MultiScience - XXX. microCAD International Multidisciplinary Scientific Conference
University of Miskolc, Hungary, 21-22 April 2016, ISBN 978-963-358-113-1

NEW STATISTICAL APPROACH FOR WATER CONTENT DETERMINATION IN SHALLOW GEOLOGICAL ENVIRONMENT

Gergely Pál Balogh
Ph.D. student
Department of Geophysics, University of Miskolc, 3515, Miskolc-Egyetemváros, Hungary

ABSTRACT

A further-developed factor analysis method is presented to estimate water content in shallow subsurface sediments. Resistivity, natural gamma-ray, neutron-thermal neutron and density logs recorded by cone penetration tools are processed simultaneously to estimate the vertical variations of factor scores along a penetration hole. After the phase of factor analysis, regression tests are performed to relate the factor variables to petrophysical parameters of subsoils. In this study, a strong linear relationship is detected between the water volume and the first statistical factor. The new factor analysis procedure is based on the iterative reweighting of data prediction errors using the highly robust most frequent value method, which improves the estimation accuracy of factor scores in case of non-Gaussian data sets.

INTRODUCTION

The engineering geophysical sounding (EGS) method has been applied to the in-situ investigation of shallow freshwater-bearing formations for more than two decades [1]. During the measurement, a cone-shaped tip is pushed into the ground, while geophysical parameters including nuclear and electric data are recorded with the sensors installed in the tube behind the cone. The well logs of observed quantities provide information on the composition, water saturation and geotechnical parameters of subsoils. EGS data processing mostly incorporates deterministic or inversion methods adapted from oilfield well log interpretation [2], [3].

Factor analysis is conventionally applied to reduce the dimensionality of statistical problems and extract latent information from the statistical sample [4]. Factor analysis of well log data was used for the calculation of shale volume in hydrocarbon reservoirs [5]. Exploratory factor analysis of EGS data was suggested in [6] to estimate the water saturation of near-surface sediments and to simulate the neutron-porosity log to missing intervals. The same method was used for the derivation of dry density from the factor scores [7]. Jöreskog suggested a fast non-iterative factor analysis technique, which gives optimal results for normally distributed data [8]. Since this method is rather sensitive to non-Gaussian noises, the classical factor analysis algorithm has been further developed for giving a more robust solution. By applying a weighting matrix including diagonal elements proportional to the deviation of the measured and calculated data (i.e. prediction error), an iteratively reweighted least squares solution for the factor scores is given.

DOI: 10.26649/musci.2016.004

The solution of classical factor analysis is further improved by incorporating the most frequent value (MFV) method used as a weighting procedure [9]. The Steiner weights have been previously used in the establishment of a robust inversion-based Fourier transformation method, which showed high noise rejection capability [10]. In this study, the MFV method-based factor analysis procedure is tested on EGS data collected from a Hungarian borehole and its result is compared to that of local inverse modelling.

INVERSION OF EGS DATA

The rock matrix of near-surface layers is composed of course and fine grain components, while their pore spaces are saturated with freshwater and some amount of air. The model vector of the inverse problem is defined as follows

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

which contains the volumetric ratios of clay (V_{cl}), sand (V_s), water (V_w) and gas (V_g), (T is matrix transpose). Usually natural gamma-ray (*GR*), density (*DEN*), neutron-porosity (*NPHI*) and resistivity (*RES*) data are measured by EGS instruments. The following response equations can be used for calculating EGS data in the forward problem

$$\mathrm{GR} = \mathrm{V_{cl}GR_{cl}} + \mathrm{V_s GR_s}, \tag{2}$$

$$\mathrm{DEN} = \mathrm{V_w}\rho_{\mathrm{w}} + \mathrm{V_{cl}}\,\rho_{\mathrm{cl}} + \mathrm{V_s}\rho_{\mathrm{s}}, \tag{3}$$

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

$$\mathrm{RES} = \mathrm{a}\left(\mathrm{V_w} + \mathrm{V_g} + \mathrm{V_{cl}}\right)^{-\mathrm{m}} \left(\frac{\mathrm{V_{cl}}/(\mathrm{V_w}+\mathrm{V_{cl}})}{\mathrm{R_{cl}}} + \frac{1-[\mathrm{V_{cl}}/(\mathrm{V_w}+\mathrm{V_{cl}})]}{\mathrm{R_w}}\right)^{-1} \left(\frac{\mathrm{V_w}+\mathrm{V_{cl}}}{\mathrm{V_w}+\mathrm{V_g}+\mathrm{V_{cl}}}\right)^{-\mathrm{n}}, \tag{5}$$

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 parameters *m*, *a*, *n* represent the cementation exponent, tortuosity factor and saturation exponent, respectively. The vector of observed data in given depth is

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

Local inverse problems are solved separately in adjacent depths. The objective function is chosen as the square of the Euclidean norm of the difference between the measured and calculated data. Since the EGS data are of different magnitudes, each deviation is normalized by the observed data [11]

$$\sum_{k=1}^{N}\left(\frac{\mathrm{d}_k^{(\mathrm{calculated})} - \mathrm{d}_k^{(\mathrm{measured})}}{\mathrm{d}_k^{(\mathrm{measured})}}\right)^2 = \min, \tag{7}$$

where N is the number of data processed in a given depth. Since the gas volume is calculated by $V_g=1-V_w-V_{cl}-V_s$, the inverse problem is overdetermined; therefore, a stable inversion procedure can be solved by using the Gaussian least squares method [12]. By the inversion process, the components of vector **m** with their standard deviations can be properly estimated.

STEINER WEIGHTED FACTOR ANALYSIS

The input of factor analysis is the N-by-K matrix of standardized EGS data (**D**), which is decomposed into two matrices

$$\mathbf{D}=\mathbf{F}\mathbf{L}^{\mathrm{T}}+\mathbf{E}, \tag{8}$$

where **F** is the N-by-M matrix of factor scores, **L** is the K-by-M matrix of factor loadings and **E** is the matrix of residuals (M is the number of extracted factors, N is the number of sampled depths and K is the number of EGS logs). The factor loadings and the factor scores are generally estimated simultaneously by the maximum likelihood method (MLM). Jöreskog proposed a fast non-iterative algorithm for calculating the factor loadings followed by an MLM estimation for the factor scores, which gives optimal results for normally distributed data [8]. To formulate the robust algorithm of factor analysis, the classical model of factor analysis defined in Eq. (8) is modified

$$\mathbf{d}=\tilde{\mathbf{L}}\mathbf{f}+\mathbf{e}, \tag{9}$$

where **d** denotes the KN column vector of EGS data, $\tilde{\mathbf{L}}$ is the NK-by-NM matrix of factor loadings, **f** is the MN vector of factor scores, **e** is the KN length vector of prediction error. In the first step of the procedure, we give an estimate to the initial values of factor loadings and scores by Jöreskog’s method. Then the starting values are refined gradually by an iterative algorithm, which takes the form in the q-th iteration step as

$$\mathbf{L}^{\mathrm{T}(q)}=\left(\mathbf{F}^{\mathrm{T}(q-1)}\mathbf{F}^{(q-1)}+\alpha^2\mathbf{I}\right)^{-1}\mathbf{F}^{\mathrm{T}(q-1)}\mathbf{D}, \tag{10}$$

$$\mathbf{f}^{(q)}=\left(\tilde{\mathbf{L}}^{\mathrm{T}(q-1)}\mathbf{W}\tilde{\mathbf{L}}^{(q-1)}\right)^{-1}\tilde{\mathbf{L}}^{\mathrm{T}(q-1)}\mathbf{W}\mathbf{d}, \tag{11}$$

where α is a properly chosen damping factor. The elements of the NK-by-NK diagonal weighting matrix **W** are proportional to the deviation of measured (**d**) and calculated data ($\tilde{\mathbf{L}}\mathbf{f}$). The elements of matrix **W** are the Steiner weights

$$W_{kk}=\frac{\varepsilon^2}{\varepsilon^2+(e_k)^2}, \tag{12}$$

where the scale parameter ε called dihesion is calculated in the following iterative procedure (k=1,2,…,KN and j is the number of iterations)

$$\varepsilon_{j+1}^{2}=\frac{3\cdot\sum_{k=1}^{KN}\frac{\left(e_{k}-M_{j}\right)^{2}}{\left[\varepsilon_{j}^{2}+\left(e_{k}-M_{j}\right)^{2}\right]^{2}}}{\sum_{k=1}^{KN}\frac{1}{\left[\varepsilon_{j}^{2}+\left(e_{k}-M_{j}\right)^{2}\right]^{2}}}\Leftrightarrow M_{j+1}=\frac{\sum_{k=1}^{KN}\frac{\varepsilon_{j+1}^{2}}{\varepsilon_{j+1}^{2}+\left(e_{k}-M_{j}\right)^{2}}e_{k}}{\sum_{k=1}^{KN}\frac{\varepsilon_{j+1}^{2}}{\varepsilon_{j+1}^{2}+\left(e_{k}-M_{j}\right)^{2}}}. \qquad (13)$$

The optimal values of dihesion and parameter *M* called the most frequent value are estimated simultaneously by the MFV method [9]. The larger the distance between the observed and predicted data in Eq. (12), the less weight given to the relevant datum. The above iterative factor analysis procedure is named MFV-FA method.

CASE STUDY

The MFV-FA method is tested in a penetration hole drilled in Bátaapáti, south-west Hungary. The following EGS data types were measured in a shallow sedimentary geological environment: cone resistance (*RCPT*), natural gamma-ray (*GR*), density (*DEN*), neutron-porosity (*NPHI*) and resistivity (*RES*). The *RCPT* log responds to the solidity of soil, the *GR* log is mainly sensitive to the clay content and lithology, the *DEN* log gives information on porosity and bulk density, the *NPHI* log gives total porosity, and the *RES* log is primarily used in water content estimation. The values of zone parameters are listed in Table 1, which were chosen similarly to [6]. The physical properties of air found in Eqs. (2)–(4) are practically set to be zero.

Table 1

The values of fixed zone parameters

Zone parameter		**Clay**	**Sand**	**Water**	**Unit**
GR	–	8.90	1.65	0	cpm
DEN	–	2.05	2.30	1.00	g/cm^3
NPHI	–	0.20	0.00	1.00	v/v
RES	–	6.00	–	6.70	ohm·m
m	1.7	–	–	–	–
a	1.0	–	–	–	–
n	2.0	–	–	–	–

The results of Jöreskog’s procedure are improved by the MFV-FA procedure using Eqs. (10)–(12). The relative distance between the measured and calculated data decreases progressively with the number of iterations. The factor loadings and factor scores are updated in 15 iteration steps. In addition, in each step of the iterative procedure the Steiner weights are recalculated in further 30 steps. In this inner loop, dihesion is decreased automatically and changed differently for each EGS log (Figure 1-a). Beside the same value of ε, the larger the distance between measured and calculated data, the smaller the weight. With the decrease of ε, bigger deviations contribute less to the solution. The optimal values of diagonal elements of the weighting matrix **W** are shown in Figure 1-b, where the weighting coefficients are represented as a function of the prediction error.

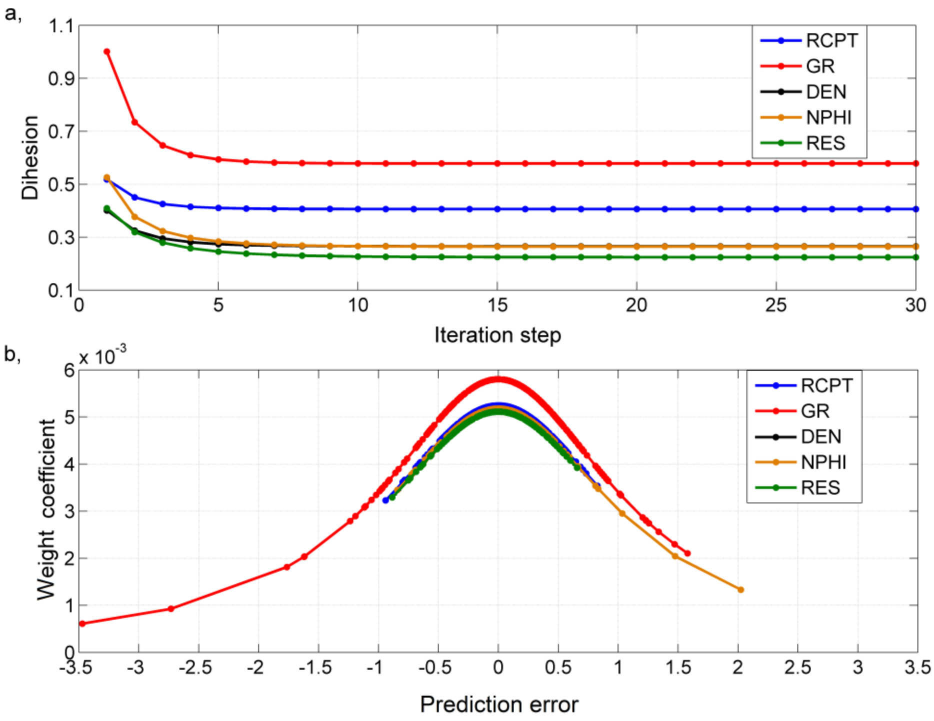

Figure 1
Decrease of dihesion during the MFV-FA procedure (a), optimal weights used in factor analysis of engineering geophysical sounding data (b).

Two independent factors are estimated by the MFV-FA procedure. The first one explains 69.5 % of the total variance of EGS data and its loadings are 0.28 (*RCPT*), 0.04 (*GR*), –0.93 (*DEN*), –0.94 (*NPHI*), 0.95 (*RES*). Between the first factor and water content-sensitive logs (*NPHI*, *RES*) a very strong correlation is indicated. The first factor (F_1) is connected to water content (V_w) in regression analysis (Figure 2-a). The regression coefficients of the linear relation are estimated with their 95 % confidence bounds (a=–0.034±0.001, b=–0.16±0.01). The Pearson's correlation coefficient (R) between the first factor and water content also shows strong relation.

Inverse modelling is useful to confirm the results of the MFV-FA procedure (Figure 2-b). The forward problem is based on Eqs. (2)–(5). After solving a set of local inverse problems along the borehole, the measured and theoretical EGS logs calculated from the resultant model show good similarity (tracks 1–4 in Figure 3). The average RMS between the observed and calculated data is 4.26 %. The computed values of volumetric parameters in vector **m** are plotted in the last track. The two factor logs are given in track 5, while the water content logs extracted by robust factor analysis (*VW_MFV-FA*) and inverse modeling (*VW_INV*) are in track 6. It is observable that the water content curve estimated by factor analysis is somewhat smoother than that of inverse modeling. The reason of it is that factor analysis gives an outlier resistant solution, while inverse modeling is more sensitive to data noises. The average RMS between the water content logs estimated by the two independent procedures is 1.35 %.

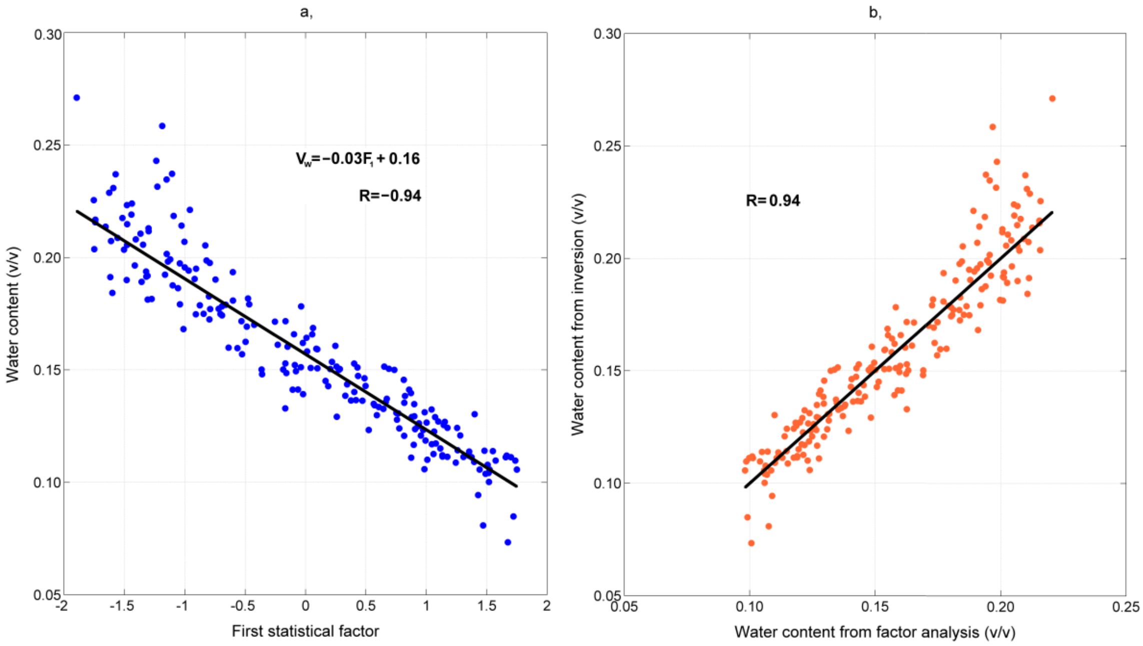

Figure 2
Linear regression relation between the first factor and water content in shallow formations (a), connection between water contents estimated by the MFV-FA procedure and local inversion (b).

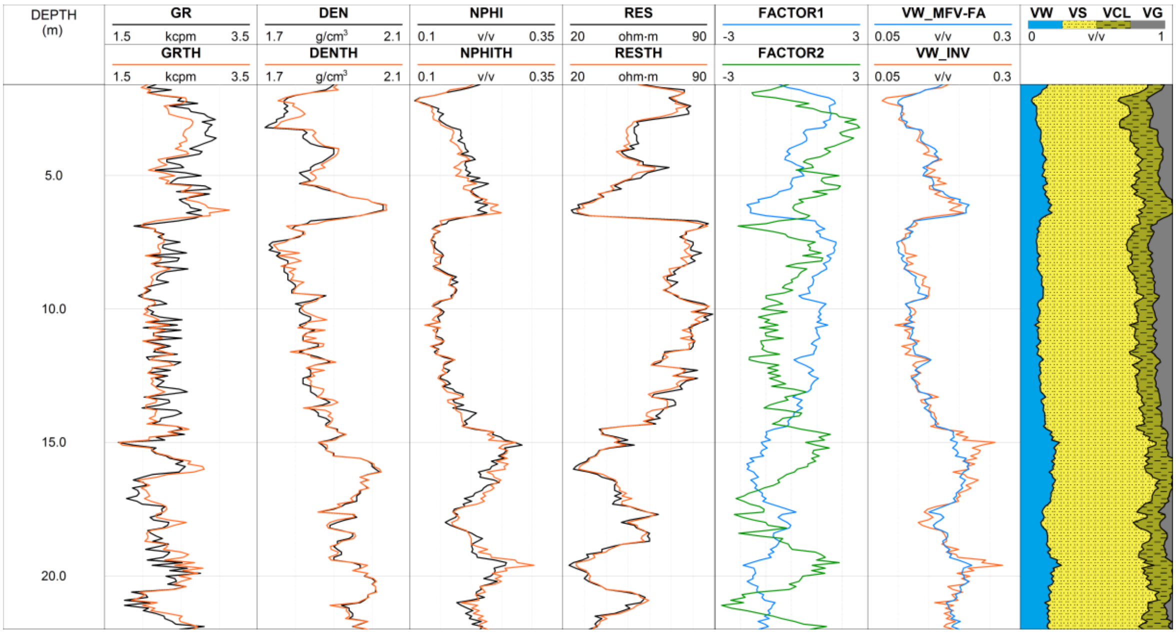

Figure 3
Results of MFV-FA and local inversion of engineering geophysical sounding data in a Hungarian borehole.

CONCLUSIONS

Engineering geophysical sounding data sets can be transformed into statistical factors accurately by using an iterative robust factor analysis procedure. By using the MFV-FA method, the uncertainty of data is also taken into account in the solution. Strong correlation is found between the first factor and water content in shallow sediments, and the linear regression function is specified in a Hungarian measurement area. The statistical method is applicable in the processing of real engineering geophysical sounding data. The results of classical factor analysis can be further improved by using the MFV-FA method. Inverse modelling also confirms the results of factor analysis. The suggested method supports the petrophysical modelling of heterogeneous shallow formations, which can be applied in solving engineering and environmental problems.

ACKNOWLEDGEMENTS

The author thanks to the support of the Hungarian Scientific Research Fund for OTKA project PD-109408, Assoc. Prof. Norbert Péter Szabó and Prof. Mihály Dobróka (University of Miskolc) for their scientific advices and Assoc. Prof. Dezső Drahos (Eötvös Loránd University) for the co-operation.

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. SEG, 321–331.

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

[3] SZABÓ, N. P., DOBRÓKA, M. **Float-encoded genetic algorithm used for the inversion processing of well-logging data**, In: Michalski A (ed.) Global Optimization: Theory, Developments and Applications: Mathematics Research Developments, Computational Mathematics and Analysis Series. New York: Nova Science Publishers, 2013. pp. 79–104.

[4] LAWLEY, D. N., MAXWELL, A. E. **Factor analysis as a statistical method**. The Statistician, 12, 1969. 209–229.

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

[6] 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.

[7] SZABÓ, N. P. **Dry density derived by factor analysis of engineering geophysical sounding measurements**. Acta Geodaetica et Geophysica Hungarica, 47:(2), 2012. pp. 161**–**171.

[8] JÖRESKOG, K. G. **Factor analysis and its extensions**. In: R. Cudeck and R. C. MacCallum (eds.), Factor analysis at 100, historical developments and future directions: Lawrence Erlbaum Associates, Publishers, 2007. 47–77.

[9] STEINER, F. **The most frequent value**. Introduction to a modern conception of statistics. Academic Press, Budapest 1991.

[10] SZEGEDI, H.., DOBRÓKA, M. **On the use of Steiner’s weights in inversion-based Fourier transformation: robustification of a previously published algorithm**. Acta Geodaetica et Geophysica, 49, 2014. 95–104.

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

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