We 09 05

Shale Indicator Derived from Multivariate Statistical Analysis of Well Logs

N.P. Szabo* (University of Miskolc), M. Dobróka (University of Miskolc), J. Somogyi Molnár (University of Miskolc) & R. Kavanda (DACHS GmbH.)

SUMMARY

In the paper a geostatistical approach is presented to estimate shale volume in shaly sand reservoirs. The factor analysis of well-logging data results in a new log, which correlates with the shale content of the formations. The connection between shale content and factor scores is quantified by an empirical relationship between the two variables. The nonlinear formula seems to be straight in different areas and gives consistent results both in water and hydrocarbon reservoirs. In the paper, statistical interpretation results of three different data sets originated from Hungary and the United States of America are compared. The results are verified by estimates of independent petrophysical interpretation.

Introduction

Petrophysical parameters of rocks are derived from well-logging data by deterministic or inversion methods. Former procedures substitute data to explicit equations in order to determine non-measurable parameters separately. For instance, there are approximately 20 different methods for the estimation of shale volume (Asquith and Krygowski 2004). Inversion techniques assume known relationships between the data and relevant petrophysical parameters. Data measured by different probes are processed in a joint inversion procedure in order to estimate several model parameters with their confidence intervals. The optimal solution is obtained by fitting theoretical data to measured ones calculated by response equations. A novel inversion method for the estimation of petrophysical parameters including shale volume was suggested by Dobróka and Szabó (2011).

Factor analysis represents a multivariable statistical method that can be used to extract non-measurable background variables from large-scale geophysical data sets. A novel application can be found in Szabó et al. (2012) to process borehole data for the characterization of shallow formations. Factor analysis can be applied to well-logging data for the determination of shale content of shaly sand reservoirs. Szabó (2011) found strong correlation between the log of one derived variable (factor) and the shale volume estimated by inversion data processing. Based on this relationship a non-linear formula for shale volume estimation was introduced that gave consistent results in different areas of the Pannonian Basin, Hungary. The method can be considered as an independent approach for shale volume estimation, which exploits information inherent in all types of well logs being sensitive to the presence of shale. In this paper, the connection between factor scores and shale content is studied and compared for three different well-logging data sets originated from Hungary and the United States of America. The statistical research revealed that the shale indicator gives consistent results to wireline logging data originating from overseas areas, too.

Shale volume estimation method

Factor analysis is the generalization of principal component analysis, a statistical method of which basic principles are detailed in the paper of Lawley and Maxwell (1962). Consider data matrix **D**, where D_{ij} element represents the datum measured by the *j*-th well-logging instrument in the *i*-th depth. The size of matrix **D** is *N*-by-*M*, where *N* is the total number of measuring (depth) points in the logged interval and *M* is the number of probe types. The factor analysis model decomposes matrix **D** as

$$\mathbf{D} = \mathbf{F}\mathbf{L}^{\mathrm{T}} + \mathbf{E} \tag{1}$$

where **F** denotes the *N*-by-*a* matrix of factor scores, **L** is the *M*-by-*a* matrix of factor loadings and **E** is the *N*-by-*M* matrix of residuals (superscript *T* denotes the transpose symbol). The number of factors is less than that of the original variables ($a<M$). The *j*-th column of **F** represents the values of the *j*-th new variable (uncorrelated factor) computed for different measuring points representing a new well log that is called a factor log. Matrix **L** represents the weights of the original variables (well-logging data) on the resultant factors. Assuming that factors are linearly independent $\mathbf{F}^{\mathrm{T}}\mathbf{F}/N = \mathbf{I}$ (where **I** is the identity matrix), Eq.(1) leads to

$$\mathbf{R} = \mathbf{L}\mathbf{L}^{\mathrm{T}} + \mathbf{\Psi}\,, \tag{2}$$

where $\mathbf{R}=\mathbf{D}^{\mathbf{T}}\mathbf{D}/N$ is the correlation matrix of the standardized original variables, and $\boldsymbol{\psi}=\mathbf{E}^{\mathbf{T}}\mathbf{E}/N$ is the diagonal matrix of specific variances. For the determination of factor loadings an approximate solution was suggested by Jöreskog (2007). The factor scores can be estimated by the maximum likelihood method to which a linear estimation was given by Bartlett (1955)

$$\mathbf{F} = \left(\mathbf{L}^{\mathrm{T}}\mathbf{\Psi}^{-1}\mathbf{L}\right)^{-1}\mathbf{L}^{\mathrm{T}}\mathbf{\Psi}^{-1}\mathbf{D}\,. \tag{3}$$

Szabó (2011) found strong correlation between the scores of the first factor (1^{st} column of matrix **F**) and shale volume estimated by inverse modeling in several deep wells. It was inferred that factor analysis is applicable to extract the shale content from wireline logging data. Assuming a non-linear connection, the following formula was introduced for shale volume estimation

$$V_{sh} = ae^{bF'_1}, \tag{4}$$

where F'_1 is the value of the scaled first factor in the given depth, a and b represent regression constants. For the sake of comparability, raw factor scores (F_1) are transformed into the same interval by using the following formula

$$F'_1 = F'_{1,min} + \frac{F'_{1,max} - F'_{1,min}}{F_{1,max} - F_{1,min}}\left(F_1 - F_{1,min}\right), \tag{5}$$

where $F_{1,min}$ and $F_{1,max}$ are the minimum and maximum value of the original factor log, respectively; $F'_{1,min}$ and $F'_{1,max}$ are the desired lower and upper limit of the scaled factor, respectively. Since shale volume ranges between 0 and 100%, the same interval for the factor scores was chosen.

Comparative study

Three prospecting sites were selected for the study. One of the well-logging data sets was collected from an Upper Miocene unconsolidated sand complex interbedded with clay and silt layers from the Pannonian Basin, Hungary (Area-1). The sequence has got high and medium porosity and is saturated with water and gas. Well logs included caliper CAL (inch), spontaneous potential SP (mV), natural gamma-ray GR (API), compensated neutron CN (%), density DEN (g/cm^3), acoustic traveltime AT (μs/ft), microlaterolog RMLL (ohmm), shallow resistivity RS (ohmm), deep resistivity RD (ohmm) logs. The second data set was chosen from the Permian Basin of Texas, USA (Asquith and Krygowski 2004). The Pennsylvanian Atoka sandstone is characterized by coarse grains with large porosity and permeability (Area-2). Pores were filled with water, gas and condensate. Data set consisted of caliper CALI (inch), spontaneous potential SP (mV), natural gamma-ray GR (API), neutron porosity NPHI (v/v), density porosity DPHI (v/v), deep induction ILD (ohmm), medium induction ILM (ohmm), shallow resistivity LL8 (ohmm) logs. The third data set originated from a Cretaceous sequence called Torok Formation situated in Alaska, USA (Area-3). The formation represents turbiditic sandstones with observable petroleum potential (Gryc 1988). Well logs used as input for factor analysis were CAL (inch), spontaneous potential SP (mV), natural gamma-ray GR (API), neutron porosity NPHI (%), density RHOB (g/cm^3), acoustic interval time DT (μs/ft), micro-spherically focused log MSFL (ohmm), shallow laterolog LLS (ohmm), deep laterolog LLD (ohmm).

Factor analysis was tested on the data sets, separately. The number of factors was specified previously. Two uncorrelated factors were extracted in all cases, because the major part of the variance of measured variables could be explained by two factors. The estimated factor loadings are shown in Table 1. The biggest weights on the first factor were caused by the lithological and resistivity logs. Therewith the contribution of other log types was considerable, too. For making the results comparable, the factors were rotated and scaled by using Eq.(5). The exponential relationships between the factor scores and shale volumes based on Eq.(4) are illustrated in Figure 1. The Spearman's rank correlation coefficient (0.98) showed strong non-linear relationships. Exponent a and b in Eq.(4) were specified resulting in the shale volume estimation formula

$$V_{sh} = 2.76e^{0.037\,F'_1}. \tag{6}$$

The logs of estimated shale volumes were compared to the results obtained by independent procedures of log analysis. In Area-1 shale volumes were estimated previously by local (point-by-point) inversion by using a weighted least squares method (Szabó 2011). In Area-2, a deterministic procedure based on more well logs was used (Asquith and Krygowski 2004). In Area-3, the GR log served as a shale indicator (Larionov 1969). The shale volume logs estimated by traditional interpretation and factor analysis can be seen in Figure 2. The overall RMS between the two estimates was 8.2%, 9.6%, 10.4% for Areas 1-3, respectively.

Area-1 Pannonian Sand (Hungary)			**Area-2 Atoka Sandstone (USA)**			**Area-3 Torok Formation (USA)**		
Well log	**Factor 1**	**Factor 2**	**Well log**	**Factor 1**	**Factor 2**	**Well log**	**Factor 1**	**Factor 2**
CAL	0.46	-0.02	CALI	0.47	0.79	CAL	0.67	0.27
CN	0.91	0.25	NPHI	0.48	0.82	NPHI	0.52	0.71
DEN	0.79	-0.60	DPHI	0.09	0.78	RHOB	0.11	0.90
AT	0.12	0.79	-	-	-	DT	0.41	0.27
GR	0.94	-0.04	GR	0.76	0.56	GR	0.72	-0.09
RD	-0.68	-0.06	ILD	-0.93	-0.31	LLD	-0.73	-0.06
RMLL	-0.72	0.57	LL8	-0.89	-0.26	MSFL	-0.73	-0.06
RS	-0.18	-0.01	ILM	-0.75	-0.30	LLS	-0.64	-0.05
SP	-0.83	-0.15	SP	0.75	-0.08	SP	0.48	0.37

Table 1 *Factor loadings for the processed well logs.*

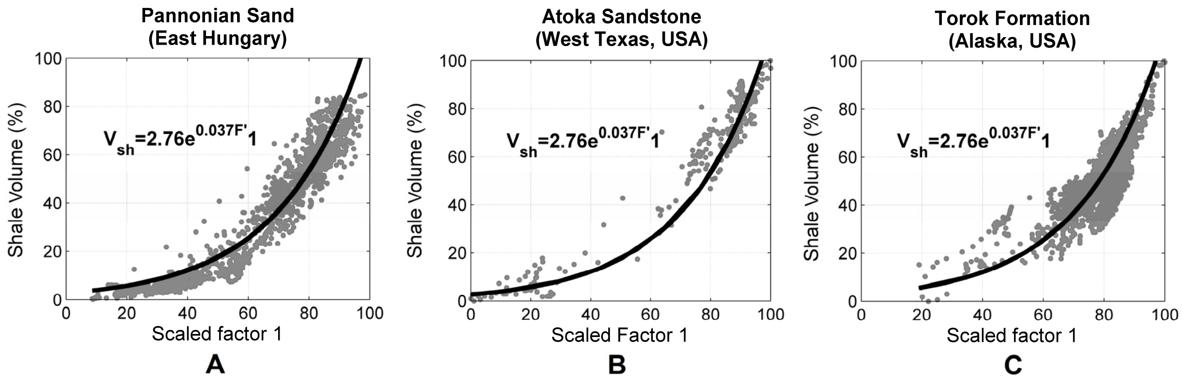

Figure 1 *Exponential relationship between the scaled factor and shale content for Area-1 (A), Area-2 (B), Area-3 (C).*

Conclusions

Factor analysis is applicable to extract the shale content from wireline logging data as basic lithological information about hydrocarbon reservoirs. Factor analysis was tested on three data sets from different locations. The results were confirmed by estimates derived from independent inverse/deterministic modeling. In this stage of research, we assume an exponential relationship between the first factor and shale volume in sedimentary geological environments. The relationship proves to be straight representing a strong correlation between the two variables. Parameter a and b in Eq.(4) seems to be approximately independent of the location of the prospecting area. For further research more data sets originating from other areas are required to be processed. The benefits of the method in a given area are: this method may confirm or bring new information about the values of shale volumes estimated from different sources; derived quantities as effective porosity, hydrocarbon saturation, effective permeability etc. can be updated; shale volume information can be obtained more accurately and reliability, because the information of all log types available at the site is used simultaneously. The estimation of shale volume from an independent source is an important issue for the success of inverse modeling, too. By giving a reliable estimate for shale volume by factor analysis, one can decrease the number of unknowns of the well-logging inverse problem. Having more data for the rest of petrophysical unknowns, a more accurate estimate for the inversion model can be given.

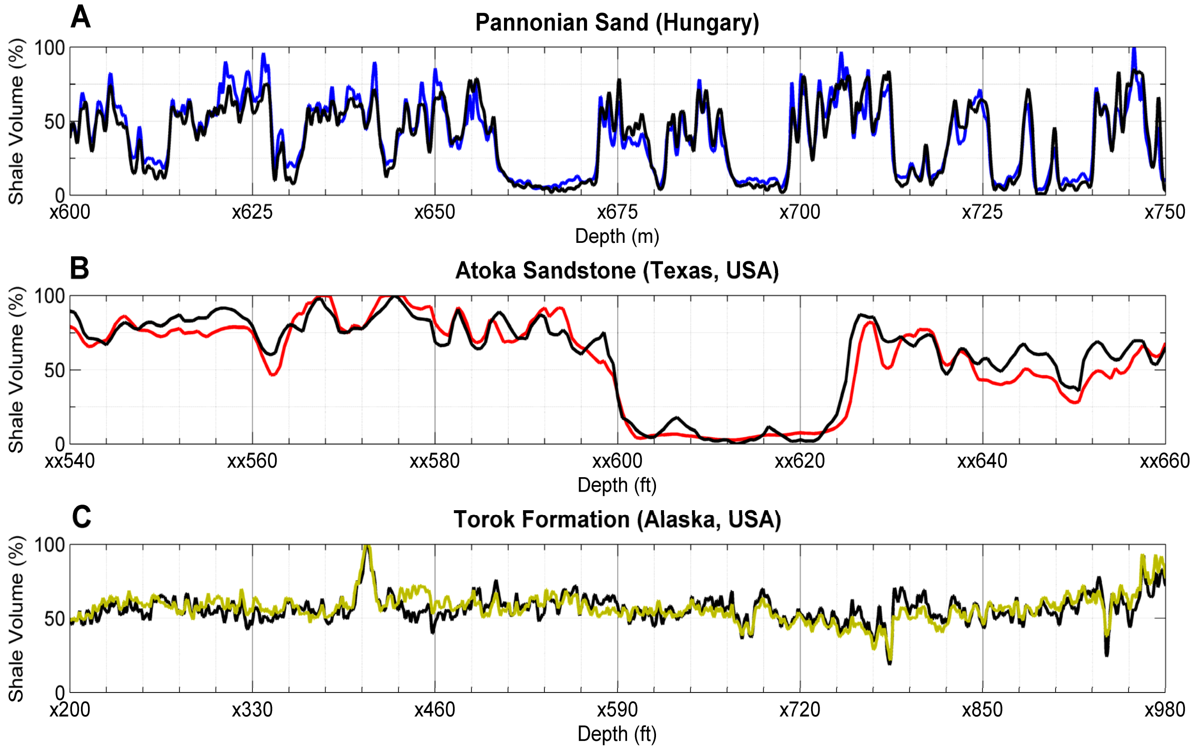

Figure 2 *Shale volume estimated by factor analysis (blue curve in A, red curve in B, green curve in C) and independent sources (black curves in A, B, C).*

Acknowledgements

The described work was carried out as part of the TÁMOP-4.2.2.A-11/1/KONV-2012-0005 project in the framework of the New Széchenyi Plan. The realization of this project is supported by the European Union, co-financed by the European Social Fund. The first author thanks to the support of the János Bolyai Research Fellowship of the Hungarian Academy of Sciences. As a member of the MTA-ME Applied Geoscience Research Group the second author is grateful for the support of the Hungarian Academy of Sciences.

References

Asquith, G. and Krygowski, D.A. [2004] *Basic well log analysis*. 2nd edition. AAPG Methods in Exploration Series, No. 16, Tulsa.
Bartlett, M.S. [1955] *An Introduction to Stochastic Processes with Special Reference to Methods and Applications.* Cambridge University Press.
Dobróka, M. and Szabó, N.P. [2011] Interval inversion of well-logging data for objective determination of textural parameters. *Acta Geophysica*, **59**, 907-934.
Gryc, G. (Ed.) [1988] Geology and exploration of the National Petroleum Reserve in Alaska, 1974 to 1982, *U.S.G S. Professional Paper*, 1399, 322 - 324.
Jöreskog, K.G. [2007] Factor analysis and its extensions. In: Cudeck, R. and MacCallum, R.C. (Eds.) *Factor analysis at 100*, *Historical developments and future directions*. Lawrence Erlbaum Associates.
Larionov, V.V. [1969] *Radiometry of boreholes (in Russian)*. Nedra, Moscow.
Lawley, D.N. and Maxwell, A.E. [1962] Factor analysis as a statistical method. *The Statistician,* **12**, 209-229.
Szabó, N.P. [2011] Shale volume estimation based on the factor analysis of well-logging data. *Acta Geophysica*, **59**, 935-953.
Szabó, N.P., Dobróka, M. and Drahos, D. [2012] Factor analysis of engineering geophysical sounding data for water saturation estimation in shallow formations. *Geophysics*, **77**, WA35-WA44.