Exploration Geophysics, Remote Sensing and Environment XXX.2 (2023)

DOI: 10.26345/EGRSE-059-23-205

COMPONENT ANALYSIS AS A NEW EFFICIENT INTERPRETATION METHOD FOR WELL-LOGGING DATA

KOMPONENTNÍ ANALÝZA JAKO NOVÁ EFEKTIVNÍ INTERPRETACE KAROTÁŽNÍCH DAT

František Ryšavý[1]

Abstract

This paper presents a guidebook on how to apply a method of the Principal Component Analysis on well-logging statistical sets. The algorithm of the problem is based on principles of the Matrix Calculus respected with all spreadsheets. It is about the method of Principal Component Analysis presenting a closed system of linear equations. For calculation, you can use processor Microsoft Excel and its implements for the solution of equations. Both ways of analysis you can solve there; both after characteristics and after the points of depth. Each of the variants yields different information. The main is analysis after characteristics which can depict the interpreted factors in form of continuous curves being like classical ones of deterministic models. However, the curves of the stochastic model we have grasp like curves of probability for the type of rock/suspense. The analysis after the points of depth is then next information about relations between characteristics; it is, however, important for well geological identification of factors.

Abstrakt

Tato práce je jakýmsi odborným návodem, jak aplikovat metodu komponentní analýzy na statistické soubory karotážních dat. Algoritmus úlohy je založen na principech maticového počtu, které respektují všechny tabulkové procesory. Jedná se o metodu hlavních komponent, která představuje uzavřený systém lineárních rovnic. Pro výpočet byl použit procesor Microsoft Excel a jeho aplikace pro řešení rovnic. Lze řešit oba způsoby analýzy, jak podle registrovaných charakteristik, tak podle hloubkových bodů. Každá z variant poskytuje různé informace. Hlavní je analýza podle charakteristik, která zobrazuje interpretované faktory formou spojitých křivek, které se podobají klasickým křivkám deterministických modelů. Avšak křivky stochastického modelu musíme spíše chápat jako pravděpodobnostní křivky pro předpokládaný typ horniny/látky. Analýza podle hloubkových bodů je potom dodatečnou informací o vztazích mezi charakteristikami; to je ale velmi důležité pro správnou geologickou identifikaci faktorů.

Keywords

Principal Component Analysis, Factor Analysis, standardized data, analysis after characteristics, analysis after the depth points

Klíčová slova

Analýza hlavních komponent, faktorová analýza, standardizovaná data, analýza podle charakteristik, analýza podle hloubkových bodů

1. Introduction – comparison between Factor and Component Analyses

Factor Analysis and Principal Component Analysis are both the statistical methods determined for interpreting multidimensional sets. In lot of cases, they are similar, however, in other things are different. Factor Analysis explains the dispersion of the studied variables with the help of a lower number of latent variables denoted as factors. We measure something that is not measurable directly. Factor Analysis began used at first in Social Sciences in psychology. Later it was enlarged too in sociology, economy and too in Nature Sciences. That holds for the world round the Czech Republic. It is surprising, that in the Czech Republic it was a bit later. Majority works was in biology. An analogous situation was for Principal Component Analysis. I was interested in what is situation in Czech geophysics, especially in Well-Logging. However, when I tracked back works in professional geophysical and geological journals in the Czech Republic, I was not successful. Nevertheless, an expansion of new measuring methods and the discovery of Artificial Intelligence present wide possibilities for new ways interpreting of well-logging data in geophysics.

There exist two main orientations of Factor Analysis; they are Explorative Factor Analysis denoted as EFA and Confirmatory Factor Analysis denoted as CFA. Explorative Factor Analysis tends to find the latent factors which are just for geological prospection important. In contrary Confirmatory Factor, Analysis studies put certain restrictions on items of linear systems of equations. Geological prospection is a very suitable method for EFA.

Explorative Factor Analysis describes every studied variable as a linear combination of influence factors through matrix operations. It is either the correlation matrix or the covariance matrix. The number of input factors can be various; however, the sense of analysis is to find the number of acceptable factors as low as possible. This aim is close to Component Analysis, but both methods have a different system of linear equations.

Method of Principal Component Analysis denoted as PCA studies too latent variables of higher level denoted as factors. However, the system of linear equations is a bit different in comparison to the system equations for Explorative Factor Analysis. Nevertheless, the large similarity of both equation systems tends in older literature to the contention that Principal Component Analysis belongs to Factor Analysis in sensu extenso. On the contrary younger authors, classify Principal Component Analysis as independent.

The method of PCA tends to reduce the number of the input variables in a way to have as well as possible explained dispersion of input variables. Factor Analysis tends to an explanation of correlation input variables. Method PCA has an advantage that offers clearly defined factor solution. Just this is why the method belongs among those favoured. Opponents of the method dispute that the before method PCA is not the one that best explains correlations between factors. Methods PCA and EFA and the comparison between them you find in the literature

DMITRIEV, et al. (1982), MELOUN, MILITKÝ (2004) and (2012), FABRIGAR (2012), FINCH (2019), JOLLIFFE (2002) and MA. Y. (2015). As for Well-Logging published papers, significant and important works are the works of NICULESCU, B.M., ANDREI, G., and CIUPERCĂ, C.L. (2016) and NICULESCU, B.M., ANDREI, G. (2016) for gas boreholes in Romania. PCA serves to expansion added information from the boreholes results gotten through standard interpretation Well Logs. Method PCA offers comparison its results with core analyses, production tests and mud low gravity solids. Both papers follow rife figures Well Logs common with PCA Logs. More in detail in the mentioned papers.

Principle Component Analysis and Factor Analysis present a highly effective ways of statistical data evaluation. It is strange enough that such analyses were not yet imply in the Czech Republic for Well-Logging, even if well-logging data present large statistic sets reflecting relations between various physical characteristics and the borehole depth. In comparison to the deterministic models, where all relations are defined; the characteristic and its relations, the stochastic models offer a probable relationship of factors to the borehole depth. These factors can copy certain the deterministic curves as total porosity, shaliness or the next ones, but they can have completely different courses as well. The stochastic models use probability and interpretation of results; it depends on the well geological experience of one who such interpretation makes. A certain indeterminateness of such models – can be the reason there exists misgiving to methods of factor analysis. Simply speaking, the stochastic models are not definite. However, they have their advantage too; they are making it possible to interpret in cased boreholes the curves closed to deterministic ones as total porosity or shaliness are.

I should like to say in advance, the crucial problem of interpretation is always the identification of factors. Not always all factors intelligibly you can identify; it is not a rare case that although factors have defined relations to every depth point, nevertheless, they have no geological identification. It needs to say that geological identification of factors presents a big problem; despite that a new possibility how to get quite current information, is attractive.

2. Model´s equations of Explorative Factor Analysis and Principal Component Analysis

It is about two models. The first is the model of method EFA; the second is model of method PCA. The first model is the open one. Such enables us to reach more variants of solutions. An interpreter separates the final variant. The opened model we denote as multivariable.

It is possible to describe that by formula representing the system of equations:

$$X_i = \sum_{j=1}^{m} a_{ij} \times Y_j + e_i, \qquad (1)$$
$$i = 1,2,\ldots,n; \;\; m \langle n,$$

where a_{ij} = the weight of j-th vector of directional cosine for the i-th characteristic X_i,

Y_j = the j-th vector of cosine,

e_i = the resting term of errors depending on X_i, and

X_i = the random i-th characteristic X_i.

The model, opened, makes it possible to get lower number of factors from *n* to *m* very rapidly and in every case.

The model of PCA is a closed one. Such a model yields an unambiguous solution on the condition that the scales of characteristics are fix. The registered characteristics you must standardize to get dimensionless data. The above model you can record as a system of equations:

$$X_i = \sum_{j=1}^{m} a_{ij} \times Y_j \,, \qquad i = 1,2,\ldots,n; \;\; m \le n. \tag{2}$$

The mentioned model uses all original factors; any of them, you cannot exclude in advance, as latent, $m = n$. However, the number of latent factors is more likely lower, $m < n$. Method of PCA has no error term, e_i, which gives bigger freedom to the system; this is why the opened one is.

In the geometric sense, it is the transformation of the n-dimensional ellipsoid from the original system of coordinates into the new orthogonal system of coordinates. The new orthogonal system can be again the n-dimensional, however, very often has a lower dimension and we receive so the m-dimensional one. The axes of such ellipsoid we characterize by vectors of directional cosines $\boldsymbol{a_i}$ and just these form the matrix of terms denoted as $\boldsymbol{a_{ij}}$. Their handheld calculation would be sure difficult, however, in the age of computers and matrix calculus it is not any problem.

3. Principal Component Analysis

I should say why I preferred Component Analysis to Factor Analysis in sensu stricto. I studied Russian works, namely, BELONIN et al. (1982) and AJVAZJAN et al. (1981). I must state this method is more frequently used in geophysics for the unambiguous solutions. However, the registered characteristics need standardisation. The mentioned work BELONIN et al. (1982) I selected like the methodical guide on how to apply method PCA on sets of well-logging data.

I needed to have well-logging data. Such data I did not have any. However, just in the mentioned work BELONIN et al. (1982) there were tables of the evaluated data yet. They were the table of input data of surface samples {14×5} and the tables of calculated vectors of the directional cosines. Therefore, I kept the numeric data of the input table; however, I changed the registered characteristics. For example, I replaced the thickness of the Jurassic cover with the mass activity of radionuclide [μBq/g] and the plain of deposit by resistivity in [Ωm] and made the next changes. For all 14 points of observations, I added the depth of the borehole as it is usual for well-logging sets.

The next problem I had for the counting software Excel 2007 through I wanted to check whether algorithm PCA is suitable for Well Log Analysis and following geological interpretation. This software is now outdated yet but ten years ago it was not. The software was not, of course, satisfactory for large data sets. I supposed that software is always only an instrument for testing suitability and whether algorithm is acceptable for Well Log Analysis and following geological interpretation. And that is possible to reach that too in the conditions you have less ideal software and small data sets. If you reach expected destination, you could expect that the same will be valid too for large data sets evaluated by more perfect software such as MATLAB or PYTHON.

All adjustments made me possible to absolve all processes of calculation and to compare the output numeric data with those of BELONIN et al. (1982). I tried to have the input numeric data in coincidence with the usual well-logging characteristics. The final interpretation of factors can seem to someone like a bit sophisticated geological surroundings; nevertheless, important, is to see how the method to apply on larger well-logging sets. And if this paper did that, you achieved aim. Information about matrix calculation I had drawn from SIGORSKIJ (1975) and REKTORYS (1968).

4. Evaluation of input data

As the well-logging sets are large, it is possible to suppose that they follow a normal distribution. However, the input characteristics need standardization in dimensionless units. What is important is the matrix of correlation of these units is identical to the matrix of correlation for dimensional characteristics. Tab. 1 presents input data of statistical sets. There are five physical characteristics registered like a function of the borehole depth. The first column is resistivity in [Ωm], the second one is magnetic susceptibility in [SI] and the third one can present either the mass activity of radionuclide in [μBq/g] or the apparent specific activity of rocks in [μg ^{238}U-eq/t]. The fourth informs about the bulk density of rocks in [g/cm^3] and, finally, the fifth is the apparent electric susceptibility that is dimensionless. In this table, there are too the means of all characteristics and their standard deviations.

The mean is denoted as $\overline{x}$ and the standard deviation as *s*. The formula of standardization is following:

$$X_{ij}^{*} = \frac{\left(X_{ij} - \overline{x}\right)}{s_j}, \tag{3}$$

$i = 1,2,...,14$ and $j = 1,2,..,5$.

Results of standardization you will find out in Tab. 2. The standardized set presents input data for the matrix of correlation. Tab. 2 presents the field of matrix $\mathbf{X}^{*}$. Dimensions of this matrix are {14×5}. Besides that, we need to have the transposed matrix denoted as $\mathbf{X}^{*\mathbf{T}}$. Its dimensions are {5×14}. It is because we have two various matrices of correlation. The first is the correlation matrix of characteristics denoted as $\mathbf{R_1}$ and the second is the correlation matrix of the depth points as $\mathbf{R_2}$. The matrix $\mathbf{R_1}$ has dimensions {5×5}, while, the matrix $\mathbf{R_2}$ has bigger dimensions {14×14}. In this case, I must note that for digital registration 10 points/m of depth, if we register an interval of 500 m of the borehole depth, we shall receive 5000 of the depth points. That means that matrix $\mathbf{R_2}$ will be extraordinarily large; its dimensions would be {5000×5000}.

The correlation matrix $\mathbf{R}$ with the help of the covariance matrix denoted as $\mathbf{K_1}$:

$$\mathbf{K_1} = \frac{\mathbf{X}^{*\mathbf{T}} \times \mathbf{X}^{*}}{\left(n - 1\right)}, \tag{4}$$

where n = 14; i.e., the number of objects (the depth points).

The matrix of correlation of the depth points, which you denote as $\mathbf{R_2}$, you can define by the formula for the covariance matrix denoted as $\mathbf{K_2}$:

$$\mathbf{K_2} = \frac{\mathbf{X}^* \times \mathbf{X}^{*\mathbf{T}}}{(n-1)}, \tag{5}$$

where n = 14, i.e., number of the depth points.

The terms of both covariance matrices are symmetric after the main diagonal. It holds that $K_{ij} = K_{ji}$. Transformation of terms in the covariance matrix on terms belonging to the correlation matrix realized with the help of relation:

$$R_{ij} = \frac{K_{ij}}{\sqrt{s_i \times s_j}}. \tag{6}$$

Both correlation matrices have all terms in the main diagonal equal to 1. The transformation from the covariance matrix to the correlation matrix you do due to the computer. All analyses you can make with the help of the table processor Microsoft Excel. The correlation matrix $\mathbf{R_1}$ is in Tab. 3, while, in Tab. 4 there is the correlation matrix $\mathbf{R_2}$. Both matrices you will receive due to succession **Implements → Data Analysis → Correlation**. For matrix $\mathbf{R_1}$ you must denote “columns” and for matrix $\mathbf{R_2}$ “rows “.

5. Analysis after characteristics

This is the analysis of the matrix $\mathbf{R_1}$. For simplification, we will use identity $\mathbf{R} = \mathbf{R_1}$. Owing to the mentioned matrix we can define the characteristic matrix $\mathbf{F}(\lambda_j)$. The formula in general form for that is the following:

$$\mathbf{F}(\lambda_j) = \mathbf{R} - \lambda_j \times \mathbf{I}, \tag{7}$$

where **I** = the unit matrix, and

λ_j = the root of the characteristic equation.

The matrix $\mathbf{F}(\lambda_j)$ you can record by terms $(1 - \lambda_j)$ in the main diagonal. All resting terms are the same as it is in the correlation matrix $\mathbf{R_1}$. The determinant of the characteristic matrix is polynomial; because matrix $\mathbf{F}(\lambda_j)$ has dimension $\{5\times5\}$ as matrix $\mathbf{R_1}$ is, we obtain the polynomial of the fifth degree. So, if we lay the condition that the determinant equals zero, we shall get an equation of the fifth degree. You must find different roots of the characteristic equation $\lambda_1, \lambda_2 \ldots \lambda_5$. All roots are real and after substitution into the characteristic matrix, you receive five matrices $\mathbf{F}(\lambda_1)$, $\mathbf{F}(\lambda_2)$ ... $\mathbf{F}(\lambda_5)$. In mathematical processes, I followed AJVAZJAN et al. (1981).

For a real solution, we use again Microsoft Excel. The first step is I must select a cell and into its content, I put the number 5. This cell after calculation will offer the value of the roots λ_j. The number 5 is not select at random. It is by the condition of solutions:

$$Tr\,\mathbf{R_1} = \sum_{j=1}^{5} \lambda_j = 5, \tag{8}$$

where Tr $\mathbf{R_1}$ = trace of the correlation matrix, i.e., the sum of all roots λ_j.

Condition (8) is all clear; $n = 5$. Tr $\mathbf{R_1}$ presents the upper limit of values for the roots, because $\lambda_j <$ Tr $\mathbf{R_1}$. Any of the roots are not higher than number 5. The second step is forming the field of the characteristic matrix. All terms of the main diagonal will be $(1 - \lambda_j)$; starting position is $\lambda_j = 5$. If the calculation finishes, we shall obtain the field of matrices $\mathbf{F}(\lambda_1)$, $\mathbf{F}(\lambda_2)$ ... $\mathbf{F}(\lambda_5)$. We begin with the creation of the characteristic matrix $\mathbf{F}(\lambda_j)$. It means to form the before field like the matrix. And we select, too, the next cell which will be the determinant of the characteristic matrix. As the starting value of $\lambda_j = 5$, we get the input value of the determinant equals 5, too. Now, we must use succession **Implements → Finding Solution**. The activated cell is the cell of determinant. The aim value is zero. The changing cell is the cell having $\lambda_j = 5$. The result will be that $\lambda_1 = 2.567$.

Tab. 1 The input characteristics as a function of depth

i	h [m]	R [Ωm]	$\chi\times10^6$[SI]	σ[μBq/g]	ρ[g/cm³]	ε
1	350	22.06	50	0	2.5	5.7
2	350.5	70	80	1.5	0.7	1.5
3	351	9.08	140	0	0.28	1.46
4	351.5	28.9	160	0.9	1.5	3.5
5	352	59.6	150	3	3	3.53
6	352.5	136	55	50.7	8.58	6.52
7	353	107.27	130	0	0	2.55
8	353.5	63	110	50.4	4.04	4.24
9	354	89.25	58	4.7	3.54	1.57
10	354.5	27.63	148	1.5	2.68	4.37
11	355	200	40	1.9	3.2	2.3
12	355.5	68	130	9.6	3.06	2.05
13	356	30	80	2.5	3.25	4.2
14	356.5	81.95	20	1.5	3.25	3.42
$\bar{x}$		70.91	96.5	9.16	2.83	3.35
s		49.57	45.37	17.07	2.01	1.51

Tab. 2 The standardized characteristics depicted as $\mathbf{X}^*$ as a field both the depth and characteristics

i	h [m]	R	$\chi\times10^6$	σ	ρ	ε
1	350	-0.9855	-1.0248	-0.5366	-0.1628	1.5592
2	350.5	-0.0184	-0.3637	-0.4487	-1.0584	-1.2283
3	351	-1.2473	0.9587	-0.5366	-1.2674	-1.2549
4	351.5	-0.8475	1.3995	-0.4838	-0.6603	0.0991
5	352	-0.2282	1.1791	-0.3608	0.0860	0.1190
6	352.5	1.3131	-0.9146	2.4343	2.8624	2.1035
7	353	0.7335	0.7383	-0.5366	-1.4067	-0.5314
8	353.5	-0.1596	0.2975	2.4167	0.6035	0.5902
9	354	0.3700	-0.8485	-0.2612	0.3547	-1.1819
10	354.5	-0.8731	1.1350	-0.4487	-0.0732	0.6765
11	355	2.6041	-1.2452	-0.4252	0.1855	-0.6974
12	355.5	-0.0587	0.7383	0.0260	0.1159	-0.8633
13	356	-0.8253	-0.3637	-0.3901	0.2104	0.5637
14	356.5	0.2227	-1.6860	-0.4487	0.2104	0.0460

The next roots you will find with the help of consecutive division. The determinant you divide by the difference in the cell content (λ_j–λ_1). Then we adjust again the starting value for λ_j = 5. We call **Finding Solution** and here is the second root λ_2 = 1.314. The content of the cell, where the determinant divided by (λ_j - λ_1) is, you must divide by (λ_j – λ_2); starting condition is again λ_j = 5. All process is cyclically repeated. The next roots are λ_3 = 0.674, λ_4 = 0.290 and, finally, λ_5 = 0.156. We must do successive sums of roots until the condition (8) is filled. It holds that (2.567+1.314+0.674+0.290+0.156) = 5.001. The solution can have only real roots!

The used implements of Microsoft Excel work in way of finding iterations. Therefore, the solution can have only real roots. If there are complex roots, the implement will announce – no solution found. Fortunately, the roots of the characteristic equation are always real. For each of roots, we reach the single matrix. The matrices F (λ_1), **F** (λ_2) ...**F** (λ_5) you need for the process of orthogonalization. We must orthogonalize vectors of the modal matrix **a**. This is the next crucial step.

Here are the conditions for orthogonalization:

$$\mathbf{F}(\lambda_j)\times\mathbf{a_j}=0,\text{and} \tag{9}$$

$$\sum_{i=1}^{n} a_{ij}^2=1. \tag{10}$$

We use again processor Microsoft Excel. The first is to form the matrix from the field of a matrix. We shall get matrix **F** (λ_j) = **F** (λ_1) having dimensions {5×5}. The next is to form the field of matrix/vector $\mathbf{a_j}$ = $\mathbf{a_1}$. As n = 5 we have 5 cells and the content of each of them will be equal to 1. The values equal to 1 are the starting ones.

Further, we create the column matrix $\mathbf{a_j}$; j = 1, 2, 5. Its dimensions are {5×1}. Multiplication of two matrices after condition (9), {5×5} × {5×1} ≡ {5×1}, presents the matrix having dimensions {5×1}. It is still to form the cell where will be the second condition of orthogonalization there, i.e., condition (10). As the starting values of all cells for matrix $\mathbf{a_j}$ are 1, the content of the cell will be value 5.

Now, we use succession **Implements → Solutionist**. We must activate an arbitrary cell of matrix **F** (λ_j) × $\mathbf{a_j}$. It is usually the first cell. The value of this cell must be zero. The changing cells are the cells of the field of matrix $\mathbf{a_j}$. Attention, they are denoted and detached by a semi-colon. Now, I shall fill limiting conditions. Here they are:

$$\sum_{i=1}^{5} \mathrm{a}_{ij}^2=1,\quad \mathrm{a}_{11}=\mathrm{a}_{21}=\mathrm{a}_{31}=\mathrm{a}_{41}=\mathrm{a}_{51}=0\,.$$

The conditions you can express like cells. What is important is that condition a_{11} = 0, doubles the condition that the activated cell is zero. However, the computer finds ways how to minimalize the numeric content of all cells and therefore can select the cell which is best for that. It is time to adjust the parameters of The **Solutionist**. We press the button **Possibilities**. **Precision** will be lower, 1E-03. **Maximal Time** and **Iterations** can be higher. **Toleranc**e will be 1%. **Extrapolation** is **Quadratic because** the model is nonlinear. The **Derivation** is **Correct,** and the **Method** is **Associated**. Then **Solutionist** announces that found the solution. It is convenient to have the cells of the matrix **F** (λ_j) × $\mathbf{a}_j$ together with the cell of condition (10) adjusted on three decimal places. I would remark that in case of higher **Precision** calculation will be more correct, but **Solutionist** will announce - no solution found out.

Tab. 3 The correlation matrix characteristics denoted as R_1

	R [Ωm]	$\chi\times10^6$[SI]	σ[μBq/g]	ρ[g/cm^3]	E
R [Ωm]	1	-0.501	0.252	0.391	-0.074
$\chi\times10^6$[SI]	-0.501	1	-0.118	-0.423	-0.178
σ[μBq/g]	0.252	-0.118	1	0.743	0.512
ρ[g/cm^3]	0.391	-0.423	0.743	1	0.653
ε	-0.074	-0.178	0.512	0.653	1

Tab. 4 The correlation matrix of denoted the depth points denoted as R_2

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1	-0.854	-0.486	-0.049	-0.169	0.464	-0.514	0.058	-0.544	0.317	-0.327	-0.860	0.885	0.412
2	-0.854	1	0.366	-0.004	-0.003	-0.508	0.789	-0.109	0.367	-0.417	0.502	0.548	-0.991	-0.308
3	-0.486	0.366	1	0.869	0.808	-0.846	0.522	0.051	-0.427	0.616	-0.570	0.762	-0.296	-0.996
4	-0.049	-0.004	0.869	1	0.931	-0.824	0.430	-0.165	-0.739	0.907	-0.714	0.435	0.109	-0.898
5	-0.169	-0.003	0.808	0.931	1	-0.828	0.389	-0.403	-0.503	0.861	-0.562	0.569	0.086	-0.815
6	0.464	-0.508	-0.846	-0.824	-0.828	1	-0.830	0.430	0.382	-0.554	0.215	-0.617	0.410	0.836
7	-0.514	0.789	0.522	0.430	0.389	-0.830	1	-0.441	-0.154	0.088	0.254	0.366	-0.701	-0.501
8	0.058	-0.109	0.051	-0.165	-0.403	0.430	-0.441	1	-0.092	-0.176	-0.412	-0.052	0.067	-0.072
9	-0.544	0.367	-0.427	-0.739	-0.503	0.382	-0.154	-0.092	1	-0.814	0.752	0.246	-0.483	0.503
10	0.317	-0.417	0.616	0.907	0.861	-0.554	0.088	-0.176	-0.814	1	-0.818	0.154	0.511	-0.665
11	-0.327	0.502	-0.570	-0.714	-0.562	0.215	0.254	-0.412	0.752	-0.818	1	-0.133	-0.553	0.625
12	-0.860	0.548	0.762	0.435	0.569	-0.617	0.366	-0.052	0.246	0.154	-0.133	1	-0.559	-0.705
13	0.885	-0.991	-0.296	0.109	0.086	0.410	-0.701	0.067	-0.483	0.511	-0.553	-0.559	1	0.232
14	0.412	-0.308	-0.996	-0.898	-0.815	0.836	-0.501	-0.072	0.503	-0.665	0.625	-0.705	0.232	1

The results for root λ_1 you copy like the column of vector $\mathbf{a}_1$. We active again the starting values for $\mathbf{a}_j$ and the cell of condition (10) will be again 5. The field of matrix **F** (λ_2) copies into matrix **F** (λ_j). And all process repeats periodically. So, we obtain vectors $\mathbf{a}_1, \mathbf{a}_2 \dots \mathbf{a}_5$. These vectors form the field of the modal matrix **a**. The above matrix has dimensions {5×5} and has characteristic properties: the sum of squared terms each row and each column, too, equals 1. I can obtain 5 factors; it holds that $m = n = 5$. However, what is important is which of the factors is more significant and which is less significant. About that there exists a criterion of weights. The field of the modal matrix **a** is in Tab. 5.

With the help of matrix, **a,** we can express all field of the matrix of directional cosines denoted as $\boldsymbol{\omega}$ having dimensions {5×5}. We use the formula as follows:

$$\omega_{ij} = a_{ij} \times \sqrt{\lambda_j}\,. \qquad (11)$$

It means all columns $\mathbf{a}_{i1}$ you must multiply by $\sqrt{\lambda_1}$, the columns $\mathbf{a}_{i2}$ by $\sqrt{\lambda_2}$ … You will get in this way the directional cosines $\boldsymbol{\omega}_1$, $\boldsymbol{\omega}_2$, $\boldsymbol{\omega}_3$, $\boldsymbol{\omega}_4$ and $\boldsymbol{\omega}_5$. They are dependent on the characteristics registered. There exist two important conditions.

For rows, it holds that,

$$\sum_{j=1}^{5} \omega_{ij}^2 = 1\,. \qquad (12)$$

Tab. 5 The modal matrix of multidimensional vectors* $\mathbf{a}_i$ *after characteristics

$\mathbf{a}_1$	$\mathbf{a}_2$	$\mathbf{a}_3$	$\mathbf{a}_4$	$\mathbf{a}_5$	
0.317	-0.640	0.413	0.540	0.180	1.0
-0.347	0.538	0.618	0.432	-0.164	1.0
0.497	0.268	0.503	-0.501	0.419	1.0
0.588	0.091	0.040	0.006	-0.803	1.0
0.432	0.471	-0.440	0.521	0.348	1.0
1.00	1.00	1.00	1.00	1.00	

Tab. 6 The matrix of the directional cosines* $\boldsymbol{\omega}$ *after characteristics

	$\boldsymbol{\omega}_1$	$\boldsymbol{\omega}_2$	$\boldsymbol{\omega}_3$	$\boldsymbol{\omega}_4$	$\boldsymbol{\omega}5$	
R [Ωm]	0.508	-0.733	0.339	0.291	0.071	1.00
$\chi \times 10^6$[SI]	-0.555	0.617	0.507	0.232	-0.065	1.00
σ[μBq/g]	0.796	0.307	0.413	-0.270	0.165	1.00
ρ [g/cm^3]	0.942	0.104	0.033	0.003	-0.317	1.00
ε	0.692	0.539	-0.361	0.280	0.137	1.00
λ_j	2.567	1.314	0.674	0.290	0.156	5.00
w [%]	51.3	26.3	13.5	5.8	3.1	100.0

For columns, it holds that,

$$\sum_{j=1}^{5} \omega_{ij}^{2} = \lambda_j \, . \qquad (13)$$

The common sum of both above sums must be Tr $\mathbf{R_1} = 5$. The field of matrix $\boldsymbol{\omega}$ is in Tab. 6. There are, too, the roots of the characteristic equation $\lambda_1, \lambda_2 \dots \lambda_5$. They are aligned from the highest, λ_1, to the lowest λ_5. And each of them has its weight. The weight is denoted as symbol *w* and defined like this:

$$w = \frac{\lambda_j}{\sum_{j=1}^{5} \lambda_j} \, . \qquad (14)$$

The weight records in percentage. The weight decides which factors can be statistically more significant or less significant. We successively sum weights of factors towards λ_5. If the sum is in 95%, it will be crucial. Tab. 6 informs us that factors $\boldsymbol{\omega_1}$, $\boldsymbol{\omega_2}$, $\boldsymbol{\omega_3}$ and $\boldsymbol{\omega_4}$ belong to more significant, while factor $\boldsymbol{\omega_5}$ less significant. The first four are in bold black type, while the second one in bold brown type.

It is still to decide the precision of calculation for directional cosines. The fundament of that is matrix $\boldsymbol{\omega}$. However, we need to have the transpose matrix $\boldsymbol{\omega}^{\mathbf{T}}$, as well. Both matrices will form two new matrices. The first is the correlation matrix $\mathbf{R}_1$.

$$\mathbf{R_1} = \boldsymbol{\omega} \times \boldsymbol{\omega}^{\mathbf{T}}, \text{for dimensions holds: } \mathbf{R_1} \equiv \{5\times5\} \times \{5\times5\} \equiv \{5\times5\}. \qquad (15)$$

The second is the fundamental matrix $\boldsymbol{\Lambda}$ depicted the roots of the characteristic equation.

$$\boldsymbol{\Lambda} = \boldsymbol{\omega}^{\mathbf{T}} \times \boldsymbol{\omega}, \text{ for dimensions holds: } \boldsymbol{\Lambda} \equiv \{5\times5\} \times \{5\times5\} \equiv \{5\times5\}. \qquad (16)$$

Now, we must adjust number of decimal places per the following conditions. For the correlation matrix, it holds that all terms of the main diagonal must be equal to 1. The fundamental matrix declares that the terms of the main diagonal are the roots of the characteristic equation; all resting terms must be zero, exactly. It holds again that Tr $\mathbf{R_1} = (2.57+1.31+0.67+0.29+0.16) = 5$. Both matrices are in Tab. 7 and Tab. 8. It is distinct that the precision of roots and the directional cosines are 1E-02, as well. The third decimal place is burden by error yet!

6. Interpretation after characteristics

From the point of geometry, we have a 5-dimensional ellipsoid having five vectors - axes F_1, F_2, F_3, F_4 and F_5. Each of these has five coordinates - characteristics: R, $\chi \times 10^6$, σ, ρ and ε. They present numerically by directional cosines. These present coefficients of correlation between input data X^* and output data Y. Calculation between X^* and Y realized like the matrix multiplication as follows:

$$\mathbf{Y} = \mathbf{X}^{*} \times \boldsymbol{\omega}. \qquad (17)$$

Tab. 7 The correlation matrix after cosines ω denoted as R_1 as multiplication of matrices $\omega \times \omega^T$

$$R_1 = \omega \times \omega^T = \begin{bmatrix} 1 & -0.50 & 0.25 & 0.39 & -0.07 \\ -0.50 & 1 & -0.12 & -0.42 & -0.18 \\ 0.25 & -0.12 & 1 & 0.74 & 0.51 \\ 0.39 & -0.42 & 0.74 & 1 & 0.65 \\ -0.07 & -0.18 & 0.51 & 0.65 & 1 \end{bmatrix}$$

Tab. 8 The fundamental matrix after cosines ω denoted as Λ as multiplication of matrices $\omega^T \times \omega$

$$\Lambda = \omega^T \times \omega = \begin{bmatrix} 2.57 & 0 & 0 & 0 & 0 \\ 0 & 1.31 & 0 & 0 & 0 \\ 0 & 0 & 0.67 & 0 & 0 \\ 0 & 0 & 0 & 0.29 & 0 \\ 0 & 0 & 0 & 0 & 0.16 \end{bmatrix}$$

Because matrix $\mathbf{X}^*$ has dimensions {14×5} and matrix **ω** presents dimensions {5×5} the final matrix **Y** has dimensions {14×5}. It means we reach values of factors F_1, F_2 … F_5 like the functions of the borehole depth. Geometrically, it is about a 5-dimensional ellipsoid having 5 vector axes and every axis has 14 coordinates – the depth points. The tabled values you can depict in form of continuous curves with the borehole depth. That depiction is close to the depiction of continuous well-logging curves, and it is very acceptable for interpretation. Even if such curves can remember, for instance, the curve of total porosity, we must keep in mind that the curves of stochastic models report relations between the registered characteristics and depth. And a small remark yet. The matrix denoted as Y form five vectors with 14 coordinates of the depth points. If we had matrix X^* with dimensions {5000×5}, the mentioned factors should have 5000 depth coordinates.

In Tab. 9 the field of matrix **Y** the field of matrix **Y** depicts by form of vector columns. This is the starting table for the depiction of continuous curves of factors F_1, F_2 … F_5 being in Fig. 1.

Tab. 9 Factors of characteristics as a function of the borehole depth

h [m]	F_1	F_2	F_3	F_4	F_5
350	0.22	0.57	-2.44	0.20	1.11
350.5	-0.78	-0.85	0.05	-1.09	0.74
351	-1.42	0.40	0.38	-1.21	-0.06
351.5	-0.84	1.00	0.25	0.81	-0.05
352	-0.35	0.65	0.49	1.17	-1.05
352.5	2.83	0.50	0.48	0.38	-0.41
353	-0.84	-0.52	0.81	1.30	1.85
353.5	1.03	1.08	1.34	-1.59	1.66
354	-0.01	-1.12	0.04	-1.21	-1.52
354.5	-0.40	1.19	-0.23	1.11	-0.60
355	0.53	-2.41	0.49	1.34	0.26
355.5	-0.35	0.04	1.01	-0.32	-1.30
356	0.02	0.45	-1.22	-0.21	-0.57
356.5	0.36	-0.98	-1.45	-0.66	-0.06

Tab. 10 Interpretation of factors after characteristics

Interpretation of factors	
F_1	Electrically non-conductive, highly polarized rock/ore having heavy radioactive minerals. The factor of solid substance.
F_2	Electrically conductive, highly polarized rock/ore having magnetic minerals. The factor of solid substance.
F_3	Electrically non-conductive rock/ore having ore minerals being neither magnetic nor radioactive. The factor of solid substance.
F_4	Electrically non-conductive rock/ore not having any ore minerals. The factor of solid substance.
F_5	The liquid substance of low density not having any ore minerals; nevertheless, is electrically conductive. The factor of fluid.

Let us try to find each of the factors. This you can do with the help of matrix **ω** after Tab. 6. As I did have no experience with the process of identification, I accepted certain principles:

- The positive sign of the directional cosines presents direct relation between factor and characteristic.
- The negative sign of the directional cosines presents indirect relation between factor and characteristic.
- The zero and those all being close to zero confirm that the registered characteristic does not influence on the factor.

On the base of those principles here are the results of geological identification.

- Factor F_1 – non-conductive, highly polarized rock/ore having heavy radioactive minerals.
- Factor F_2 – conductive, highly polarized rock/ore having magnetic minerals.
- Factor F_3 – non-conductive rock/ore having ore minerals not being radioactive and magnetic.
- Factor F_4 – non-conductive rock/ore not having any ore minerals.
- Factor F5 – a substance of low density. The radioactive and magnetic minerals are missing.

Tab. 11 The transformed factors of characteristics as a function of the borehole depth

h [m]	F_1^*	F_2^*	F_3^*	F_4^*	F_5^*	Sum	$F_1^*+F_2^*+F_3^*$
350	0.006	0.043	0.784	0.005	0.162	1.000	0.833
350.5	0.200	0.237	0.001	0.387	0.176	1.000	0.437
351	0.534	0.043	0.037	0.385	0.001	1.000	0.614
351.5	0.288	0.415	0.025	0.271	0.001	1.000	0.728
352	0.037	0.131	0.075	0.420	0.337	1.000	0.243
352.5	0.911	0.028	0.026	0.017	0.019	1.000	0.965
353	0.105	0.040	0.098	0.250	0.508	1.000	0.242
353.5	0.115	0.126	0.192	0.272	0.296	1.000	0.432
354	0.000	0.251	0.000	0.289	0.459	1.000	0.252
354.5	0.050	0.440	0.016	0.381	0.113	1.000	0.506
355	0.035	0.709	0.030	0.218	0.008	1.000	0.773
355.5	0.042	0.001	0.347	0.036	0.575	1.000	0.390
356	0.000	0.097	0.723	0.021	0.159	1.000	0.820
356.5	0.036	0.267	0.575	0.121	0.001	1.000	0.878

The last substance is polarizable electrically, too. It could be the formation water or mud. Factor F_5 is a bit special. Because the resting factors form the group being significant, factor F_5 belongs to the other group being insignificant. Identification of factors is in Tab. 10. Fig. 1 offers a view of how distinct types of rock factors change with the borehole depth. However, for concrete information about the properties of rocks, it is not enough. Fortunately, Tab. 9 can be transformed in such a way as to resemble more curves of deterministic models.

Now, I must something say about graphical outputs interpreted data. Generally, holds that in processing laboratory data when you have point data, the depicted points are link by refracted lines. Depiction of good logs has, however, continuous curves. The recorded points of the physical characteristic with the depth are close to one another. You can it visualize as a dotted continuous curve with the borehole depth. Therefore, you see on the screen a continuous analogue curve. And just such analogues curve's physical characteristics with the borehole depth print out on paper as work material for Log Analysts and geologists. It is a long-term custom and in archives, there is a big amount of such paper records needed for comparing old and new records. It was the main reason the depicted curves of factors have a continuous form. Now however go to transformation calculated factors.

Tab. 12 Geological interpretation of transformed factors after characteristics

Geological interpretation of factors	
$\mathbf{F_1^*}$	Disseminated radioactive ore in rock; electrically non-conductive
$\mathbf{F_2^*}$	Compact magnetic ore or mafic rock; electrically conductive
$\mathbf{F_3^*}$	Disseminated nonmagnetic and nonradioactive ore in rock; electrically non-conductive
$\mathbf{F_4^*}$	The waste rock matrix forms of non-ore minerals; electrically non-conductive
$\mathbf{F_5^*}$	Water content of rock

We can transform each row with depth after the formula:

$$Y^*_{ij} = \frac{Y^2_{ij}}{\sum_{j=1}^{5} Y^2_{ij}} . \tag{18}$$

Input data are in Tab. 9. After transformation, the sum of all terms Y^*_{ij} in the row is 1; see Tab. 11. Tab. 11 presents the calculation made after formula (18); the transformed factors and their geological identification are in Tab. 12. This table says that holds:

- F^*_1 - Radioactive ore/rock being electrically non-conductive. (Disseminated ore formed from radioactive minerals in the waste rock.)
- F^*_2 - Magnetic ore/rock being electrically conductive. (It presents Fe, Co, Ni ores or mafic rocks as gabbro, basalts and phonolites are.)

- F_3^* - Nonradioactive Nonmagnetic ore/rock being electrically non-conductive. (It can be the rock impregnated with iron pyrites grains.)
- F_4^* - The waste rock being electrically non-conductive.
- F_5^* - Water content of rock; its high values are too an indicator of shale. It is about water, which is fluid, not solid matter, therefore the variant.

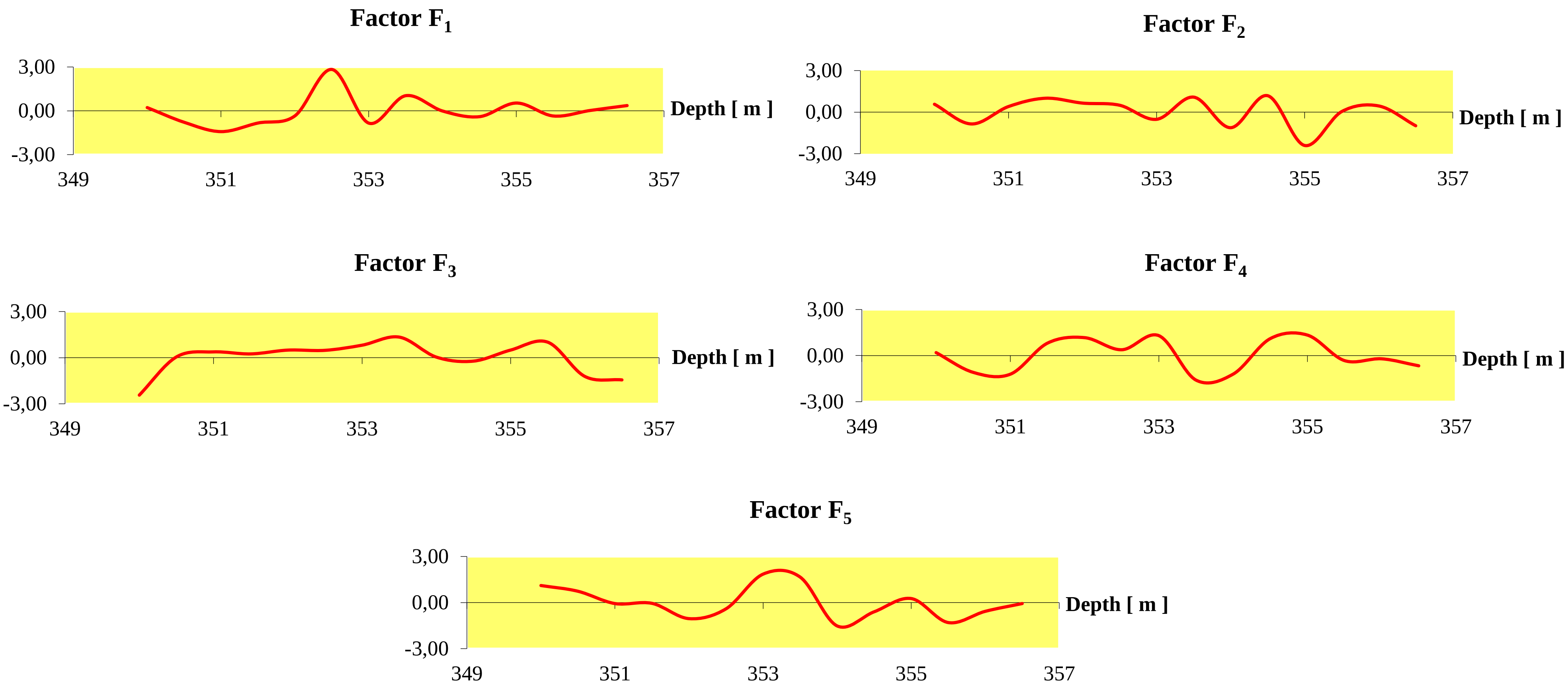

Fig. 1 Factors of rocks depicted as the function of depth in form of continuous curves

The sum of factors F_1^*, F_2^* and F_3^* is a function and presents the bulk content of ore in the rock; together with factor F_4^* they form the rock matrix. The remaining value of 1 presents the water content of the rock. The results of transformed factors depict in Fig. 2 and Fig. 3. Tab. 13 presents curves of ore content and water content of rocks for Fig. 3. What is in Fig. 3 is an independent virtual model that you can offer for real geological interpretation. Important is to find a plausible explanation between the model and reality. Radioactive can be potassic feldspars as well as uraninite; magnetic again can be basalts instead of the iron ore deposits. The Log Analyst as a geologist must be very watchful and highly smart.

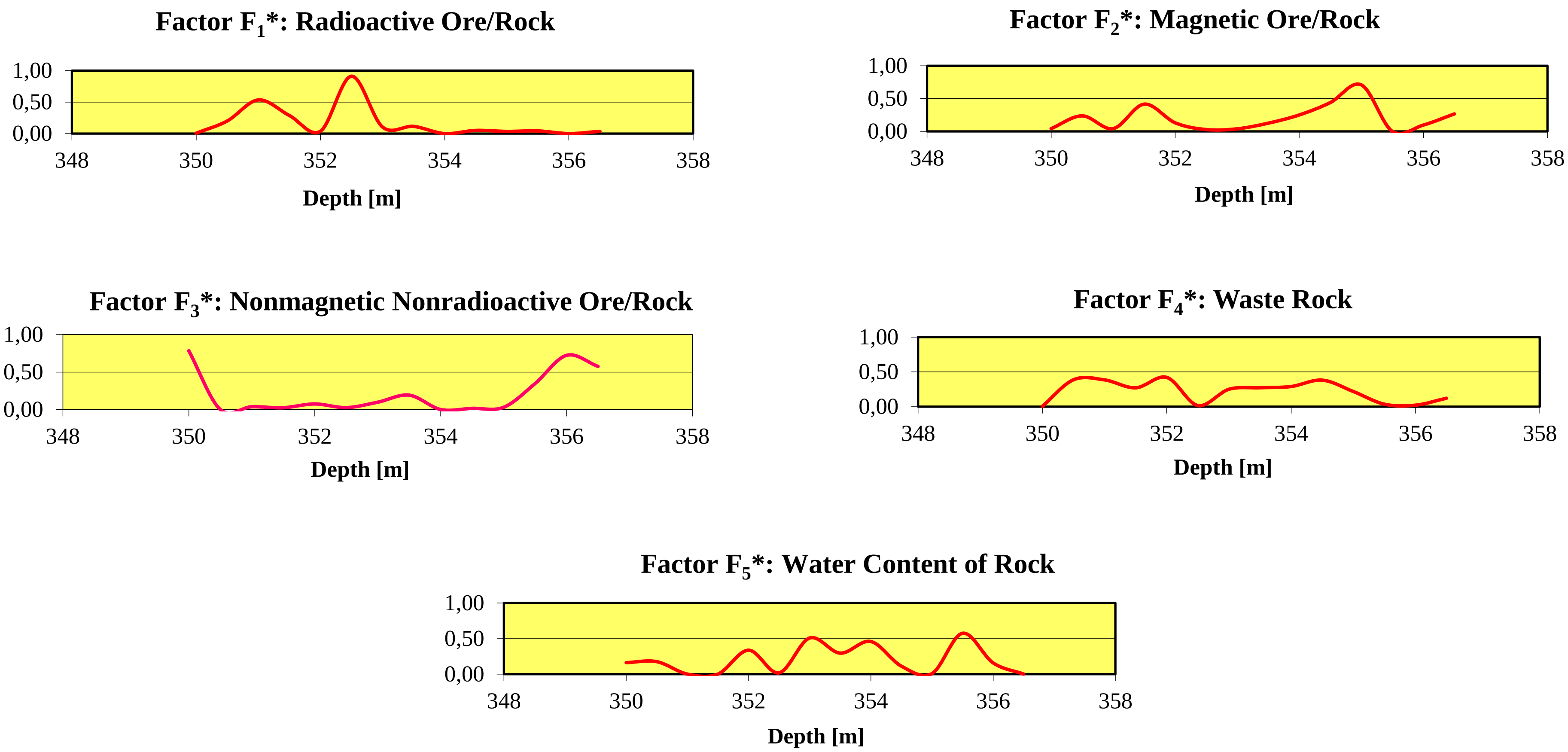

Fig. 2 The continuous curves of transformed factors closed to reservoir properties

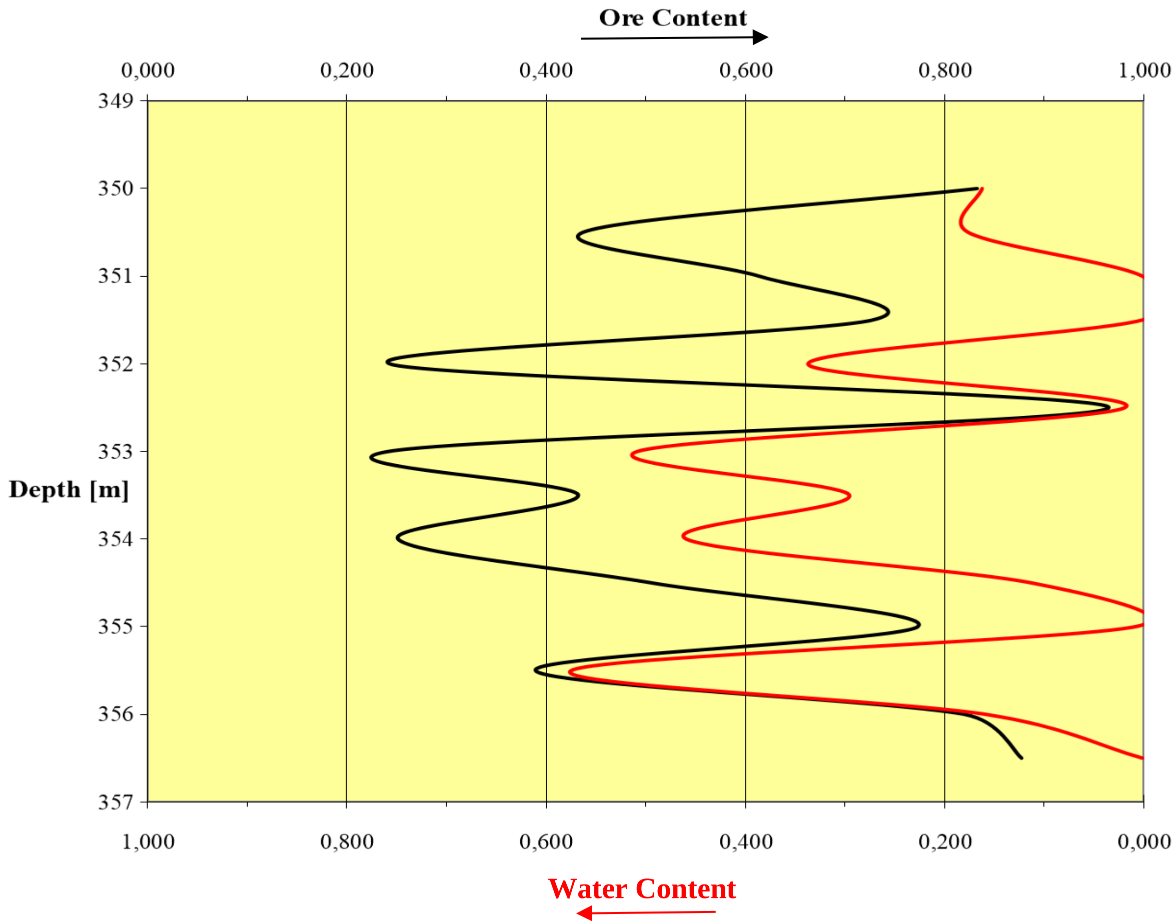

Fig. 3 The curves of the rock matrix, ore content and water content made after Tab. 13

Tab. 13 The continuous curves of ore content and water content of rocks with the borehole depth

Ore Content and Water Content		
h [m]	**$F_1^*+F_2^*+F_3^*$**	**F_5^***
350	0.833	0.162
350.5	0.437	0.176
351	0.614	0.001
351.5	0.728	0.001
352	0.243	0.337
352.5	0.965	0.019
353	0.242	0.508
353.5	0.432	0.296
354	0.252	0.459
354.5	0.506	0.113
355	0.773	0.008
355.5	0.390	0.575
356	0.820	0.159
356.5	0.878	0.001

Types of rocks after Tab. 12 could quite well characterize the rocks being in Czech Cretaceous Basin. Further, you can make to interpretation in the cased borehole. If I have enough methods available to register through steel Colonna, I shall have curves reporting water content, shaliness or saturation by hydrocarbons. This is for a deterministic model almost impossible. However, for that, I need to have non-electrical logs such as the Density Log, Acoustic Log, Gamma-Ray Log and one, two or all three variants of the Neutron Log.

Fig. 3 presents the curves of the ore content and water content of rocks. The ore content is the ore-component of the matrix of rocks; has three partial components: the radiation, the magnetic and the electric ones. The red curve belongs to the water content of rocks. The supplement of the curve up to value one presents the sum of both the ore-component and the non-ore-components of the matrix of rocks. This common sum is the total matrix of rocks.

7. Analysis after depth points

In this case, I use the correlation matrix $\mathbf{R}_2$ depicted in Tab. 4. This matrix has dimension {14×14} which is difference because the correlation matrix $\mathbf{R}_1$ said before to be {5×5}. The process of calculation of the roots for the characteristic equation is fully identical to that described in the chapter about analysis after characteristics. However, condition (8) you must adjust, because the new matrix has other dimensions. It holds that:

$$Tr\ \mathbf{R_2} = \sum_{j=1}^{14} \lambda_j = 14. \tag{19}$$

This condition presents that any of the roots of the characteristic equation cannot be equal or higher than 14, because there holds that $\lambda_j < 14$. We shall use again succession **Implements → Finding Solution**. The computer makes again successive divisions of determinant for the characteristic matrix **F** (λ_j). So, you will get the roots: $\lambda_1 = 6.655$, $\lambda_2 = 4.770$, $\lambda_3 = 1.626$ and $\lambda_4 = 0.950$. It holds that Tr $\mathbf{R}_2$ = (6.655+4.770+1.626+0.950) =14.001. Their sum per condition (19) must be 14. That means to stop the next division, even if **Finding Solution** offers the next real roots. This is the main difference in comparison to the calculation of roots having a relation to matrix $\mathbf{R}_1$. The roots are only four instead of five and simultaneously with roots; we reach four matrices **F** (λ_1), **F** (λ_2), **F** (λ_3) and **F** (λ_4).

The process of orthogonalization is in the chapter about analysis after characteristics. In this case, the matrix **F** (λ_j) has its dimensions {14×14}. The column matrix/vector remarked as $\mathbf{a_j}$ has dimensions {14×1}. And the same dimensions are valid for the multiplication of **F** (λ_j) × $\mathbf{a_j}$. It holds that **F** (λ_j) × $\mathbf{a_j}$ ≡ {14×14} × {14×1} ≡ {14×1}. We use again succession **Implements → Solutionist** and conditions (9) and (10). You will get to the modal matrix **a** having dimension {14×4} as you work with 4 roots! What is characteristic of that is the sum of squared terms in columns is equal to1, however, the sum of squared terms in rows is extremely lower than 1. This is a difference in comparison to the former modal matrix because $m = 4 < 14$! You see it well in Tab. 14.

If we form the matrix of directional cosines **ω** with the help of formula (11), we shall reach the matrix being {14×4}. And for this matrix there hold both conditions (12) and (13). Here is a coincidence yet. All factors F_1, F_2, F_3 and F_4 are more significant; therefore, they are in bold black letters. The matrix **ω** is in Tab. 15. Precision of calculation you control after matrix $\mathbf{R_2}$, see formula (15), and after matrix **Λ**, formula (16). Both matrices are in Tab. 16 and Tab. 17.

It holds again: Tr $\mathbf{R_2}$ = (6.7+4.8+1.6+0.9) = 14. Precision is lower, 1E-01, and this holds both for $\mathbf{R_2}$ and **Λ**. It is distinct; the original 14 factors you reduced to 4. Here is a well visible advantage of The Method of Principal Components. It needs to emphasize this is analysis after depth points not after characteristics. Therefore, we have only four factors F_1, F_2, F_3 and F_4, while analysis after characteristics presents five ones: F_1, F_2, F_3, F_4 and F_5. The next differences are in characterizing factors; for example, the factor of water content in Tab. 10 denoted as F_5 and is significant, but in Tab. 19, on contrary, is as F_2 and is insignificant.

Tab. 14 The modal matrix of multidimensional vectors $\mathbf{a}_i$ after the depth points

$\mathbf{a}_1$	$\mathbf{a}_2$	$\mathbf{a}_3$	$\mathbf{a}_4$	**Sum**
-0.153	-0.393	-0.212	0.207	0.265
0.125	0.417	0.013	0.274	0.265
0.374	0.025	0.202	0.024	0.182
0.363	-0.154	-0.055	0.030	0.159
0.351	-0.103	-0.145	-0.312	0.252
-0.363	-0.101	0.212	-0.067	0.192
0.238	0.251	-0.328	0.401	0.388
-0.072	-0.083	0.701	0.376	0.645
-0.195	0.327	0.100	-0.478	0.383
0.276	-0.309	-0.110	-0.108	0.195
-0.201	0.341	-0.324	-0.016	0.262
0.264	0.222	0.277	-0.436	0.386
-0.090	-0.435	-0.064	-0.203	0.243
-0.374	0.016	-0.195	-0.066	0.183
1.000	1.000	1.000	1.000	

Tab. 15 The matrix of directional cosines ω after the depth points

	ω_1	ω_2	ω_3	ω_4	
1	-0.394	-0.857	-0.270	0.202	1.0
2	0.322	0.910	0.016	0.267	1.0
3	0.965	0.055	0.257	0.023	1.0
4	0.936	-0.337	-0.071	0.029	1.0
5	0.906	-0.226	-0.184	-0.305	1.0
6	-0.936	-0.220	0.271	-0.065	1.0
7	0.615	0.547	-0.418	0.391	1.0
8	-0.185	-0.181	0.894	0.367	1.0
9	-0.502	0.714	0.127	-0.466	1.0
10	0.713	-0.674	-0.141	-0.105	1.0
11	-0.518	0.745	-0.413	-0.015	1.0
12	0.681	0.485	0.353	-0.425	1.0
13	-0.232	-0.950	-0.081	-0.198	1.0
14	-0.966	0.034	-0.249	-0.064	1.0
λ_j	6.655	4.770	1.626	0.950	14.00
w [%]	47.5	34.1	11.6	6.8	100.0

8. Interpretation after depth points

From the point of geometry here is the 4-dimensional ellipsoid having 4 factors presenting the column vectors-axes. Each of the axes has 5 coordinates answering to the measured characteristics. All four factors have significant character, they are in bold black letters. Calculation of values of matrix **Y** is after the formula as follows:

$$\mathbf{Y} = \mathbf{X}^{*\mathbf{T}} \times \boldsymbol{\omega}. \tag{20}$$

The transpose matrix has dimensions {5×14} and matrix **ω** has {14×4}. That is why the matrix **Y** has dimensions {5×4}. The precision of directional cosines will be only 1E-01. This matrix depicts the relations of four factors to the registered characteristics. The field of matrix **Y** is possible to find in Tab. 18. Further, note please too in Tab. 19, factors are only 4, not 5, and signature factors are other than it was in Tab. 10. For example, the factor of water content in Tab. 19 is as F_2, in brown colour, the second most important, while in Tab. 10 it is as F_5, the least important.

Now, it is a time for the geological identification of factors. There hold the same three principles as it was before. The classification of factors is in Tab. 19. F_1 presents the factor of the magnetic ore minerals, while F_2 is the factor of the rock water content. F_3 classifies the factor of the radioactive ore minerals and F_4 is the factor of the nonmagnetic and nonradioactive ore minerals. So, the sum of factors $F_1 + F_3 + F_4$ presents the factor of ore content in the rocks. You could remember that this factor and the factor of water content together decide sufficiently about a statistical set. What is interesting is that the matrix of rock is not present here, but it is possible to reach it to be as the supplement of water content. The results of Tab. 18 are in Fig. 4. It is like a stripe plot giving particularly valuable information about mutual relations. Such information is not the main one, however, an important other one is.

Tab. 16 The correlation matrix after cosines ω denoted as R₂ as multiplication of matrices ω ×ω ᵀ

$\mathbf{R_2} = \boldsymbol{\omega} \times \boldsymbol{\omega}^T =$

1	-0.9	-0.5	-0.1	-0.2	0.5	-0.5	0.1	-0.5	0.3	-0.3	-0.9	0.9	0.4
-0.9	1	0.4	0.0	0.0	-0.5	0.8	-0.1	0.4	-0.4	0.5	0.6	-1.0	-0.3
-0.5	0.4	1	0.9	0.8	-0.8	0.5	0.0	-0.4	0.6	-0.6	0.8	-0.3	-1.0
-0.1	0.0	0.9	1	0.9	-0.8	0.4	-0.2	-0.7	0.9	-0.7	0.4	0.1	-0.9
-0.2	0.0	0.8	0.9	1	-0.8	0.4	-0.4	-0.5	0.9	-0.6	0.6	0.1	-0.8
0.5	-0.5	-0.8	-0.8	-0.8	1	-0.8	0.4	0.4	-0.6	0.2	-0.6	0.4	0.8
-0.5	0.8	0.5	0.4	0.4	-0.8	1	-0.4	-0.2	0.1	0.3	0.4	-0.7	-0.5
0.1	-0.1	0.0	-0.2	-0.4	0.4	-0.4	1	-0.1	-0.2	-0.4	-0.1	0.1	-0.1
-0.5	0.4	-0.4	-0.7	-0.5	0.4	-0.2	-0.1	1	-0.8	0.7	0.2	-0.5	0.5
0.3	-0.4	0.6	0.9	0.9	-0.6	0.1	-0.2	-0.8	1	-0.8	0.2	0.5	-0.7
-0.3	0.5	-0.6	-0.7	-0.6	0.2	0.3	-0.4	0.7	-0.8	1	-0.1	-0.6	0.6
-0.9	0.6	0.8	0.4	0.6	-0.6	0.4	-0.1	0.2	0.2	-0.1	1	-0.6	-0.7
0.9	-1.0	-0.3	0.1	0.1	0.4	-0.7	0.1	-0.5	0.5	-0.6	-0.6	1	0.2
0.4	-0.3	-1.0	-0.9	-0.8	0.8	-0.5	-0.1	0.5	-0.7	0.6	-0.7	0.2	1

Tab. 17 The fundamental matrix Λ, made after the directional cosines ω, as multiplication of matrices ωᵀ × ω

$\boldsymbol{\Lambda} = \boldsymbol{\omega}^T \times \boldsymbol{\omega} =$

6.7	0	0	0
0	4.8	0	0
0	0	1.6	0
0	0	0	0.9

Tab. 18 Factors of the depth points in the form contributions of characteristics

	F_1	F_2	F_3	F_4	$F_1+F_3+F_4$
R [Ωm]	-0.7	1.0	-0.6	0.0	-1.3
$\chi\times10^6$[SI]	1.3	-0.3	0.5	0.0	+1.8
σ[μBq/g]	-0.6	-0.2	2.2	0.7	+2.3
ρ[g/cm^3]	-0.9	-0.4	1.0	-1.2	-1.1
ε	-0.6	-1.3	0.1	0.6	+0.1

Tab. 19 Interpretation of factors after points of depth

Interpretation of factors	
F_1	Factor of ore magnetic minerals.
F_2	Factor of water content of rock.
F_3	Factor of ore radioactive minerals.
F_4	Factor of ore minerals nonmagnetic and nonradioactive.
$F_1+F_3+F_4$	Factor of ore content.

9. Next possibility of evaluation of geological factors

All, what was speaking before, is only the fundament of the next evaluation. There exist next highly effective ways of interpretation. I can mention the methods of taxonomy using 2-dimensional, 3-dimensional, or more-dimensional depictions of factors. It is possible to apply the methods of regression analysis, discriminate analysis or disperse analysis, too. The extent choice of methods is exceptionally large. One can solve tasks about extremes, as well. However, just method of factor analysis looks to be for the interpretation of well-logging data significant and this is why I should like to let someone who will be interested in studies of this method for well-logging data. In this paper I tried only to write a guidebook of the Method of Principal Components applied for a small well-logging statistical set. New way of sophisticated

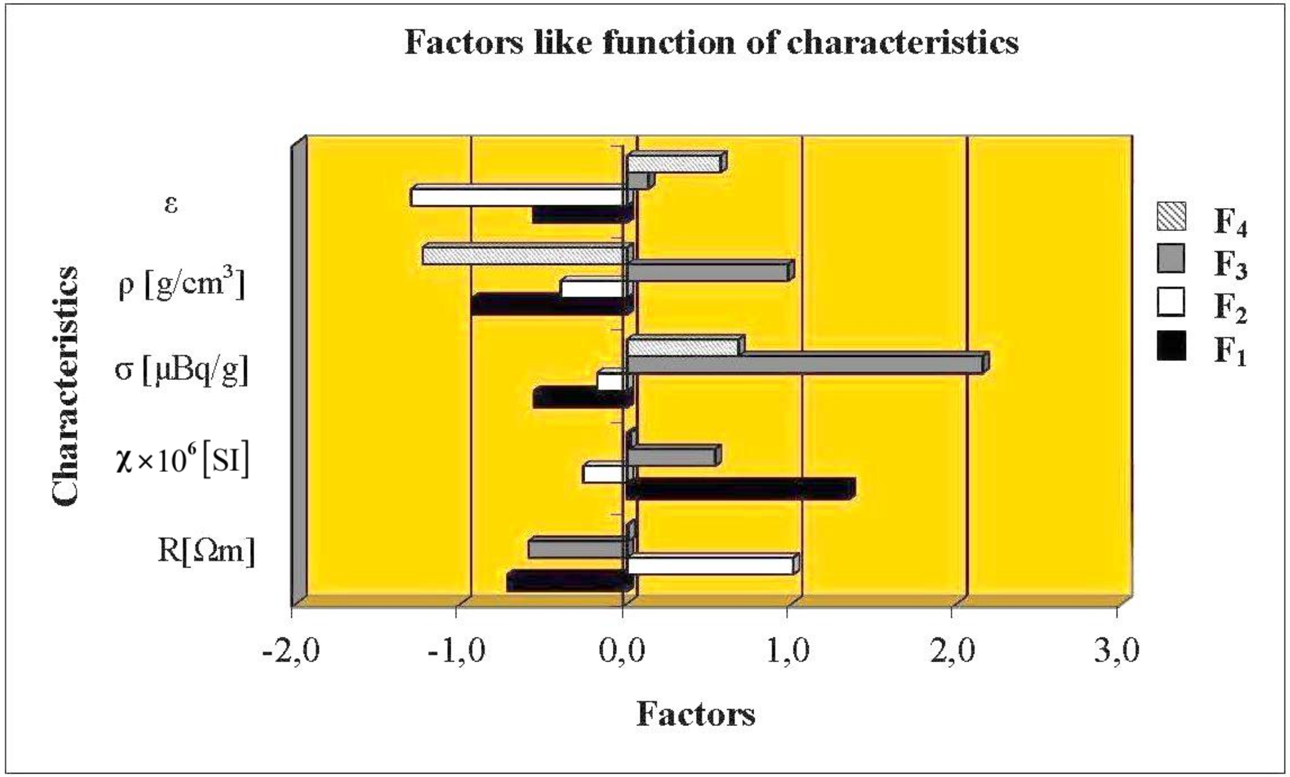

Fig. 4 Factors of depth points and characteristics depicted in form of stripe plot

interpretation of well-logging data exists; however, big well-logging statistical sets need a bit other statistical way; it is about ways for big data set processing.

10. Conclusions

Due to one of two methods of the factor analysis remarked Principal Components Analysis applied for a well-logging statistical set I can present these conclusions:

- Method PCA is well-available not only for well-logging registration, but for geophysical registration along with profile lines being on the surface of the earth, as well. It only needs to exchange the column of depth points for the column of profile points.
- The input data you must standardize as relative ones in Tab. 2 which presents the basic table for the next data processing.
- The analysis is after two different matrices of correlation. The first is the matrix of characteristics; the second is the matrix of the depth points.
- For calculation of small statistical sets, you can use the processor Microsoft Excel, having Implements Solutionist and Finding Solution, in particular.
- For the calculation of large statistical sets is advantageous to use processors such as MATLAB or Python.
- The results of orthogonalization are modal matrices making it possible to form the matrices of directional cosines.
- As result of the analysis after characteristics, there is a table of factors like the function of the borehole depth. This enables us to depict the factors in the form of classical continuous curves with the depth that presents the output depiction used by Log Analysts and Geologists.
- Factors F_j* can depict logs as well with depth and you must them geologically reliably interpret.
- Fig. 3 offers an interpretation as three-component ore and waste rock; fact, it can be quite well sands and sandstones having facies with pyrite, basalt gravels/cobbles and weathered potassium feldspars.
- And not always the geological interpretation of factors successful is, even if the above factors reliably exist.
- Next transformation makes the factors interpreted like curves of rock properties like those curves of the deterministic model.
- As result of the analysis after the depth points there is the table of factors like relations of various characteristics. It is possible to depict them like stripe plots; all factors you can find geologically, however, not always is possible. The stripe plots present how which factors influences the corresponding characteristics, whether it is positive, negative or zero relations.

References

AJVAZJAN, S., BEŽAJEVOVÁ, Z., STAROVEROV, O. *Metody vícerozměrné analýzy*. Prague, SNTL, 1981, 252 s.

ASFAHANI et al., Radioactive characterization of Ar – Rassafeh Badyieh area (Area-2), Syria by using Statistical factor analysis technique. *Contribution to Geophysics and Geodesy*, 2018, *Vol. 48/*2, p. 113–132.

BELONIN, M. D., GOLUBEVA, V. A., SKUBLOV, G. T. *Faktornyi analyz v geologii*. Moscow, Nedra, 1982, 265 s.
BLAHUŠ, P. *Faktorová analýza a její zobecnění*. Praha, SNTL, 1985, 354 s.
FABRIGAR, L. R. *Exploratory Factor Analysis*. Oxford, Oxford University Press Inc., 2012.
FINCH, W.H. *Exploratory Factor Analysis*. Sage Pub., 2019.
DMITRIEV, V. I. et al. *Vyčislitelnyje matematika i technika v razvedocnoj geofyzike, Handbook*. Moskva, Nedra, 1982, 218 s.
JOLLIFFE, I. T. Principal Component Analysis. *Springer-Verlag New-York Inc.*, New-York, 2002, 488 s.
MA, Y. Generalized Principal Component Analysis. *Springer-Verlag New-York Inc.*, New-York, 2015, 559 s.
MELOUN, M., MILITKÝ, J. *Statistická analýza experimentálních dat*. Praha, ACADEMIA,2004, 953 s.
MELOUN, M., MILITKÝ, J. *Interaktivní statistická analýza dat*. Praha, Nakladatelství Karolinum, 2012, 960 s.
NICULESCU, B.M., ANDREI, G., CIUPERCĂ C.L. Improved Formation Evaluation through Principal Component Analysis. *European Association of Geoscientists & Engineers* (*EAGE*) *78th Conference Exhibition including SPE EUROPEC 2016* "*Efficient Use of Technology – Unlocking Potential*", Vienna, Austria, Conference Proceedings, paper Tu STZ2 15, 2016, ISBN 978-94-6282-185-9, ISSN 2214-4609, DOI: 10.3997/2214-4609.201600899, pp. 4.
NICULESCU, B.M., ANDREI, G. Principal Component Analysis as a Tool for Enhanced Well Log Interpretation. *Revue Roumaine de Géophysique / Romanian Geophysical Journal*, tome 60 (2016), Bucharest, Romania, Publishing House of the Romanian Academy, 2016, ISSN 1220-5303, pp. 49–61.
REKTORYS, K. et al. *Přehled užité matematiky,* druhé opravené vydání, SNTL, Prague, 1968, 1135 s.
SIGORSKIJ, V. P. *Matematitcheskyi apparat ingenyera*, Technika, Kyjev, 1975, 753 s.
SHERIFF, R.E. *Encyclopedic Dictionary of Exploration Geophysics,* Dictionary, Fourth Edition, Oklahoma, Tulsa, Society of Exploration Geophysicists, 2002.
ŠKRÁŠEK, J., TICHÝ, Z. *Základy aplikované matematiky, část 3*. Praha, SNTL, 1986, p. 853.

Author:
[1] RNDr. František Ryšavý, Lesní 3, 695 03 Hodonín, Czech Republic, rysavy.frantisek@seznam.cz