Exploration Geophysics, Remote Sensing and Environment *XXXI.2 (2024)*

DOI: 10.26345/EGRSE-001-24-201

STATISTICAL METHODE OF MATHEMATICAL GROUPING FOR THE CREATION OF THE BOREHOLE LITHOLOGICAL PROFILE

STATISTICKÁ METODA MATEMATICKÉHO GRUPOVÁNÍ PRO VYTVOŘENÍ LITOLOGICKÉHO PROFILU VRTU

František Ryšavý [1]

Abstract

This work presents operating an algorithm of one Statistical Method of Mathematical Grouping for classification numerical well-logging data. The algorithm uses the correlation matrices of input data. The registered data are well-logging characteristics presented like function of the borehole depth. The algorithm is based on principles of grouping and its evaluation. Results of interpretation are probable types of formations characterized by their physical properties and they are linked to the borehole depth. So, we form the borehole profile after geophysical characteristics which can be compared with the geological profile.

Abstrakt

Práce představuje operace s algoritmem jedné statistické metody matematického grupování, pro klasifikaci numerických karotážních údajů. Tento algoritmus používá korelační matice vstupních dat. Registrované údaje jsou karotážní charakteristiky presentované jako funkce hloubky vrtu. Algoritmus je založen na principech grupování a jeho vyhodnocení. Výsledky interpretace představují pravděpodobné typy hornin, které jsou charakterizovány svými fyzikálními vlastnostmi a které jsou navázány na hloubku vrtu. Vytvářejí se tak profily vrtu podle geofyzikálních charakteristik, které se mohou srovnávat s geologickým profilem.

Keywords

Mathematical Grouping, heuristic methods, well- logging data, numerical data, correlation matrix

Klíčová slova

matematické grupování, heuristické metody, karotážní data, numerická data, korelační matice

1 Introduction

Statistics can be used in various mathematical analyses. One of the main methods is Mathematical Grouping. You can evaluate not only numeric data, but too logic data as are “true” denoted with symbol “1” and “false” with symbol “0”. It is heuristic cases when classification works with logic data are marked with logic symbols “1” and “0”; in the next evaluation they will change in numerical values with numerical symbols 1 and 0 and respect all mathematical operations of method Mathematical Grouping.

Each geological investigation of various objects presents a set of data which needs to be evaluated. You can work with both numeric and logic data. The main problem of geological tasks is grouping well-logging data into groups and assign them to geological objects. As for heuristic cases SIROTINSKAYA, (1986) distinguish three groups of these methods. Their relations tend often to methods Mathematical Logic.

Newer methods of Mathematical Logic use highly professional instruments as branched logic networks. These simulate human neural nets and imitate human cerebration, fuzzy logic. Modern algorithms are constructed as self-taught systems and tend to artificial intelligence, AI. System of geophysical data is combined with next geological systems such as petrographic, mineralogic, hydrogeologic and next ones are. They are often highly specialized on the selected geological area being perspective for next prospection and mining. Such new published works are, for example, CHEN et al, (2020), VUKADIN (2023), KARPENKO (2022) and CORINA (2018).

2 Input data

Input data are all identical with those having been used in the work investigating the Component Analysis in well-logging; more is in RYŠAVÝ, (2023). It offers to compare results of Mathematical Logic with results of that method. Basic is tab.1 that is identical to tab.1 of the mentioned work. It presents five physical characteristics registered simultaneously with the borehole depth. They are the following ones:

- The apparent resistivity of formations/ores; R [Ωm]
- The apparent magnetic susceptibility of formations/ores; $\chi \times 10^6$ [SI],
- The apparent bulk activity of radionuclides being in formations/ores; σ [μBq/g]
- The apparent bulk density of formations/ores; ρ [g/cm^3]
- The apparent chargeability formations/ores; ε, characteristic is dimensionless.

The tab.1 has two correlation matrices denoted as tab.3 and tab.7. Numeric values of mentioned three tables are same as data tables tab1, tab.3 and tab.7 of work RYŠAVÝ (1923). The first correlation matrix is matrix of characteristics, tab.3; the second one is the correlation matrix of the depth points, tab.7. The last is more important than the first. All operations as grouping and forming the partial matrices for partial groups of objects are proceeding from those two matrices. The submitted work needed urgently real well-logging data for verification of algorithm. However, such data I did not have. Luckily, in the work BELONIN et al. (1982) there was the table of input data of surface measurements {14×5}; 14 points of observing and 5 characteristics. This table served as a base for Tab.1of work RYŠAVÝ (1923). The numeric data of the input table stuck the same, but the registered characteristics was changed. For example, I replaced the thickness of the Jurassic cover with the mass activity of radionuclide [μBq/g] and were the next changes. All 14 points of observations were attached to the depth of the borehole. Thus, the virtual practice data-set section of a virtual borehole emerged. Such borehole section was needed for performing how to imply the algorithm of logic method into interpretation well-logging data set, and how to make geological interpretation. It had not any possibility to obtain well-logging logs, because the firm´s records are guarded because of competition. This was the only possibility how to have a short section of virtual borehole section. Thanks to the virtual dataset of tab.1 processing of implication of the logic method was possible. Without the virtual dataset it would not be possible to demonstrate how logical algorithm to imply for well-logging data. It was and remains by the aim of this work.

Tab.1 has some data, which need short commentary. They are columns having characteristics σ and ρ. Three data of radioactivity σ are zero, σ = 0. And the same is for characteristic of density. One data of density ρ has too zero value, ρ = 0. In first view it cannot be, but who knows well of well-logging data records cannot such zero value all exclude. It is about case of failure recording, for example, unexpected voltage drops. It is not excluded that in the bigger datasets the depth points having in columns value zero will form own subgroups with common alphabetic character "0" inside of the main groups. Such subgroups can then be denoted as not available.

Tab.1 The input characteristics depicted as data of well logs

i	h [m]	R [Ωm]	$\chi\times10^6$[SI]	σ[μBq/g]	ρ[g/cm^3]	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

Next commentary is again for characteristic ρ. It relates to high and low values of the rock density. They are following values: ρ = 0.70g/cm^3 and ρ = 4.04g/cm^3. Both values are no extra ones. They are densities that are known in geology. As for the lower value, it can be the water highly saturated with gas, denoted too as a gas-cut water, or the associated water when simultaneously you drill water and oil; $\rho_g \sim 0$ g/cm^3, $\rho_o \sim 0.7$–0.9g/cm^3. Such cases can be in caverns with higher diameter formed in the drilling. It can be often in Flysch Formation, and it is not an extra case. The higher value can be ores of the heavy chemical element as Hg, Pb, Fe, Ba, and the next ones. It can happen in geology. I can present some ore and their densities; as are HgS: ρ = 8.1g/cm^3, PbS: ρ = 7.5 g/cm^3, FeS_2: ρ = 5.1 g/cm^3, $BaSO_4$ ρ = 4.5 g/cm^3… An existence of FeS_2 in rocks can be often repeated case. It is the main raw material for production of oil vitriol, H_2SO_4. Tab.1 and tabs. 3 and 7 served with all numeric data as fundament for interpretation in the work RYŠAVÝ, (2023). I decided all three before tables to use again for interpretation with Method Mathematical Grouping.

3 Principles of classification of objects and their deposition into groups

Classification of well-logging data is made in this paper in accordance with algorithm "Class" published by SIROTINSKAYA, (1986). This method works with the correlation matrix of well-logging data. As an input set there has been used the set of data before used for the paper about application of the Component Analysis. Therefore, I refer to this work. It is big advantage for mutual comparison between two different methods solving identical problem. Algorithm "Class" works on principles of Mathematical Grouping.

We know to work with the classic correlation matrix when the data are presented like numerical quantified values. However, the correlation matrix can be presented, as well, with data of logic values using two basic figures 0 and 1. They belong to values of quality as an assortment of colours and next similar characteristics are. Just this is also published in work SIROTINSKAYA, (1986)**,** input table can have all terms expressed figures like 0 and 1 being in m-rows and in n-columns. The correlation matrix is defined then with the help of **the measure of coincidence** E ($\boldsymbol{X_i}$**,** $\boldsymbol{X_j}$) for i-th object and j-th characteristic of the above matrix of the input table. Characteristics $\boldsymbol{X_i}$ and $\boldsymbol{X_j}$ are vectors of the input matrix having dimension denoted as $\{m \times n\}$. Calculation of the mentioned characteristic E ($\boldsymbol{X_i}$**,** $\boldsymbol{X_j}$) tends to be determining elements of correlation matrix having *m* rows and columns.

$$E(\mathbf{X_i},\mathbf{X_j}) = \cos(\mathbf{X_i},\mathbf{X_j}) = \frac{\sum_{p=1}^{n} \alpha_{ip} \times \alpha_{jp}}{\sqrt{\sum_{p=1}^{n} \alpha_{ip}^2 \times \sum_{p=1}^{n} \alpha_{jp}^2}}, \quad i,\ j = 1,2..,m, \qquad (1)$$

where α_{ip}, α_{jp} = the coordinates expressed in form of figures 0 and 1; *m* is the number of rows and *n* is the number of columns of input matrix with figures 0 and 1.

The formula is saying that numerator of fraction is the sum of multiplication of figures 0 and 1 for selected terms of the input table, whereas the denominator of fraction presents square root of multiplication of two sums squared terms of 0 and 1. Attention, numeric calculation after formula (1) is calculation in decimal system with two numerical values 0 and 1. The figures here are not logic values.

In case tab.1 when we use classic correlation matrix expressed with figures presenting **numerical values**, the situation is simpler. As

Tab.2 The classification of correlation coefficients

Degree of correlation	**Interval**
Low	$0 \le \lvert r \rvert < 0.3$
Mild	$0.3 \le \lvert r \rvert < 0.5$
Significant	$0.5 \le \lvert r \rvert < 0.7$
High	$0.7 \le \lvert r \rvert < 0.9$
Very high	$0.9 \le \lvert r \rvert < 1$

the input set is depicted like relations of the characteristics registered to the borehole depth, we receive two matrices: after characteristics and after the points of depth. So there exist again two different variances of evaluation completing information one another.

Because it needs to evaluate the degree of correlation between two objects, there uses classification according to CHADDOCK, in ŠKRÁŠEK and TICHÝ, (1983), part 3. It is in tab.2.

4 Interpretation of the correlation matrix after characteristics

4.1 Mathematical processing

This is problem of Grouping. The correlation matrix has dimensions {5×5}, tab.3. All terms of this matrix are numbers. The **first step of grouping** is you must **mark** the **maximal values** expressed **in their absolute values** being in **each of rows**. I emphasize, they are in rows. We do not distinguish sign of terms. Maximal values on the main diagonal denoted as 1 are crossed out. The input matrix with denoted values **in bold type** is presented by tab.3.

The second step is we calculate the **average coefficient** of **correlation** of **all terms** lying **under the main diagonal**. We receive $n_0 = 4+3+2+1 = 10$ terms. The formula of calculation is following:

$$r_0 = \frac{1}{n_0} \times \sum_{j=1}^{(n-1)} \sum_{i=j+1}^{n} r_{ij} , \qquad (2)$$

$$n_0 = \sum_{i=1}^{(n-1)} i , \qquad (3)$$

where r_0 = the average coefficient of correlation

In this case $n_0 = 10$ and $| r_0 | = | + 0.126 |$. The next step is testing of maximal values denoted as $r_{ij}^{(max)}$ towards $| r_0 |$. We are allowed to group if all maximal values are bigger than $| r_0 |$. It is condition that:

$$|r_0| \langle | r_{ij}^{(max)} | \ldots \text{for all } r_{ij}^{(max)} . \qquad (4)$$

If the only one of values does not fill condition (4), you must not make grouping of characteristics.

We have got three $r_{ij}^{(max)}$: $|-0.501|$, $|+0.743|$ and $|+0.653|$. Because all three characteristics are bigger than $| r_0 |$ we are allowed to group. The rows and columns are numbered as it is in tab.3. We

Tab.3 The correlation matrix after characteristics with denoted maximal values

		R [Ωm]	χ ×10^6[SI]	σ[μBq/g]	ρ [g/cm^3]	ε
		1	**2**	**3**	**4**	**5**
R [Ωm]	**1**	1	**-0.501**	0.252	0.391	-0.074
χ ×10^6[SI]	**2**	**-0.501**	1	-0.118	-0.423	-0.178
σ[μBq/g]	**3**	0.252	-0.118	1	**0.743**	0.512
ρ [g/cm^3]	**4**	0.391	-0.423	**0.743**	1	0.653
ε	**5**	-0.074	-0.178	0.512	**0.653**	1

create succession of numbers according to the following way. The underlined number is **the number** of **columns**. The next **numbers** of each succession being not underlined are numbers **of rows** of **those terms bold typed** in **the column**. This process is made for all columns. Some of them can have the only underlined number, because there are no terms having maximal values, the other have two or more terms of maximal values. The result of grouping is in tab.4. Process of grouping is made in the way; we choose group having most figures in column and we find common figures in other successions.

The figures can have either direct relation or mediated one. It is our case for 1 and 2 of group {A} and for 3, 4 and 5 of group {B}. The group {A} belongs to the direct relation, whereas group {B} presents the mediated relation. However, grouping is all sure closed. The underlined figures in column "Succession" of tab.4 denote columns in tab.3.

So, we have two groups: {A} = {R, χ} and {B} = {σ, ρ, ε}. They are the groups of the first order. Now, we must calculate the correlation coefficient between both groups; it holds {A} = {1,2}, {B} = {3,4,5}.

$$r_{AB} = \frac{1}{6} \times \left(r_{13} + r_{14} + r_{15} + r_{23} + r_{24} + r_{25} \right) = -0.025 . \qquad (5)$$

Number coefficients of correlation present 2×3= 6 terms. We are to form the matrix of groups {A} and {B}. It is in tab.5. The only maximal value of both rows is | -0.025 |. We evaluate the value denoted like $|r_{ij}^{(max)}|$ to $|r_0|$ after condition (4). As this condition does not hold, the grouping is over. The values of the correlation coefficients are in tab.6.

Tab.4 The denoted successions and groups after characteristics

Succession	Group
1,2	**A** = {1,2}
2,1	**B** = {3,4,5}
3,4	
4,3,5	
5	

Tab.5 The correlation matrix of groups {A} and {B} after characteristics

	A	**B**
A	1	**-0.025**
B	**-0.025**	1

Tab.6 The table of correlation coefficients $\mathbf{r_0}$ and $\mathbf{r_{AB}}$ valid for characteristics

$\mathbf{r_0}$	0.126
$\mathbf{r_{AB}}$	-0.025

4.2 Results and discussion

We obtained two basic groups {A} and {B}. In such case you can construct the map of relations between characteristics. Both groups present object characterized as the sort of rock/ore. Results of mapping you will find on fig.1. Both groups you can characterize. The group {A} can be massive ore or mafic rock being electrically conductive and magnetic. It is electric conductor. However, the group {B} presentsdielectric. This dielectric is radioactive and has high density. It seems as if radioactive ore have been dispersed in the matrix of sedimentary rock.

The degree of correlation between {A} and {B} is low, I should say there is not any correlation. Correlation inside of group {A} between χ and R is denoted as significant, tab.2, and confirms electrical conductivity of object. For group {B} it holds that correlation between σ and ρ is high, whereas, resting two relations are only significant. See again tab.2. It needs to say that such identification of objects confirms only relations among them.

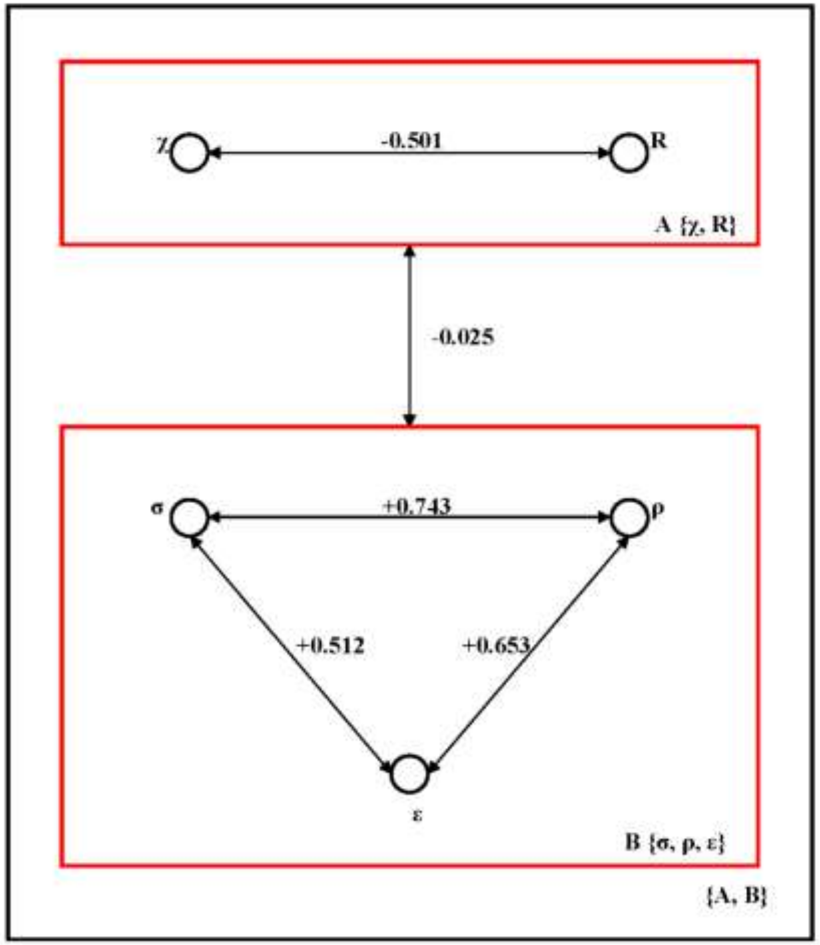

Fig.1 The map of relations between characteristics

5 Interpretation of the correlation matrix after the depth points

5.1 Mathematical processing

The input data are presented by the correlation matrix in tab.7. These data are calculated due to the table denoted as tab.1 and published before in the work interested in interpretation of Component Analysis; RYŠAVÝ, (2023). Maximal values of the correlation coefficients regardless of sign are typed in tab.7 with bold figures for every row. Grouping is made in accordance with principles explained in the chapter before.

Coefficient of correlation from tab.10, $|r_0| = |-0{,}065|$, was again counted as an average sum all terms being below the main diagonal of tab.7. **Number of all terms below the diagonal** is $n_0 = 13+12+\ldots+2+1 = 91$. The above conclusion after condition (4) holds

Tab.7 The correlation matrix after the depth points with denoted maximal values; the last row presents partial sums

	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
Sum	-1.670	0.231	0.491	-0.528	-1.207	0.272	-1.089	-0.736	0.203	-0.818	-0.062	-1.264	0.232	

again for grouping after maximal values in absolute value. Maximal values of the correlation coefficients regardless of sign are typed with **bold figures** for every row and are higher than r_0. So, we have three basic groups of depth points denoted as A, B and C; all present the main groups; {A}= {1,2,12,13}, further {B}= {4,5,9,10,11} and {C}= {3,6,7,8,14}. Successions and groups are depicted in tab.8. We find what the depth points connect among.

Tab.8 The denoted successions and groups for the depth points

Succession	Group
1,12	**A** ={ 1,2,12,13 }
2,13	**B** = { 4,5,9,10,11 }
3,6,14	**C** = { 3,6,7,8,14 }
4,5,10	
5,4	
6,7	
7,8	
8,	
9,	
10,9,11	
11,	
12,	
13,1,2	
14,3	

Tab.9 The correction matrix of groups {A}, {B} and {C} after the depth points

	A	**B**	**C**
A	1	**0.031**	-0.019
B	0.031	1	**-0.072**
C	-0.019	**-0.072**	1

Tab.10 The correlation coefficients r_0, r_{AB}, r_{AC} and r_{BC} valid for the depth points

r_0	-0.065
r_{AB}	0.031
r_{AC}	-0.019
r_{BC}	-0.072

There in tab.9 are values of the correlation coefficients r_{AB}, r_{AC} and r_{BC}. This table presents the correlation matrix of groups {A}, {B}, and {C}. The numerical values of the correlation coefficients after formulas and the average coefficient of correlation are in tab.10. Number coefficients of correlation present following terms; r_{AB}: 4×5 =20 terms, r_{AC}: 4×5 =20 terms and r_{BC}: 5×5 =25 terms. We test again after condition (4). Coefficient of correlation is $| r_0 |$ = |-0.065|. For $r_{AB}^{(max)}$ = |+0.031| and $r_{AC}^{(max)}$ = |-0.019| the test of inequality is not filled, only for $r_{BC}^{(max)}$ = |-0.072| it is filled. That means that grouping is over. Groups {A}, {B}, and {C} are final ones.

Groups {A}, {B}, and {C} have subgroups of the depth points. Tab.11 presents the correlation matrix of {A}. The maximal positive values in rows are denoted in bold. Process of grouping presented with successions and subgroups is in tab.12. There exist two subgroups $\{A_1\}$ = {1, 13} and then $\{A_2\}$ = {2, 12}. The same has been made with groups {B} and {C}. The correlation matrix of {B} is in tab.13, while successions and subgroups in tab.14. We have got again two subgroups: $\{B_1\}$ = {4, 5, 10} and $\{B_2\}$ = {9, 11}. The correlation matrix of {C} is in tab.15 and its successions and subgroups in tab.16. The group {C} was divided into two subgroups: $\{C_1\}$ = {6, 8, 14} and $\{C_2\}$ = {3,7}.

$$r_{AB} = \frac{1}{20} \times \left(r_{1,4} + r_{1,5} + r_{1,9} + r_{1,10} + r_{1,11} + r_{2,4} + r_{2,5} + r_{2,9} + r_{2,10} + r_{2,11} + r_{12,4} + r_{12,5} + r_{12,9} + r_{12,10} + r_{12,11} + r_{13,4} + r_{13,5} + r_{13,9} + r_{13,10} + r_{13,11} \right). \quad (6)$$

$$r_{AC} = \frac{1}{20} \times \left(r_{1,3} + r_{1,6} + r_{1,7} + r_{1,8} + r_{1,14} + r_{2,3} + r_{2,6} + r_{2,7} + r_{2,8} + r_{2,14} + r_{12,3} + r_{12,6} + r_{12,7} + r_{12,8} + r_{12,14} + r_{13,3} + r_{13,6} + r_{13,7} + r_{13,8} + r_{13,14} \right). \quad (7)$$

$$r_{BC} = \frac{1}{25} \times \left(\begin{array}{l} r_{4,3} + r_{4,6} + r_{4,7} + r_{4,8} + r_{4,14} + r_{5,3} + r_{5,6} + r_{5,7} + r_{5,8} + r_{5,14} + r_{9,3} + r_{9,6} + r_{9,7} + r_{9,8} + r_{9,14} + r_{10,3} + r_{10,6} + r_{10,7} + r_{10,8} + r_{10,14} + \\ r_{11,3} + r_{11,6} + r_{11,7} + r_{11,8} + r_{11,14} \end{array} \right). \quad (8)$$

Groups {A}, {B}, and {C} have subgroups of the depth points. Tab.11 presents the correlation matrix of {A}. The maximal positive values in rows are denoted in bold. Process of grouping presented with successions and subgroups is in tab.12. There exist two subgroups $\{A_1\}$ = {1, 13} and then $\{A_2\}$ = {2, 12}. The same has been made with groups {B} and {C}. The correlation matrix of {B} is in tab.13, while successions and subgroups in tab.14. We have got again two subgroups: $\{B_1\}$ = {4, 5, 10} and $\{B_2\}$ = {9, 11}. The correlation matrix of {C} is in tab.15 and its successions and subgroups in tab.16. The group {C} was divided into two subgroups: $\{C_1\}$ = {6, 8, 14} and $\{C_2\}$ = {3,7}.

Here holds again condition (4), however adjusted, $r(X_1, X_2) < r_{ij}^{(max)}$; X ≡ {A, B, C}. The correlation coefficients are all negative, see tab.17, however, maximal values of tables are all positive; see tabs. 11, 13 and 15. Therefore adjusted condition is valid. Relations between subgroups inside each group must be evaluated by the degree of correlation. Here are formulas used for $r(X_1, X_2)$, where holds X ≡ {A, B, C}:

$$r(A_1, A_2) = \frac{1}{4} \times \left(r_{1,2} + r_{1,12} + r_{13,2} + r_{13,12} \right). \quad (9)$$

$$r(B_1, B_2) = \frac{1}{6} \times \left(r_{4,9} + r_{4,11} + r_{5,9} + r_{5,11} + r_{10,9} + r_{10,11} \right). \quad (10)$$

$$r(C_1,C_2)=\frac{1}{6}\times\left(r_{6,3}+r_{6,7}+r_{8,3}+r_{8,7}+r_{14,3}+r_{14,7}\right). \quad (11)$$

The values of these coefficients are in tab.17. This says that $r(A_1,A_2)$ = -0.816 and is classified like high, whereas, both resting coefficients $r(B_1,B_2)$ = -0.692 and $r(C_1,C_2)$ = -0.594 are significant only. But, in general, we can announce that relations between subgroups are much tighter than it is between groups. What is also interesting is all relations between subgroups are negative. Imagination of relations among the depth points can be made with the help of the map of relations.

5.2 Results and discussion

Results of mapping are presented in fig.2. Relations between the depth points inside of subgroups, tab.2, are very high, high, or significant. They are positive; the only exception is negative relation between points 8 and 14. However, the subgroup $\{C_1\}$ is a bit out of the picture of subgroups as we shall see later.

The graphic output of subgroups is presented by figures 3, 4, 5. For their construction you need to have the table of fundamental data presenting changes of the registered characteristics with the borehole depth. It is tab.1; more in RYŠAVÝ, (2023). As the characteristics have various units, the data has been calculated for comparison expressed in percentage. It is presented in tab.18. This table then serves as the basic one for the stripe plots of subgroups depicted as fig.3, fig.4 and fig.5.

Differences between subgroups are well visible for $\{A_1\}$ and $\{A_2\}$. There holds the rule: the more, the less. Subgroup $\{A_1\}$ has high ε, low R and χ is low too. In contrary, subgroup $\{A_2\}$ has low ε, but R and χ are high. It explains well negative relation between $\{A_1\}$ and $\{A_2\}$.

Significant relation is visible between subgroups $\{B_1\}$ and $\{B_2\}$. The degree of correlation between $\{C_1\}$ and $\{C_2\}$ is too classified like significant, but absolute value of - 0.594 is the lower. Interesting is subgroup $\{C_2\}$ with two depths´ points, **3** and **7**. Both have characteristic σ = 0, and then the point **7** has ρ = 0 too. Algorithm formed from them subgroup $\{C_2\}$ because they are two. In contrary the depths´ point **1** has also σ = 0, but an own subgroup has not, because is single. It is possible that for two such points inside of group A we would have next special subgroup too. It seems that in subgroup $\{C_2\}$is a tendency to grouping depths' points having figure "0".

Tab.11 The correlation matrix of group {A} valid for the depth points

A =

	1	**2**	**12**	**13**
1	1	-0.854	-0.860	**0.885**
2	-0.854	1	**0.548**	-0.991
12	-0.860	**0.548**	1	-0.559
13	**0.885**	-0.991	-0.559	1

Tab.12 The denoted successions and subgroups {A_1} and {A_2} for the depth points

Succession	**Group**
1,13	A_1 = { 1,13 }
2,12	A_2 = { 2,12 }
12,2	
13,1	

Tab.13 The correlation matrix of group {B} valid for the depth points

B =

	4	**5**	**9**	**10**	**11**
4	1	**0.931**	-0.739	0.907	-0.714
5	**0.931**	1	-0.503	0.861	-0.562
9	-0.739	-0.503	1	-0.814	**0.752**
10	**0.907**	0.861	-0.814	1	-0.818
11	-0.714	-0.562	**0.752**	-0.818	1

Tab.14 The denoted successions and subgroups {B_1} and {B_2} for the depth points

Succession	**Group**
4,5,10	B_1 = { 4,5,10 }
5,4	B_2 = { 9,11 }
9,11	
10,	
11,9	

Tab.15 The correlation matrix of group {C} valid for the depth points

C =

	3	**6**	**7**	**8**	**14**
3	1	-0.846	**0.522**	0.051	-0.996
6	-0.846	1	-0.830	0.430	**0.836**
7	**0.522**	-0.830	1	-0.441	-0.501
8	0.051	**0.430**	-0.441	1	-0.072
14	-0.996	**0.836**	-0.501	-0.072	1

Tab.16 The denoted successions and subgroups {C_1} and {C_2} for the depth points

Succession	**Group**
3,7	C_1 = { 6,8,14 }
6,8,14	C_2 = { 3,7 }
7,3	
8,	
14,6	

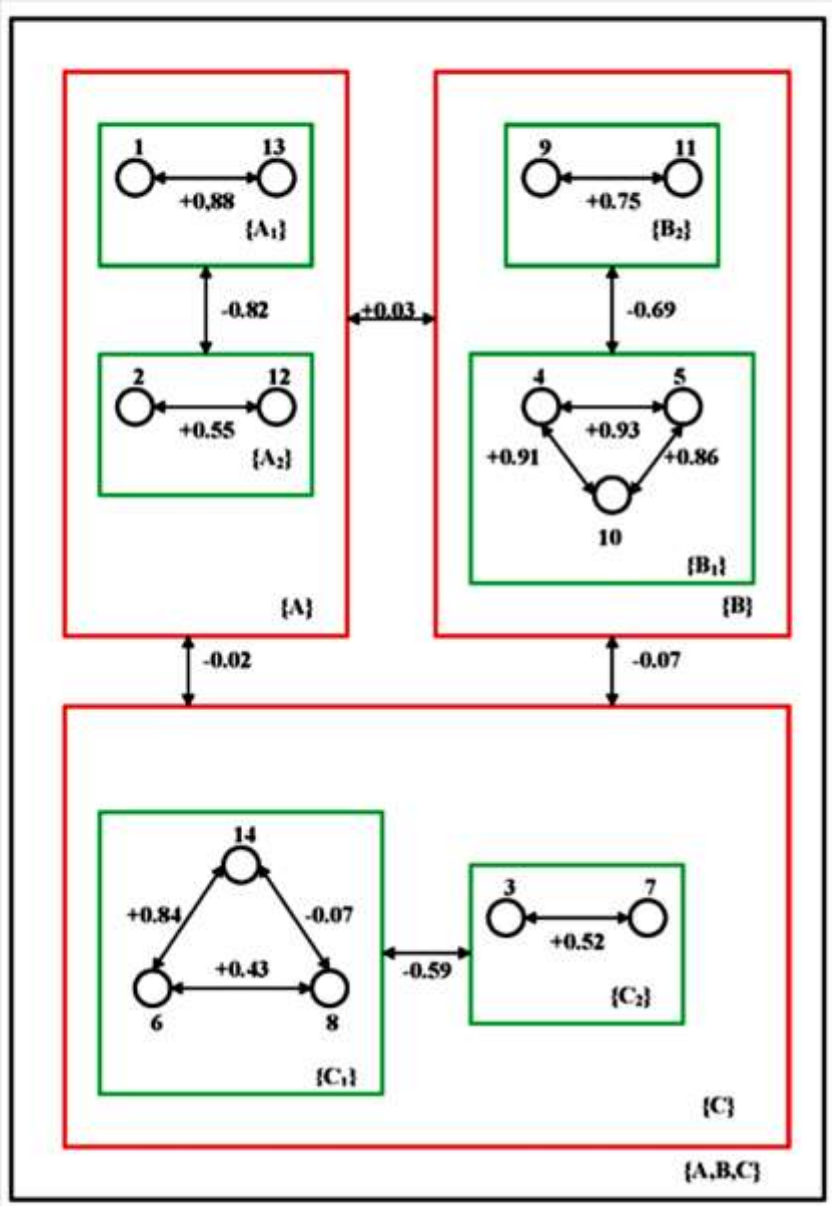

Fig.2 The map of relations between the depth points

r(A_1,A_2)	-0.816
r(B_1,B_2)	-0.692
r(C_1,C_2)	-0.594

Tab.17 The table of correlation coefficients **r (A_1, A_2), r (B_1, B_2)** ***and*** **r (C_1, C_2)**

Tab.18 The input characteristics as data of well logs depicted in percentage

i	**R** [%]	**χ** [%]	**σ** [%]	**ρ** [%]	**ε** [%]
1	2.2	3.7	0.0	6.3	12.2
2	7.1	5.9	1.2	1.8	3.2
3	0.9	10.4	0.0	0.7	3.1
4	2.9	11.8	0.7	3.8	7.5
5	6.0	11.1	2.3	7.6	7.5
6	13.7	4.1	39.5	21.7	13.9
7	10.8	9.6	0.0	0.0	5.4
8	6.3	8.1	39.3	10.2	9.0
9	9.0	4.3	3.7	8.9	3.3
10	2.8	11.0	1.2	6.8	9.3
11	20.1	3.0	1.5	8.1	4.9
12	6.8	9.6	7.5	7.7	4.4
13	3.0	5.9	2.0	8.2	9.0
14	8.3	1.5	1.2	8.2	7.3
Sum	100.0	100.0	100.0	100.0	100.0

Tab.19 Interpreted subgroups of rocks after the depth points

	Interpretation of groups
A_1	Rock being high electrically polarized and conductive, weak radioactive and weak magnetic with higher density
A_2	Rock being weak electrically polarized and non-conductive, having variable radioactivity and magnetic susceptibility, with low density
B_1	Rock being high electrically polarized and conductive, weak radioactive and high magnetic, having variable density
B_2	Rock being weak electrically polarized and non-conductive, weak radioactive and magnetic with high density
C_1	Rock being high electrically polarized , weak magnetic, having high variable radioactivity, high density and having variable conductivity
C_2	Rock being weak electrically polarized, non-radioactive, high magnetic, having low density and variable conductivity

Tab20 Data of borehole profile sections 1 and 2

h [m]	Section 1	Section 2
350.0	A	A_1
350.5	A	A_2
351.0	C	C_2
351.5	B	B_1
352.0	B	B_1
352.5	C	C_1
353.0	C	C_2
353.5	C	C_1
354.0	B	B_2
354.5	B	B_1
355.0	B	B_2
355.5	A	A_2
356.0	A	A_1
356.5	C	C_1

Relations between groups {A}, {B}and {C} are extremely low, fig.2. Formations of all three groups could be either massive ores and mafic rocks, or dielectric rocks as classified the analysis of virtual characteristics. With the help of the map of correlation and measured characteristics we can make the classification of subgroups as the types of formations. It is made in tab.19. The formations complete the depth points into groups of equivalent properties. I accent; this is classification made strictly after virtual physical characteristics. However, these types of rocks are only virtual, because to have real well-logging records was not possible. Thanks to the mentioned operation we are close to construction of the borehole profile only after subgroups having been interpreted. Data for the borehole profile you find in tab.20. There exist two sections, 1 and 2, made per the borehole depth. Section 1 is constructed with groups {A}, {B} and {C}and is characterized like global, while section 2 is presented with subgroups {A_1}, {A_2}, {B_1}, {B_2}, {C_1}and {C_2}; is more detailed. Both sections can be compared to the real geological one for the correct identification of the rock types interpreted. This result can be further used for solving geological cycles in the borehole sections. It is visible in section 1.

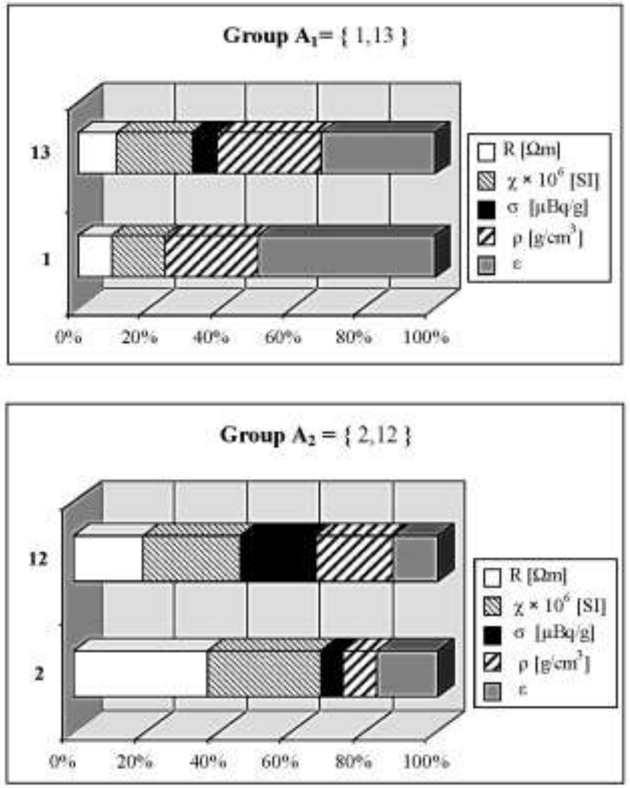

Fig.3 The stripe plot of group A

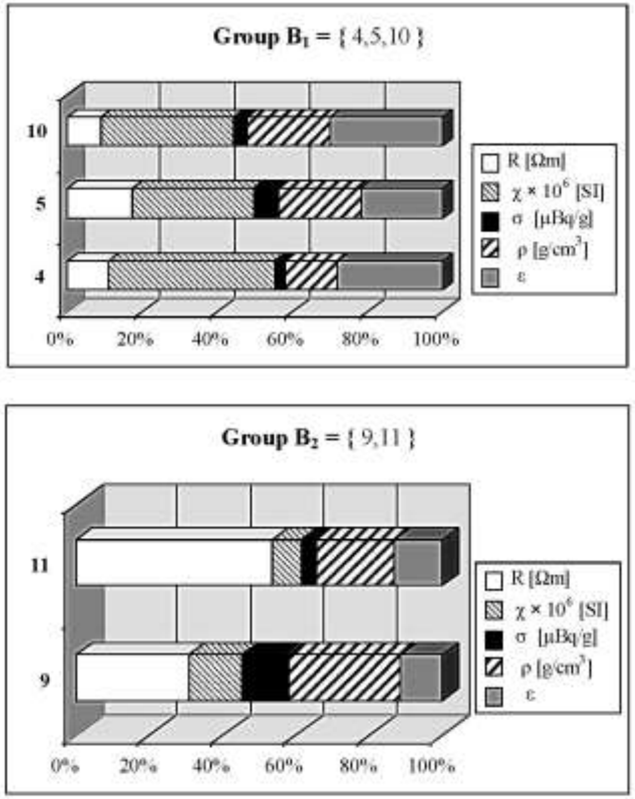

Fig.4 The stripe plot of group B

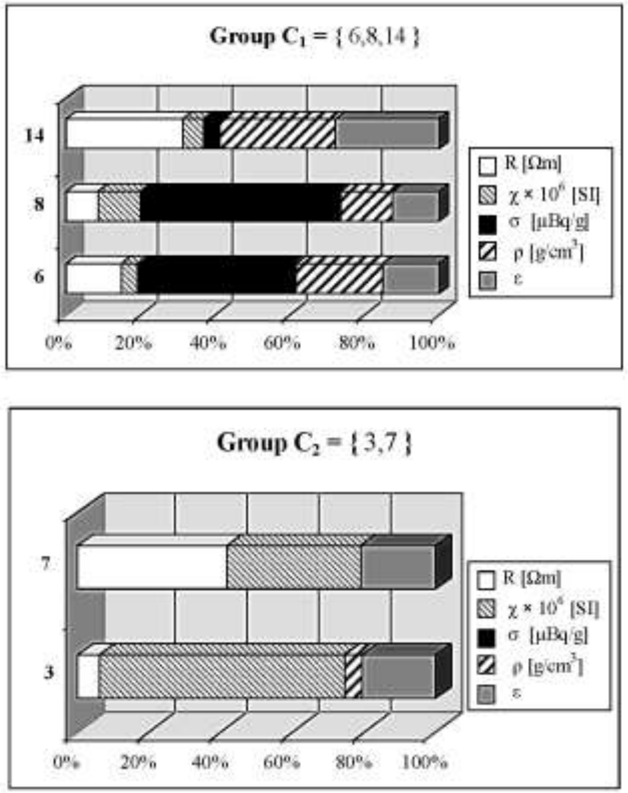

Fig.5 The stripe plot of group C

6 Conclusion

On the base of this analysis, I came to the following conclusions:

- Heuristic algorithm “Class,” published by SIROTINSKAYA, (1986), is well-applied for classification of the registered characteristics as function of the borehole depth into groups.
- The same algorithm is well-usable for numeric values of this paper. It can be used not only for boreholes, but for the surface geophysical profiles, as well. It only must exchange the column of depth for the column of profile length.
- Algorithm uses numeric data for mathematical grouping of maximal values of the correlation coefficients in rows of the correlation matrix.
- Two correlation matrices are studied. Input data are presented with the help of the correlation matrices constructed after the characteristics and after the depth points.
- Groups present the types of formations characterized by their physical properties: R, χ, σ, ρ, ε.

- As the input table of data, Tab.1, is identical to the input table of data in the work RYŠAVÝ, (2023). So, you can gain next added information about examined borehole.

References

CHEN, G., CHEN, M., HONG, G., LU, Y., ZHOU, B., GAO, Y. *A new method of lithology classification based on convolutional neural network algorithm by utilizing string vibration data,* – Energies, 2020 – mdpi.com_

CORINA, A.N., HOVDA, S. *Automatic lithology prediction from well-logging using kernel density estimation,* – Journal of Petroleum Science and Engineering, 2018, – Elsevier

KARPENKO, O., MYRONTSOV, M., ARPILOVA, Y. – … *Application of discriminant analysis in the interpretation of well-logging data,* Decision and Control in Energy III, 2022 – Springer

VUKADIN, D., ČOGELJA, Z., VIDAČEK, R., BRKIČ, V. *Lithology and Porosity Distribution of High-Porosity Sandstone Reservoir in North Adriatic Using Machine Learning Synthetic Well Catalogue,* – Applied Sciences, 2023 –mapi.com

MAREŠ, J. *Matematická logika*, ČVUT Praha, Praha, 2009, p.92

PETKOV, P. P. *Mathematical Logic*, Springer – Verlag New York Inc., New York, 2011, p.420

RACLAVSKÝ, J. *Úvod do logiky: Klasická výroková logika*, Masarykova Universita, Brno, 2015, p. 240

RYŠAVÝ, F. *Component Analysis as a new efficient interpretation method for well-logging data*, EGRSE Journal, vol. XXX, no.2, 2023, p. 59- 81

SIGORSKIJ, V. P. *Matematitcheskyi apparat ingenyera*, Technika, Kyjev, 1975, 753s

SIROTINSKAYA, S. V. *Logitcheskye metody analyza geologitcheskoy informatzii*, Nedra, Moscow, 1986, p. 156

ŠKRÁŠEK, J., TICHÝ, Z. *Základy aplikované matematiky, část 3*, SNTL, Praha, 1986, p. 853

Author

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