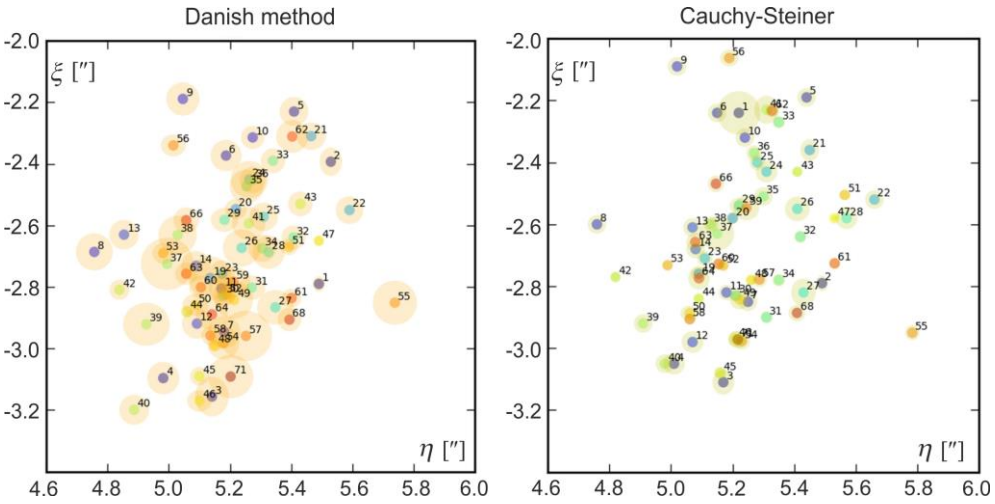

Figure 6

DOV inversion results of 72 QDaedalus measurement series by the Danish method vs. Cauchy-Steiner weights at Pistahegy Station. Outlying observations are not shown. Numbers of series are given on the plots. Ellipses indicate estimated inversion errors.

It is clear from *Figure 6* that inversion with Cauchy-Steiner weights generally gives smaller standard deviation of DOV components. Another useful type of information on the quality of inversion is the correlation norm, the average of off-diagonal elements of the parameter correlation matrix. The distribution of the absolute value of the correlation norm for the two inversion methods is shown in *Figure 7*. We can see that inversion by Cauchy-Steiner weights usually gives smaller correlation between the parameters.

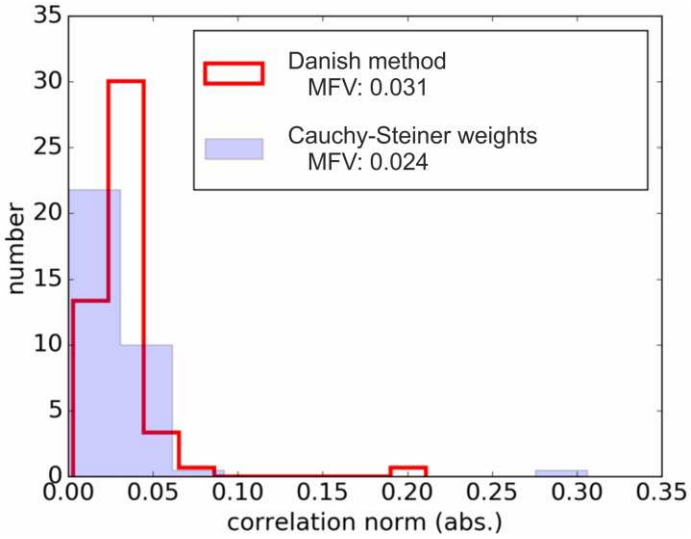

Figure 7
Distribution of absolute value of the correlation norm of DOV inversion results of 72 QDaedalus measurement series by Danish method vs. Cauchy-Steiner weights at Pistahegy Station

5.1. Distribution of inversion residuals

We made tests on whether the distribution of inversion residuals of the Danish method follows Gaussian distribution. *Figure 5* shows that if this is the case, the Danish method gives 100% efficiency of estimation. We performed one-sample Kolmogorov-Smirnov (K-S) tests to see whether the distribution of inversion residuals follows a Gaussian theoretical distribution. Empirical cumulative distribution (ECDF) of residuals for both vertical (V) and horizontal (Hz) residuals were computed and best fitting Gaussian CDFs are determined. It is important to take into account the fact that when the distribution parameters are estimated from data, the K-S test will be distribution dependent (KEUTELIAN [12]). Hence proper test statistic distributions and confidence levels have to be determined by Monte Carlo simulation.

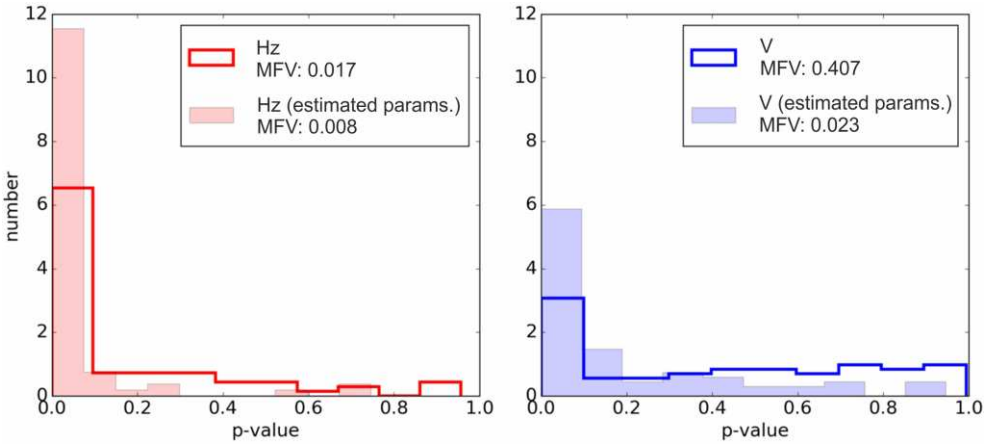

Figure 8

Distribution of p-values of the Kolmogorov-Smirnov test with fixed and estimated parameters for both horizontal and vertical angle measurement residuals for 72 measurement series. Most frequent values (MFV) of p-values are also shown. Monte Carlo simulation was done using 10,000 samples.

The distribution of K-S test statistic *p*-values is shown in *Figure 8* for horizontal and vertical measurement residuals. The most frequent values of the K-S test *p*-values for the estimated parameters are 0.8% and 2.3% for horizontal and vertical measurement residuals, respectively. These low probabilities indicate that inversion results actually do not follow a Gaussian distribution for these 72 QDaedalus measurements. If, instead, we mistakenly treat distribution parameters (mean, std of Gaussian) as fixed, these probabilities will increase to 1.7% and 40.7%, respectively. This may lead to the (false) acceptance of a Gaussian PDF, at least for V residuals.

These test results are in accordance with our observation that Cauchy-Steiner inversion with P_k-norm yields consistently better results than the traditional least-squares estimation applied in geodesy, even if the estimation is made resistant with a re-weighting scheme like the Danish method. M-inversion gives higher statistical efficiencies for distributions significantly differing from Gaussian distribution (cf. *Figure 5*).

5.2. Accuracy dependence on the number of measurement epochs

If the law of large numbers is fulfilled, it is expected that by increasing the number of observations the accuracy can be increased, too. Hence we made tests to see how increasing the number of star measurements (number of measurement epochs) affects the estimated accuracy of the DOV components. We chose the longest continuous QDaedalus measurement No. 47 at Pistahegy, which lasted for 301 minutes. This data set contained 3,746 measurement epochs and we selected 3,085 epochs for processing. We note that to increase accuracy, our standard QDaedalus

processing excludes each first measurement of a star from a series of continuous measurements to the same star. In this particular case we rejected the first from a series of 5 measurements.

The accuracy of inversion by the Danish method and Cauchy-Steiner weights for this particular measurement for DOV components ξ, η is shown in *Figure 9*. The horizontal axes of Figure 9 are scaled as $1/\sqrt{n}$, since if the law of large numbers is fulfilled, data points would be located along straight lines. We can see only partial fulfillment of this law, so other factors must contribute here. What is clearly seen, however, is the difference between the accuracy of the two methods. Inversion with Cauchy-Steiner weights consistently gives higher accuracies, and we can estimate that 0.05″ accuracy in this case can be reached after about 300 epochs or 30 minutes of measurement time.

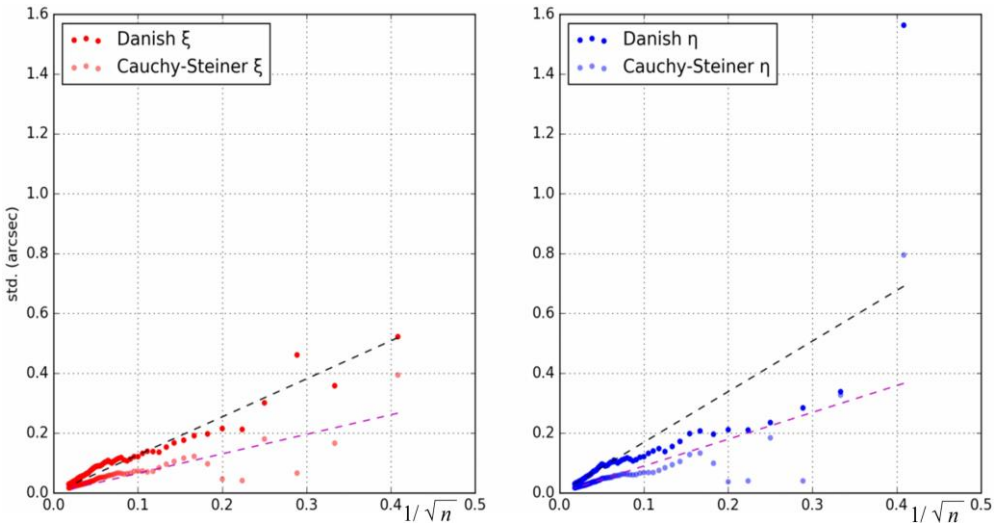

Figure 9

Accuracy of DOV components ξ, η determined by inversion as a function of the number of measurement epochs for measurement No. 47 at Pistahegy. If the law of large numbers were fulfilled, points would have to be located along a straight line. Best-fitting lines are also shown

6. CONCLUSIONS

We have compared two robust and resistant methods for the inversion of QDaedalus zenith direction determination. M-inversion using Cauchy-Steiner weights [13] yielded consistently better inversion results in terms of accuracy and correlation norm based on more than 70 independent measurement series. This is probably due to the higher statistical efficiency of M-inversion in cases of non-Gaussian distribution of residuals compared with the robust E-estimation with Danish method commonly used in geodetic data processing. By analyzing the accuracy of deflection of

the vertical components we can state that 15–30 minutes of measurement time by QDaedalus is probably enough to reach DOV accuracy below 0.1″.

ACKNOWLEDGEMENTS

Our investigations was supported by the National Scientific Research Fund, Project numbers OTKA K-124286.

LIST OF SYMBOLS

Symbol	Description	Unit
CDF	cumulative distribution function	–
DOV	deflection of the vertical, its components are: ξ, η	arcsec [″]
ECDF	empirical cumulative distribution function	–
ICRS	International Celestial Reference System	–
ITRS	International Terrestrial Reference System	–
MFV	most frequent value	–
K-S	Kolmogorov-Smirnov	–
PDF	probability distribution function	–
α, δ	equatorial coordinates (right ascension and declination)	°
δ_{refr}	refraction correction of vertical angles	°
$\underline{\underline{B}}$	first Jacobian matrix of the implicit measurement function	–
ε	dihesion of residuals	arcsec [″], pixel
$\vec{g}$	model function (implicit)	–
$\underline{\underline{G}}$	second Jacobian matrix of the implicit measurement function	–
ξ, η	N-S and E-W components of deflection of the vertical	arcsec [″]
$\vec{e}$	measurement residual vector	arcsec [″], pixel
$\vec{e'}$	projected measurement residual vector	arcsec [″], pixel
i_z	index error of the total station	arcsec [″]
$\vec{\lambda}$	vector of Lagrange multipliers	–
$\vec{m}$	parameter vector	°
ℓ, z	horizontal and vertical angles of the pointing axis of the telescope	°
ℓ^*, z^*	horizontal and vertical angles of a star	°

Symbol	Description	Unit
x_0, y_0	position of the image of the pointing axis of the telescope	pixel
x, y	position of the image of a star on the CCD sensor	pixel
φ, λ	geodetic latitude and longitude	°
Φ, Λ	astronomical latitude and longitude	°
$\underline{\underline{W}}$	weight matrix	–
$\underline{\underline{W'}}$	matrix of Cauchy-Steiner weights	–
ω	orientation correction	°

REFERENCES

[1] HIRT, C.–BÜRKI, B.–SOMIESKI, A.–SEEBER, G.: Modern determination of vertical deflections using digital zenith cameras. *Journal of Surveying Engineering*, 2010, 136(1), 1–12, DOI: doi:10.1061/_ASCE_SU.1943-5428.0000009.

[2] BÜRKI, B.–GUILLAUME, S.–SORBER, P.–OESCH, P. H.: Daedalus: A Versatile Usable Digital Clip-on Measuring System for Total Stations. *2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 15–17 September 2010, Zürich, Switzerland.

[3] HIRT C.–SEEBER G.: Accuracy analysis of vertical deflection data observed with the Hannover Digital Zenith Camera System TZK2-D. *Journal of Geodesy*, 2008, 82(6), 347–56.

[4] VÖLGYESI L.–TÓTH G.: Possibilities of determinations of the vertical deflections, *Geomatikai Közlemények*, 2015, XVIII (2), pp. 75–83. (In Hungarian)

[5] VÖLGYESI L–TÓTH GY.: Possibilities of geodetic application of the QDaedalus system, *Geodézia és Kartográfia*, 2016, LXVIII (9–10) pp. 11–17. (In Hungarian)

[6] VANICEK, P.–KRAKIWSKY, E. J.: *Geodesy: The Concepts*. (2nd Edition), Elsevier Science: North-Holland, Amsterdam, 1986.

[7] HADLEY, G.: *Nonlinear and Dynamic Programming*. Addison-Wesley: Reading, Massachusets, 1964.

[8] STEINER, F.: Possibilities to realize higher efficiency in geophysical interpretation. *Geophysical Transactions,* 1987, 33(1), 3–9

[9] KRARUP, T.–KUBIK, K.: The Danish method, experience and philosophy. Seminar Mathematical models of geodetic photogrammetric point determination with regard to outliers and systematic errors. Ed.: F. E. Ackermann, München. *Deutsche Geodätische Kommission Reihe A*, 1983, 98, 131–134

[10] HAJAGOS, B.–STEINER, F.: The P*-norm. Acta Geodaetica Geophysica et Montanistica Hungarica, 1991, 26(1–4), 153–182

[11] MORITZ, H.: Geodetic Reference System 1980. *Journal of Geodesy*, 2000, 74(1), 128–162, doi:10.1007/S001900050278

[12] KEUTELIAN, H.: The Kolmogorov-Smirnov Test When Parameters Are Estimated From Data Fermilab report, 1991, DCF/EWK/DOC/PUBLIC/1285, 1–13.

[13] STEINER, F.: Most frequent value procedures (a short monograph). *Geophysical Transactions,* 1988, 34, (2–3), 139–260.