Improved recognition of rock formation on the basis of well logging and laboratory experiments results using factor analysis

Edyta Puskarczyk[1] · Jadwiga A. Jarzyna[1] · Kamila Wawrzyniak-Guz[1] · Paulina I. Krakowska[1] · Marcin Zych[1]

Abstract
Several data sets from the Silurian and Ordovician formations from three wells on the shore of Baltic Basin in Northern Poland prepared on the basis of well logging data and results of their comprehensive interpretation were used in factor analysis. The goal of statistical analysis was structure recognition of data and proper selection of parameters to limit the number of variables in study. The top priority of research was recognition of specific features of claystone/mudstone formations predisposing them to be potential shale gas deposits. The identified data scheme based on data from one well, was then applied to: 1) well 2 and well 3 separately, 2) combined data from three wells, 3) depth intervals treated as *sweet spots*, i.e., formations of high hydrocarbon potential. Numbers of samples from well logging were proportional to number of laboratory data from individual formations. The extended data set comprising all available log samples in explored formations was also prepared. Outcomes from standard (Triple Combo—natural gamma log, resistivity log, neutron log and bulk density log and Quad Combo—with addition of sonic log and spectral gamma log) and sophisticated (GEM™—Elemental Analysis Tool, Wave Sonic and Nuclear Magnetic Resonance—NMR) logs were the basis for data sets. Finally, laboratory data set of huge amount of variables from elemental, mineralogical, geochemical and petrophysical laboratory experiments was built and verified in FA to select the most informative components. Conclusions on the data set size, number of factors and type of variables were drawn.

Keywords Rock formation · Petrophysical characteristics · Well logging · Laboratory experiments · Factor Analysis

Introduction

Rock recognition on the basis of well logging supported by petrophysical laboratory experiments is an important part of the qualitative and quantitative geological interpretations. Great number of logs is always the goal in planning borehole investigations to obtain results with uncertainty level as low as possible. Processing of various logs and their comprehensive interpretation sometimes generates technical and petrophysical problems with unambiguous treatment of results because modern well logs provide interpreters with great amount of data. Sophisticated statistical tools are useful in smart data management to get maximum indispensable geological information without problems with/of unambiguity (Asfahani et al. 2005; Kaźmierczuk and Jarzyna 2006; Wawrzyniak-Guz et al. 2016; Puskarczyk 2018).

Factor analysis (FA) belongs to the group of statistical procedures enabling mutual relationships investigations between great numbers of data and revealing hidden relations between unknowns which prove necessity of analysis of selected factors. Factors in analysis have substantive interpretation related to the considered problem and preserve information included in primary variables (Szabó 2011).

The goal of using factor analysis in the presented case was to recognize structure of data and meaningful factors in large numbers of variables originated from well logging and laboratory data. Many crucial decisions in FA were subjective, i.e., interpreters decided on number of factors, rotational method, interpreting loadings; so, in the presented examples the geological and petrophysical knowledge of authors and their experience in the comprehensive interpretation of well logs and laboratory data are very important.

In the paper, there are presented results of factor analysis aimed to make reduction and proper selection of variables

✉ Edyta Puskarczyk
puskar@agh.edu.pl

[1] AGH University of Science and Technology, al. A. Mickiewicza 30, 30-059 Kraków, Poland
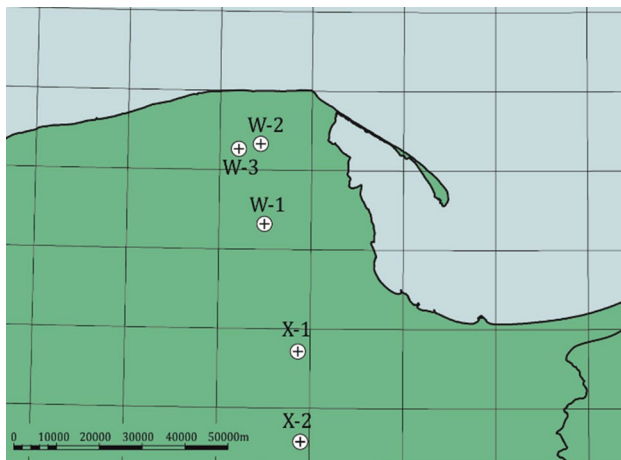
Fig. 1 Schematic map of Northern-Central part of Poland with marked positions of wells in study

in data sets. As training data were used, variables from well logging and laboratory experiments from formations differentiated as regards their lithology, petrophysical (reservoir and elastic) properties and total organic content. Methodological and petrophysical conclusions were discussed in parallel to show that similar analyses may be applied to other geological data. Because well logging provided a huge amount of data (logs were depth-sampled at each 0.1 m), it was possible to make FA on data sets consisting of different number of variables and show that combining limited laboratory results with almost unlimited log data is important in the proper selection of variables.

Materials

Data sets for factor analysis were built of well logs from three boreholes W-1, W-2, and W-3 (Fig. 1), closely located within themselves on the shore of Baltic Basin in Northern Poland. In each well, the Silurian and Ordovician formations were studied at similar depths, with particular emphasis on Ja Mb and Sa Fm considered as the so-called *sweet spots,* shale formations of high hydrocarbon potential. A sequence of geological formations in the study is presented

in Table 1. Close proximity of wells and similar sampling depths were the reasons enabling combining variables into numerous data sets. Variables represented quantities logged in wells and results of their processing and interpretation. Well-log data comprised standard curves, i.e., resistivity in ohm-m (LLD, LLS), natural radioactivity (standard gamma ray log—GR [API] and spectral gamma ray—POTA [%], URAN [ppm], THOR [ppm] and total signal from three energetic windows—GRTO [API] and sum of potassium and thorium windows—GRKT [API]), acoustic (transit interval time from BHC device—DT [μs/m], DTP [μs/m] and DTSX [μs/m]—transit interval times of P and S waves, respectively, from the modern full-wave sonic instrument—Wave Sonic device), caliper with subtracted bit size (DCAL [m]), bulk density—RHOB [g/cc] and photoelectric absorption index—Pe [barn/electron], measured by spectral gamma–gamma tool. Results of the GEM log interpretation were also included as follows: organic carbon content—DKER [wt%], volume of water bound in clay minerals—VCBW, volume of free water—VWF, volume of gas—VGAS, PHIE—porosity comprising VWF + VGAS and PHIG comprising PHIE + VCBW. Lithological components from ULTRA program (Halliburton), i.e., volume of shale, sandstone, limestone, dolomite, pyrite and kerogen—VSH, VKWA, VLIM, VDOL, VPIR and VKER, respectively, and porosities, total—PHI and effective—PHIE, worked also as variables. All volumes were presented in fractures. In the analyses, transit interval times of P—DPEQ and S—DSEQ waves and bulk density—RHEQ estimated from Biot–Gassmann model were also included (Bała and Cichy 2007). Lists of variables in each of the data sets were different depending on the range of measurements and interpretation made in individual wells. Depth, H was included into FA, but this variable turned out to be not informative. (The considered depth range was similar in the three wells.)

Laboratory measurements results obtained from selected sections of the Silurian and Ordovician formations in three boreholes (66 samples) constituted the other data sets for factor analysis (Jarzyna J et al. 2017; Jarzyna and Wawrzyniak-Guz 2017). Cored depth intervals in boreholes in study were between 2870 and 3235 m. Majority of samples were composed of claystones/mudstones. Several sandstones

Table 1 Stratigraphy and lithology of geological formations (after Jarzyna and Wawrzyniak-Guz 2017)

Formation	Stratigraphy	Type of rock
Pe Fm	Silurian, Ludlow, Wenlock	Claystone/mudstone
Pa Fm	Silurian, Llandovery	Claystone/mudstone
Ja Mb (bottom of Pa Fm)	Silurian, Llandovery	Bituminous claystone
Pr Fm	Ordovician, Ashgillian	Marl, Limestone, Claystone
Sa Fm	Ordovician, Caradocian, Llanvirnian	Bituminous claystone, claystone/mudstone
Ko Fm	Ordovician, Llanvirnian, Llandeilian	Limestone

and tuffs were also included. Samples from the claystone/mudstone (Pe Fm, Pa Fm) and calcareous (Pr Fm, Ko Fm) formations were considered with independent approach to the Ja Mb at the bottom of Pa Fm and Sa Fm built of bituminous claystones, which were treated as *sweet spots*, i.e., potentially prospective shale gas beds. Data set was composed of laboratory outcomes from Mercury Injection Porosimetry, Helium Porosimetry, special total porosity measurements—water and kerosene immersion porosimetry (Topór et al. 2016), Nuclear Magnetic Resonance experiments, Rock Eval geochemical measurements, Nitrogen Absorption/Desorption Method, Pressure Decay Permeability method. Results of the elemental analyses and mineral components interpretation made on the same geological samples were also available. Variables finally selected for factor analyses were described in detail in subchapter Laboratory data set in factor analysis.

Because numerical amount of well logging and laboratory data sets was different, and there were different proportions between laboratory data from various formations, we randomly selected numbers of well logging samples proportional to laboratory data in the way described in Table 2. Well logging and laboratory data were selected from the same depth intervals.

Most of the presented FA results were obtained using randomly selected well logging data sets from W-1, W-2 and W-3 wells (limited data sets). Only results presented in the *factor analysis on extended data set from W-1, W-2 and W-3 wells comprising large number of samples* subchapter were determined on the extended well logging data set. Our aim was to check whether the limited number of data influenced the results of the FA.

Methodology of factor analysis (FA)

Analyses were made using Statistica 13.3 software (AGH UST Licence 2018). Studies were carried for the total data set comprising samples from the geological formations: Pe Fm, Pa Fm including Ja Member at the bottom, Pr Fm, Sa Fm and Ko Fm from three wells (Fig. 1, Table 1). Analyses for log data from individual boreholes comprising all the above-mentioned formations and for Ja Mb and Sa Fm from three wells combined together were also performed. Separately, data set comprising laboratory data was constructed. It was not too numerous, so was processed as a body.

The following algorithm of analyses was assumed for all data sets. Variables were standardized using average values and standard deviation, so FA worked on the correlation matrix or equivalently the standardized variables of variances equal to 1. All sets of samples were treated as raw data in Statistica 13.3 software. Multiple regression was adopted as a method for data analyses. Maximum Likelihood factors were selected as the most suitable for analyses of geological properties. The available normalized rotation techniques were tested: Varimax, enabling minimization of variables number with high factor loadings and simplifying the factors interpretation; Quartimax, allowing minimization of factors number necessary to explain each variable and simplifying interpretation of the observed variables; and Equamax, being the combination of both upper-mentioned techniques making easier interpretation of factors and variables interpretation. Rotation types were considered also as regards the number of factors to be retained to rotation and interpreted. At the beginning, the maximum number of factors was adopted as 10 and the minimum eigenvalues equaled to 0. In the next steps, number of factors was lowered to 6 or 5. In the spreadsheets of eigenvalues, percentage of total variance, cumulative eigenvalues, and cumulative percentage were analyzed. According to the Kaiser criterion (Kaiser 1958), there were retained factors with an eigenvalue greater than 1. Scree plots were also analyzed. The point where the continuous drop in eigenvalues levels suggested the cutoff (elbow method) was considered. On the basis of the two presented criteria, the numbers of factors useful in the next steps of analyses and being interpretable were adopted. Factor loadings were interpreted as the correlations between the input variables and factors, and represented the most important information for interpretation. The first factor showed most of the highest loadings, and successive factors accounted for less and less ones. The signs of the factor loadings showed only the way that variables with opposite loadings on the same factor relate the data to the factors by inverse proportionality. Firstly, the data sets were processed without any

Table 2 Proportions between numbers of laboratory and well logging data sets

Data	Laboratory				Well logging: available/randomly selected			
Wells	W-1	W-2	W-3	Total	W-1	W-2	W-3	Total
Total	23	25	18	66	865/423	1062/604	1145/508	2920/1555
Pa Fm	4	0	1	5	475/56	450/136	680/43	1605/319
Ja Mb	5	5	7	17	140/140	120/120	130/130	390/390
Pr Fm	2	0	4	6	55/32	80/24	80/80	215/136
Sa Fm	12	20	6	38	195/195	260/260	255/255	710/710

rotation. Finally, standardized Equamax rotation was applied (Hair et al. 2006).

Results

Firstly, factor analysis was performed on the well-log data set from W-1 well to adopt proper algorithm for the main analyses. Data set from W-1 well was selected because it comprised samples from considered formations (Pa Fm, Ja Mb, Pr Fm and Sa Fm). Total number of well logging samples, proportional to the number of laboratory outcomes in the same depth, was relatively high. The proportion well projected relationships between samples from considered formations.

W-1 well logging data set in factor analysis

At the beginning, 10 factors and the minimum eigenvalue equal to 0 were assumed in the analyses. Results for the Principal Components (left) and Maximum Likelihood factors (right) as factor extraction methods are presented in Table 3. The included spreadsheets consist of the eigenvalues (EV), percentage of total variance (%TV), cumulative eigenvalues (CEV), and cumulative percentage of explained variance (CV).

Selecting finally the Maximum Likelihood factoring, we assumed the known number of factors (10, 6 or 5). Then, Statistica 13.3 software estimated the loadings and communalities that maximized the probability of the observed correlation matrix. Chi-square tests of the goodness of fit were available (Statistica 13.3 Help 2019) and analyzed. Costello and Osborne (2005) stated that if "data are relatively normally distributed, Maximum Likelihood factors are the best choice because they allow for the computation of a wide range of indexes of the goodness of fit of the model and

permit statistical significance testing of factor loadings and correlations among factors and the computation of confidence intervals." So, despite lower cumulated percentage of explained variance in the case of Maximum Likelihood factors, we decided to apply this choice for geology origin (normal distributed) data. We selected the Principal Components to compare to Maximum Likelihood Factors because PCA method and FA are similar and PCA is frequently used as a data reduction tool. PCA is a quicker alternative to FA but is computed without regard to structure caused by latent variables. Components are calculated using all of the variance of the variables, and all of that variance appears in the solution. During factor calculation, the shared variance of a variable is partitioned to unique and error variances to reveal the underlying factor structure and only shared variance appears in the solution. PCA does not discriminate between shared and unique variance. Since FA only analyzes shared variance, it yields the same solution and also avoids the inflation of estimates of variance accounted (Costello and Osborne 2005).

According to the scree plot and Kaiser Criterion, 6 or 5 factors were adopted in the next analyses. Scree plot for the analyzed case of 10 Maximum Likelihood factors from W-1 well is presented in Fig. 2.

Explained variance in the Principal components and Maximum Likelihood factors cases is visible in the lowest right cells in Table 3 (95,96—PC and 87,53—ML). Value for ML is lower than that for PC choice but satisfactory. Factor loadings in two cases (10 factors—left and 5 factors—right) for the normalized Equamax rotation and Maximum Likelihood factoring are compared in Table 4. Loadings higher than 0.7 are red marked. In the presented solutions, one variable is explained by one factor with correlation higher than 0.7.

The first factor, F1 in both parts of Table 4 transmits the information included in the NPHI, DT, DTP, DTSX, DPEQ variables. Quantities DT, DTP, DTP and DPEQ are related to elastic properties of rock, all of them are provided by

Table 3 Comparison of results for principal components and maximum likelihood factors analyses

| No.	EV	TV (%)	CEV	CV	No.	EV	TV (%)	CEV	CV
		10 factors, min EV = 0, principal components					10 factors, min EV = 0, maximum likelihood factors		
1	10.52	31.88	10.52	31.88	1	9.86	29.87	9.86	29.87
2	7.87	23.86	18.39	55.73	2	7.37	22.33	17.23	52.20
3	5.29	16.04	23.68	71.77	3	5.05	15.32	22.28	67.52
4	2.52	7.62	26.20	79.40	4	2.29	6.95	24.57	74.47
5	1.41	4.26	27.61	83.66	5	1.37	4.14	25.94	78.61
6	1.16	3.52	28.77	87.18	6	0.97	2.94	26.91	81.56
7	0.98	2.97	29.75	90.15	7	0.82	2.48	27.73	84.04
8	0.87	2.65	30.62	92.79	8	0.51	1.55	28.25	85.59
9	0.61	1.85	31.23	94.65	9	0.35	1.06	28.60	86.65
10	0.44	1.32	31.67	95.96	10	0.29	0.88	28.89	87.53
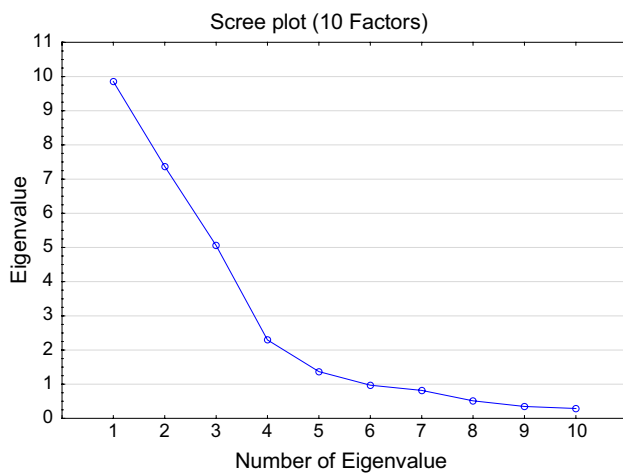
Fig. 2 Scree plot for well-log data, W-1 well

acoustic log. In the next analyses, only one of these variables, characterizing the highest loading, was used. The second factor, F2 in both parts of Table 4 provided the information carried by the variables: GRKT, POTA, THOR, VCBW. All listed quantities are parameters measured in gamma ray logs and are indicators of clay mineral presence in rock formation. The third factor, F3 is related to GR, GRTO, URAN and VKER. These variables are also indicators of natural radioactivity of rocks but mostly in the part responsible for uranium content and organic matter represented by volume of kerogen (parameter from the comprehensive interpretation). In the 10-factor solution, F4 and F5 did not carry important information at all, F6 transmits information about apparent resistivity from LLS and LLD devices, F7 presents high correlation to VDOL, volume of dolomite. F8-F10 did not provide essential information. Information on effective porosity (PHIE) was transferred in the first factor F1 (10-factors solution) and in the fourth factor F4 (5-factors solution). High correlation to volume of quartz visible in F5 in the last mentioned case is similar to volume of dolomite in the previous one. It is not crucial in the discussed case, because credible information about mineral content can be obtained from other geological sources. Table 4 shows that many of the variables included in the analysis were not informative. Firstly, in this group, results of the comprehensive interpretation: PHI, VWF, VSH, VLIM, VPIR, should be pointed out. H, DCAL, DKER were also not useful.

W-3 well data set in factor analysis

Experience gained during the FA on W-1 well data set was used in the next two wells. Geological profile in W-3 borehole comprised the same formations like in W-1. Quantities from all available logs were used as variables. Maximum Likelihood factoring and normalized Equamax rotation were

adopted. The comparison of results was obtained for 10- and 5-factor solutions. Scree plot is presented in Fig. 3. Test of the goodness of fit showed 77.31% of explained variance for the 5-factor solution.

Comparison of factor loadings obtained in two solutions (10 and 5 factors) is presented in Table 5. It is modified compared to Table 4. Factors F9 and F10 in 10-factor solution were removed because they did not transport information about variables. There were also removed variables (in both parts of Table 5) because correlation coefficients between them and any factor are below 0.7. Comparing Tables 4 and 5, similarities and differences are visible. In Table 5, factors F1 in both solutions transmitted information about NPHI, GRKT, POTA, THOR, URAN and VCBW on the same level. Elastic properties indicators: DPEQ and DSEQ, are also included to the group of variables related to F1. Factor F1 transmitted information on NPHI and DPEQ variables similarly as in W-1 well data set (Table 4). In W-1 well data analysis, F1 factor transmitted information of group of porosity indicators in the 10-factor solution. In W-3, well data analysis porosity representatives (PHI and PHIT) were present in F3 in 10-factor solution and in F4 in 5 factors one. In Table 5, it is visible that no factors carried information about mineral composition. Only, VSH was included in F1 and F2 (in 10-factor analysis) and F3 (in 5 factors one) carried information on VKER. Generally, lower number of factors caused more consistent representation between factors and variables.

Factor analysis on common data set from W-1, W-2 and W-3 wells, number of samples proportional to number of laboratory results

FA on the data set from three wells comprising all analyzed geological formations was performed using 10 factors and 6 factors. In 10-factor solution, F8 and F10 did not carry any information on variables. Factors F3, F4 and F9 were transmitted information only on single variables: VLIM, VDOL and VPIR, respectively, while F1 explained 5 variables and F2—9, F7—3, F6—2 in 10-factor solution. In the 5-factor solution, F1 transmitted information on 3 variables, F2—9, the same as in previous one, F3—3, F4—2. Table 6 illustrates the obtained results. Similar to Table 5, only rows and columns with significant representation of quantities are presented. Now, cells with loadings lower than |0,20| are empty.

The first factor, F1 in 10-factor solution in three wells data set is similar to F1 in W-1 well data set (Table 4). F1 transmitted the information carried by density (RHOB, RHEQ) and acoustic (DT, DTP) logs and porosity (PHI, PHIE, PHIT, PHIG), F2—GRKT, POTA, THOR, VCBW. In three wells data set, F2 transmitted also the information

Table 4 Factor loadings in two solutions (10 factors and 5 factors). W-1 well

Var	10 factors									5 factors					
	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F1	F2	F3	F4	F5
H	0.08	−0.17	0.05	−0.68	0.03	0.12	0.02	0.03	0.12	−0.10	−0.13	−0.06	0.04	0.17	−0.62
LLS	0.02	−0.24	0.35	0.00	0.00	−0.87	−0.01	0.01	0.00	−0.05	−0.02	−0.23	0.61	−0.12	−0.10
LLD	−0.05	−0.22	0.36	−0.04	0.00	−0.87	0.02	−0.01	−0.01	−0.08	−0.08	−0.21	0.62	−0.15	−0.11
DT	0.85	0.07	0.19	0.25	0.21	−0.04	0.15	−0.03	−0.05	0.00	0.91	0.01	0.23	0.12	−0.01
RHOB	−0.73	0.32	−0.32	0.29	−0.07	0.20	−0.05	−0.04	−0.03	0.10	−0.53	0.28	−0.45	−0.31	0.45
Pe	−0.47	0.43	−0.13	0.48	−0.43	0.14	−0.02	−0.02	0.00	−0.05	−0.34	0.35	−0.28	0.01	0.70
NPHI	0.71	0.38	−0.01	0.40	0.10	0.08	0.17	−0.04	0.13	−0.09	0.85	0.25	−0.07	0.12	0.24
GR	0.18	0.42	0.83	0.03	0.05	−0.08	0.00	0.01	0.00	−0.02	0.20	0.55	0.74	0.09	0.08
GRKT	0.04	0.95	0.20	0.02	0.04	0.05	0.00	0.08	−0.05	−0.08	0.16	0.95	0.01	0.01	0.11
GRTO	0.12	0.47	0.84	−0.01	−0.03	−0.08	−0.03	0.00	0.02	0.00	0.11	0.61	0.73	0.14	0.08
POTA	0.03	0.90	0.26	0.04	0.11	0.02	0.03	0.09	−0.07	−0.12	0.17	0.92	0.09	−0.07	0.09
THOR	0.06	0.94	0.09	0.02	−0.08	0.10	−0.05	0.00	−0.02	0.00	0.12	0.93	−0.11	0.13	0.13
URAN	0.15	0.02	0.95	−0.04	0.00	−0.15	−0.02	−0.04	0.04	0.04	0.08	0.20	0.93	0.13	−0.01
DCAL	0.01	0.38	−0.42	0.27	−0.01	0.06	0.04	0.08	0.19	0.08	0.15	0.24	−0.48	−0.05	0.29
DTP	0.81	−0.02	0.20	0.29	0.22	−0.09	0.18	−0.04	−0.06	0.01	0.89	−0.08	0.27	0.07	0.01
DTSX	0.85	0.14	−0.12	−0.02	0.05	0.13	0.07	0.00	0.24	0.00	0.78	0.05	−0.14	0.36	−0.15
DKER	0.48	−0.31	0.46	0.15	0.14	−0.49	0.11	−0.02	−0.11	−0.07	0.47	−0.30	0.65	−0.02	−0.08
VCBW	−0.11	0.89	−0.09	0.15	−0.06	0.08	0.06	−0.25	0.01	0.02	0.05	0.82	−0.28	−0.07	0.24
VWF	0.51	0.09	−0.36	−0.34	−0.16	0.39	−0.05	−0.02	0.47	0.14	0.30	0.03	−0.46	0.52	−0.25
VGAS	0.74	−0.01	0.42	−0.08	−0.27	−0.10	−0.17	0.26	−0.18	0.09	0.47	0.02	0.45	0.63	−0.13
PHIE	0.83	0.03	0.18	−0.22	−0.29	0.10	−0.16	0.20	0.07	0.13	0.51	0.03	0.16	0.76	−0.21
PHIG	0.71	0.52	0.12	−0.12	−0.31	0.13	−0.12	0.04	0.07	0.14	0.50	0.47	−0.01	0.67	−0.07
PHI	0.68	−0.37	−0.25	−0.37	−0.14	0.00	−0.22	−0.19	−0.15	−0.17	0.34	−0.42	−0.14	0.50	−0.51
PHIT	0.85	−0.11	−0.31	−0.15	−0.03	0.07	−0.13	−0.19	−0.11	−0.15	0.63	−0.21	−0.24	0.44	−0.36
VSH	0.25	0.65	−0.08	0.53	0.28	0.15	0.24	0.02	0.10	0.07	0.59	0.53	−0.21	−0.25	0.43
VLIM	−0.51	−0.09	−0.31	0.14	−0.70	0.05	0.17	−0.01	0.04	−0.22	−0.54	−0.15	−0.35	0.11	0.40
VDOL	−0.44	−0.17	0.08	0.01	0.08	0.05	−0.82	0.01	0.01	−0.06	−0.59	−0.08	0.08	0.08	0.06
VKWA	0.29	−0.48	0.06	−0.68	0.23	−0.19	0.20	−0.02	−0.16	0.20	0.11	−0.40	0.21	0.01	−0.78
VPIR	−0.05	−0.06	−0.04	0.02	0.04	0.06	0.02	0.00	0.01	0.43	−0.04	−0.07	−0.04	−0.01	0.05
VKER	0.24	0.00	0.90	−0.01	0.08	−0.19	0.00	0.05	0.02	−0.05	0.20	0.16	0.92	0.09	−0.03
DPEQ	0.71	0.38	0.11	0.23	0.30	0.05	0.34	0.08	0.04	0.01	0.90	0.31	0.07	−0.04	0.03
DSEQ	0.53	0.51	0.05	0.50	0.10	0.07	0.31	0.11	0.08	−0.08	0.78	0.39	−0.03	−0.04	0.39
RHEQ	−0.79	0.27	−0.29	0.29	0.02	0.15	−0.10	0.09	0.13	0.13	−0.57	0.24	−0.40	−0.36	0.46

Var—variable. F1, …, F10—factors

carried by NPHI, while in W-1 well data set NPHI is represented by F1. Considering together the presented results, we can say that two factors transmitted essential information on rock formation provided by logs.

Factor analysis on extended data set from W-1, W-2 and W-3 wells comprising large number of samples

Factor analysis was also carried out for a larger data set, built of all available log samples in three wells. Different results were obtained. Extended data set comprised apart from Pa Fm with Ja Mb at bottom, Pr Fm and Sa Fm samples data from Pe Fm and Ko Fm (Table 1) because in laboratory

data sets there were samples from these formations. Only Ja Mb and Sa Fm parts of the extended data set consisted of the same number of samples comparing to the previous analyses. Numerous sets of Pe Fm and Pa Fm samples and also quite big representation of Pr Fm and Ko Fm were available. Two last formations are built of marls and carbonates, while Pe Fm is similar in lithology to Pa Fm. Results are presented in Table 7.

The first factor, F1, transmitted information carried by variables based on elastic properties (DT, DTP, DTSX, DPEQ, DSEQ). It can be explained by differences in velocity of elastic waves in mudstone/claystone formations (Pe Fm, Pa Fm and Sa Fm) and calcareous/limestone
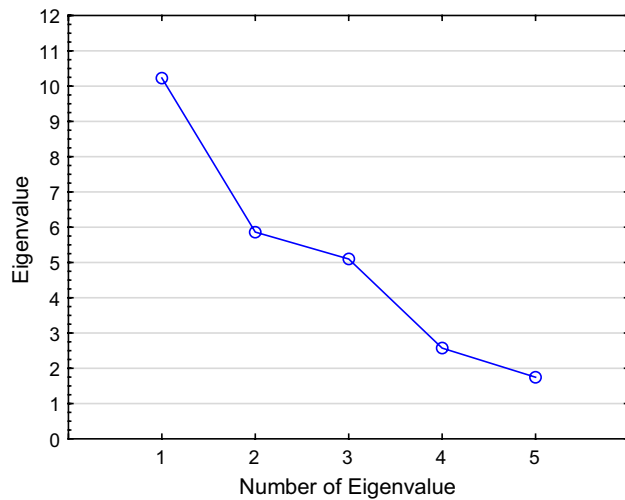
Fig. 3 Scree plot, W-3 well, 5 factors

(soft) and carbonate (hard) sections. F1 transmitted also information from GR log and VSH, which means that the difference in clay minerals volume in various formations may be sufficiently explained on the basis of these variables loading F1. F2 transmitted information carried by resistivity logs, which is differentiating low-resistivity claystone/mudstone rocks and high-resistivity carbonates. Information on organic matter presence in *sweet spots* is visible in F2 transmitting the variables DKER and VKER. Lithology differentiation is also marked in F1 transmitting information included in VLIM (calcareous formations) and VSH (claystone/mudstone formations). Volume of quartz is visible only just in F4. Table 7 shows that in the discussed case, lithology differences are presented by two first factors. In the discussed solution, differences in lithology are distinct, so they played an important role in FA. Porosity indicators (PHI and PHIE) are included in F3. Summarizing, we can say that two or three factors transmitted sufficient information. Standard parameters from logs are useful and sufficient for simple petrophysical interpretation. Comparing the results presented in Tables 6 and 7, outcome obtained in 5-factor solution in data set

formations (Pr Fm and Ko Fm). What is interesting is that in the discussed solution, DCAL is present. It can be also explained by differences in borehole diameters in claystone

Table 5 Factor loadings in two solutions (10 factors and 5 factors), W-3 well

Var	10 factors								5 factors				
	F1	F2	F3	F4	F5	F6	F7	F8	F1	F2	F3	F4	F5
LLS	−0.19	0.44	0.03	−0.13	−0.16	−0.05	−0.82	−0.06	−0.22	−0.21	−0.67	0.05	−0.18
LLD	−0.22	0.39	0.01	−0.12	−0.18	−0.05	−0.82	−0.06	−0.25	−0.24	−0.63	0.02	−0.19
RHOB	−0.01	−0.64	−0.53	0.41	−0.16	−0.05	0.20	0.04	−0.03	−0.24	0.76	−0.51	0.13
NPHI	0.79	0.31	0.37	0.01	0.08	−0.05	−0.04	−0.16	0.80	0.12	−0.27	0.33	−0.10
GR	0.55	0.78	−0.09	0.02	0.07	−0.01	−0.04	0.10	0.62	0.14	−0.68	−0.16	0.19
GKUT	0.56	0.78	−0.11	0.03	0.07	0.01	−0.03	0.10	0.63	0.13	−0.68	−0.18	0.21
GRKT	0.94	0.02	−0.11	0.06	0.05	0.02	0.07	0.10	0.94	0.01	0.08	−0.10	0.06
POTA	0.90	0.08	−0.08	0.07	0.05	0.06	0.06	0.20	0.90	0.01	0.01	−0.07	0.15
THOR	0.92	−0.07	−0.15	0.03	0.05	−0.02	0.09	−0.05	0.92	0.02	0.16	−0.13	−0.09
URAN	0.11	0.96	−0.07	0.01	0.06	0.00	−0.09	0.05	0.20	0.16	−0.89	−0.16	0.21
DKER	0.03	0.65	0.16	−0.17	−0.29	−0.01	−0.48	−0.07	0.04	−0.28	−0.81	0.15	−0.11
VCBW	0.86	−0.04	0.01	0.15	−0.05	0.00	0.09	0.03	0.86	−0.07	0.15	0.00	0.08
VGAS	0.05	0.36	0.14	−0.05	0.83	−0.04	−0.04	−0.08	0.08	0.79	−0.32	0.05	−0.06
PHIE	0.03	0.09	0.26	0.02	0.92	0.02	0.13	0.02	0.06	0.95	0.02	0.19	0.04
PHIG	0.30	0.07	0.25	0.08	0.86	0.02	0.16	0.03	0.34	0.89	0.07	0.18	0.07
PHI	−0.23	−0.08	0.91	0.02	0.23	−0.06	0.01	0.05	−0.24	0.31	0.04	0.87	0.09
PHIT	0.12	−0.10	0.92	0.02	0.21	−0.08	0.03	0.04	0.11	0.27	0.08	0.90	0.06
VSH	0.93	−0.03	−0.17	0.01	−0.12	−0.04	0.03	−0.02	0.92	−0.17	0.10	−0.13	−0.10
VLIM	−0.30	−0.29	−0.18	−0.16	0.00	0.20	−0.07	−0.82	−0.34	−0.02	0.18	−0.19	−0.70
VDOL	−0.53	−0.02	0.06	−0.09	−0.07	−0.79	−0.11	−0.02	−0.55	−0.09	−0.07	0.06	−0.07
VKWA	−0.41	0.04	0.19	0.08	0.09	0.43	0.10	0.73	−0.37	0.11	0.00	0.20	0.62
VPIR	0.02	−0.15	−0.25	0.84	0.12	0.09	0.22	0.31	0.07	0.13	0.42	−0.35	0.67
VKER	0.17	0.94	0.00	0.02	0.13	0.00	−0.08	0.07	0.25	0.23	−0.87	−0.10	0.24
DPEQ	0.88	0.26	0.17	−0.19	0.03	0.15	0.00	0.03	0.89	0.04	−0.24	0.19	−0.09
DSEQ	0.81	0.17	0.05	−0.28	−0.02	0.01	−0.10	−0.35	0.80	−0.03	−0.22	0.08	−0.49

Table 6 Factor loadings in two solutions (10 factors and 5 factors), W-1, W-2 and W-3 wells

Var	10 factors					5 factors				
	F1	F2	F3	F4	F5	F1	F2	F3	F4	F6
DT	0.75	0.47					0.69	0.33	−0.51	
RHOB	−0.73								0.59	
NPHI	0.60	0.71					0.87		−0.34	
GR		0.57					0.63			0.20
GRKT		0.72					0.74		0.23	
GRTO		0.80					0.73		0.21	
POTA		0.90					0.83			
THOR		0.92					0.82	−0.25	0.21	
URAN						0.26	0.27	0.86	0.15	
DTP	0.73	0.42					0.65	0.36		
DKER	0.44							0.81		
VCBW		0.89					0.77	−0.31		
VGAS	0.36					0.71		0.29	−0.26	
PHIE	0.46				0.86	0.86			−0.45	
PHIG	0.37				0.80	0.78	0.47		−0.34	
PHI	0.81			0.24					−0.94	
PHIT	0.90			0.20					−0.96	
VSH		0.92					0.92	−0.23		
VLIM		−0.26	−0.78				−0.36	−0.23	0.30	−0.74
VDOL		−0.41		−0.87			−0.50			
VKER						0.29	0.31	0.86		0.21
DPEQ		0.76					0.91		−0.23	
DSEQ		0.80	−0.23				0.90			
RHEQ	−0.82					−0.20	−0.14	−0.62	0.71	

Table 7 Factor loadings in extended data set from W-1, W-2 and W-3 wells

Var	F1	F2	F3	F4	F5	F6
LLS		0.89				
LLD		0.86				
DT	0.87	0.23	0.23			0.22
Pe	−0.34			0.73		−0.40
NPHI	0.92			0.06		−0.28
GR	0.74	0.25		−0.28		
GKUT	0.36			−0.27	0.39	0.75
POTA	0.51					0.75
DCAL					0.87	
DTP	0.86	0.24	0.20			0.23
DTSX	0.83		0.36			
DKER	0.37	0.83				
VWF	0.06	−0.38	0.71			
PHIE	0.15		0.96			
PHI	0.26		0.74		0.27	
VSH	0.85	−0.33	−0.33			
VLIM	−0.72			0.58		
VKWA				−0.88		
VKER	0.31	0.72		−0.34		0.24
DPEQ	0.94					
DSEQ	0.86			0.37		

proportional to number of laboratory data turned out to be the best, distinctly combining factors with variables responsible for petrophysical information.

Factor analysis results in sweet spots, Ja Mb and Sa Fm

Ja Mb occurring at the bottom of Pa Fm and Sa Fm were selected as *sweet spots* in the analyses carried out to characterize shale gas potential of the Silurian and Ordovician formation in the Baltic Basin (Jarzyna and Wawrzyniak-Guz 2017). Special data sets from the log samples in those formations in three wells were prepared. Selected results of factor analysis performed on Ja Mb and Sa Fm data sets as regards spreadsheets consisting of the eigenvalues (EV), percentage of total variance (%TV), cumulative eigenvalues (CEV), and cumulative percentage of explained variance (CV) are presented in Table 8.

Scree plots are included in Fig. 4. According to Kaiser criterion and scree plot, 6 factors were adopted for further analyses. Factor loadings for 6 factors analysis are included in Table 9 and in Fig. 5. Figure 5 presents only the first

three most meaningful factors, which have factor loadings above |0.7|. Merely loadings for variables and factors which turned out to be essential are presented. Variable names and loadings which occurred only in one analysis are presented in Italic.

In both analyses H, Pe, Dcal, VPIR and PHI variables were not essential. What is interesting is that in the FA for data set from Sa Fm, distinctly lower number of variables is present. There is lack of: LLS, LLD, DTSX, VGAS, PHIT, VSH, VLIM, VDOL variables. In Ja Mb data set, only VWF and VKWA variables were not engaged in FA. In Ja Mb data set, F1 transmitted information on 8 variables carrying the knowledge about bulk density (RHOB, RHEQ), elastic properties (DT, DTP, DTSX), neutron (NPHI) and total (PHI) porosities and volume of kerogen (VKER). F2 transmitted information about shaliness. All variables obtained from gamma ray logs (standard, GR and spectral, GRKT, GRTO, POTA, THOR and URAN) and from interpretation (VCBW, VSH) are represented. In F2, VKER is also observed. F3 transmitted the information carrying by resistivity logs, LLS and LLD. Porosity represented by variables PHIE and PHIG is visible only just

Table 8 Comparison of selected parameters in factor analysis for Ja Mb and Sa Fm in sets comprising log data from W-1, W-2 and W-3 wells

Ja Mb					Sa Fm				
No.	EV	TV (%)	CEV	CV	No.	EV	TV (%)	CEV	CV
1	14.41	43.67	14.41	43.67	1	9.07	27.48	9.07	27.48
2	5.83	17.67	20.24	61.34	2	6.09	18.47	15.16	45.95
3	3.22	9.76	23.46	71.09	3	4.65	14.09	19.81	60.04
4	1.32	4.00	24.78	75.09	4	3.02	9.14	22.83	69.18
5	1.20	3.63	25.98	78.72	5	1.73	5.24	24.56	74.42
6	1.09	3.29	27.07	82.02	6	1.16	3.52	25.72	77.94
7	0.71	2.16	27.78	84.17	7	0.91	2.77	26.64	80.71
8	0.66	2.01	28.44	86.18	8	0.59	1.80	27.23	82.51
9	0.53	1.60	28.97	87.78	9	0.58	1.75	27.81	84.26
10	0.39	1.17	29.35	88.95	10	0.54	1.63	28.35	85.89

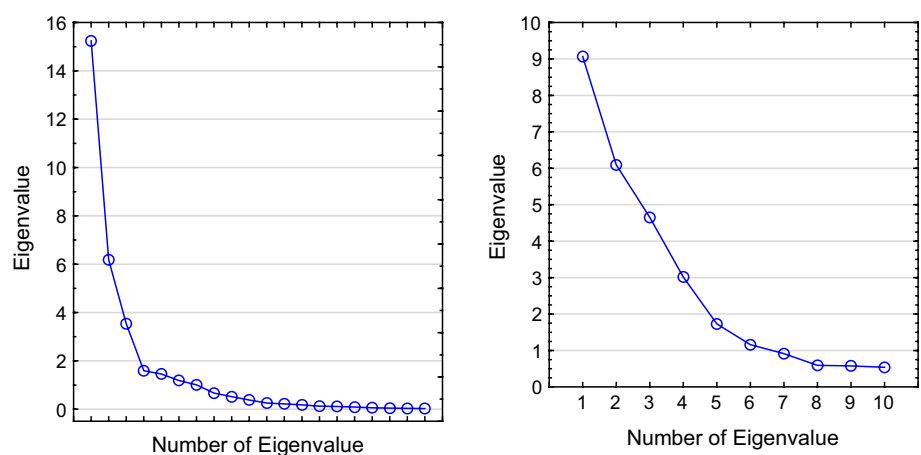**Fig. 4** Scree plots, factor analysis on data sets from Ja Mb (left) and Sa Fm (right)

Table 9 Factor loadings for FA in Ja Mb and Sa Fm data sets from W-1, W-2 and W-3 wells

| Var	Ja Mb						Sa Fm					
	F1	F2	F3	F4	F5	F6	F1	F2	F3	F4	F5	F6
LLS	0.47	0.16	0.82	−0.09	−0.09	−0.02	−0.39	−0.36	−0.41	−0.37	0.02	0.22
LLD	0.42	0.16	0.84	−0.06	−0.12	−0.01	−0.35	−0.40	−0.42	−0.37	0.01	0.21
DT	0.85	0.39	0.04	0.09	0.15	−0.01	−0.04	0.25	−0.12	0.74	0.06	0.24
RHOB	−0.80	−0.43	−0.17	0.02	−0.14	−0.09	0.53	−0.21	0.34	−0.24	−0.02	−0.60
NPHI	0.80	0.53	0.02	−0.02	0.04	−0.03	0.18	0.25	0.02	0.83	−0.16	0.09
GR	0.21	0.91	0.04	0.06	0.08	0.02	0.38	0.00	−0.88	0.01	0.01	0.01
GRKT	0.09	0.94	−0.04	0.00	−0.02	0.02	0.59	0.05	−0.51	0.13	−0.12	−0.06
GRTO	0.09	0.92	0.03	0.08	0.19	0.02	0.76	0.05	−0.33	0.06	0.10	−0.02
POTA	−0.05	0.92	0.00	0.12	0.13	0.01	0.90	−0.01	−0.04	0.21	0.02	−0.10
THOR	0.09	0.92	−0.06	−0.05	0.03	0.00	0.91	0.22	0.05	0.20	−0.01	−0.01
URAN	0.30	0.84	0.08	0.02	0.02	0.05	−0.04	−0.03	−0.95	−0.16	−0.01	0.05
DTP	0.85	0.35	0.06	0.09	0.12	0.04	−0.12	0.22	−0.10	0.72	0.16	0.23
DTSX	0.87	0.27	0.08	0.05	0.10	0.10	0.10	0.57	0.22	0.57	0.08	0.19
DKER	0.78	0.18	0.42	−0.09	−0.07	0.03	−0.53	−0.46	−0.49	−0.03	0.03	0.24
VCBW	0.04	0.81	−0.03	0.07	−0.30	−0.02	0.86	0.04	0.12	0.25	−0.07	−0.03
VWF	−0.22	−0.55	−0.39	0.17	0.18	0.13	0.04	0.79	0.31	0.23	0.05	−0.04
VGAS	0.42	0.09	0.05	−0.01	0.79	−0.11	−0.04	0.63	−0.47	0.07	0.06	0.11
PHIE	0.44	0.02	−0.11	0.03	0.86	0.06	−0.01	0.94	−0.14	0.18	0.03	0.06
PHIG	0.47	0.35	−0.12	0.10	0.74	0.06	0.32	0.87	−0.10	0.24	−0.03	0.04
PHIT	0.83	−0.37	−0.09	0.20	0.16	−0.15	−0.11	0.47	0.42	0.39	0.00	0.57
VSH	0.34	0.85	−0.04	0.05	−0.03	−0.07	0.52	−0.15	0.18	0.67	−0.02	−0.39
VLIM	0.15	−0.47	0.16	−0.77	−0.20	0.11	0.00	0.02	0.14	−0.19	−0.63	−0.06
VDOL	−0.36	−0.26	0.03	0.03	−0.03	−0.81	−0.23	−0.06	−0.19	−0.49	−0.38	0.19
VKWA	−0.38	−0.37	−0.14	0.59	0.15	0.55	−0.37	0.07	0.11	−0.21	0.82	0.24
VKER	0.26	0.84	0.09	0.02	0.17	0.05	−0.09	0.01	−0.94	−0.07	0.04	0.10
DPEQ	0.56	0.74	−0.01	0.03	0.13	0.11	0.29	0.06	0.12	0.90	0.16	0.04
DSEQ	0.64	0.62	0.09	−0.26	0.10	−0.01	0.29	0.00	0.10	0.86	−0.29	−0.05
RHEQ	−0.89	−0.20	−0.09	0.02	−0.16	−0.12	0.42	−0.21	0.23	−0.26	0.00	−0.77
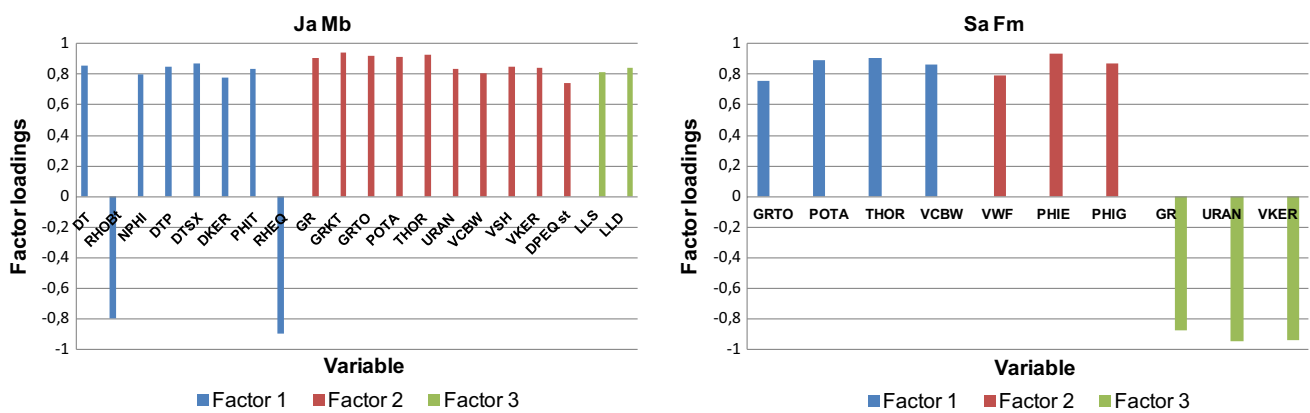
Fig. 5 Factor loadings on data sets from Ja Mb (left) and Sa Fm (right)

in F4. Such image from FA showed that in Ja Mb sweet spot, the most important is clay minerals composition and gas related in majority to mineral components, influencing

strongly elastic properties. Gas is also in porosity but in lower degree. Results of FA made in Sa Fm data set showed that F1 transmitted information on clay minerals

carried by variables: GRTO, POTA and THOR, indicating the presence of clay minerals and volume of clay bound water, VCBW. F2 transmitted information carried by porosity—effective, PHIE and total, PHIG and volume of free water, VFW. F3 transmitted information carried by variables related to organic matter (kerogen) presence. Information on elastic properties was transmitted by F4. It means that gas in Sa Fm sweet spot was in majority located in porous space.

In the next step, factor analyses were made on the data sets from Ja Mb and Sa Fm using 6 factors but including only variables selected in the earlier analyses, as shown in Table 9. Tests of the goodness of fit showed 87.37% (Ja Mb) and 88.93% (Sa Fm) of explained variance. Exemplary results were shown in the form of factor loadings (Table 10) for Sa Fm analysis.

Only 5 factors transmitted information carried by selected variables. Now, the factors with consecutive numbers transmitted information carried by other variables (in comparison with results presented in Table 9), but all important features were included in the analysis. F1 transmitted information on porosity, F2—on elastic properties, F3—on organic matter content. Information carried by gamma ray logs was transmitted only just by F4. The last factor, F5 in Table 10 plays no important role. The conclusion is that information on porosity in that analysis is the most important. In Ja Mb data set, the analysis made using only selected variables gave similar solution. Concluding this part of factor analyses, it is worth to point out that Ja Mb and Sa Fm are claystone/mudstone rocks of monotonous lithology. Differences in their petrophysical characteristics were based on volume and position of

Table 10 Factor loadings for Sa Fm data set, 6 factors analysis, selected variables, W-1, W-2, W-3 wells

Var	F1	F2	F3	F4	F5
GRTO	0.05	0.04	−0.32	0.84	−0.03
POTA	0.01	0.20	0.02	0.92	0.07
THOR	0.24	0.20	0.10	0.92	0.06
URAN	−0.04	−0.11	−0.97	−0.02	0.05
NPHI	0.25	0.83	0.03	0.17	0.18
DT	0.27	0.75	−0.14	−0.04	−0.08
VCBW	0.05	0.24	0.16	0.88	0.10
VWF	0.80	0.22	0.32	0.02	−0.06
PHIE	0.96	0.18	−0.14	−0.01	−0.04
PHIG	0.90	0.24	−0.08	0.32	0.05
VKER	0.00	−0.02	−0.99	−0.06	−0.02
DPEQ	0.07	0.92	0.17	0.30	−0.15
DSEQ	0.01	0.87	0.17	0.28	0.32
RHEQ	−0.24	−0.31	0.31	0.45	0.21
VKWA	0.06	−0.19	0.03	−0.34	−0.91

organic matter, considered as gas hydrocarbons in porous space and stuck on to rock matrix and kerogen being the part of rock matrix.

Laboratory data set in factor analysis

At first approach, the full laboratory-obtained parameters data set was the basis for factor analysis. There were considered 77 variables consisted of elements (weight content) such as: K, Na, Ca, Al, Fe, Mn, Mg, S, B, Cl, V and others such as Sm, Eu, Gd, chemical compounds such as Al_2O_3, Fe_2O_3, Na_2O, K_2O and others, minerals such as quartz, calcite, pyrite, group of illite and smectite and others, organic matter and group of parameters characterizing porous space obtained from the Mercury Injection Porosimetry, Helium Porosimetry, water and kerosene immersion porosimetry, geochemical special analyses, nitrogen adsorption/desorption method. In the consecutive steps, low loading variables were eliminated and final data set consisted of 38 variables. Factor loadings for the solution of 6 factors are presented in Table 11 and in Fig. 6. Figure 6 shows 6 factors, but presents only the variable, which has factor loadings above |0.7|. The most important factor F1 transmitted the information carried by 4 elements (aluminum, boron, chlorine, thorium), 3 chemical compounds (Al2O3, Na2O, K2O), 1 mineral, cation exchange capacity, special parameter characterizing clay minerals, CEC [mval/100 g] and 6 porosity parameters. Factor F2 transmitted the information carried by 2 elements (calcium, manganese), 3 chemical compounds (SiO_2, CaO, MnO), one mineral, total inorganic carbon, TIC and carbonate content, CC. F3 transmitted the information carried by two elements: vanadium, V and uranium, U, organic matter content, OM, total organic carbon, TOC. F4 transmitted information carried by Fe, S and Fe_2O_3, pyrite and total sulfur content, TS. Remaining factors transmitted the information dispersed from single elements (for instance F6 transmitted the information carried by samar, Sm, europium, Eu, gadolinium, Gd).

What is interesting is that F1 transmitted only one mineral, i.e., group of illite and smectite which carried the information on clay minerals content, which is in agreement with data set built in majority from laboratory outcomes of claystone/mudstone lithology and F2 transmitted information on calcite, related to calcareous/carbonate samples. It is also worth to mention the occurrence of parameters characterizing porosity only from the nitrogen adsorption/desorption experiments (Thommes et al. 2015). F1 factor transmitted the information on volumes of pores of various diameters, i.e., volume of micro-pores determined according to Dubinin—Raduszkiewicz methodology comprising pores with diameter less than 2 nm, V_DR, volume of mezzo-pores (diameter 2–50 nm) determined according to Barrett–Joyner–Halenda methodology,

Table 11 Factor loadings for three wells, for all formation data set from laboratory experiments

Var	F1	F2	F3	F4	F5	F6
Al.	0.81	0.23	0.16	0.07	0.27	0.06
Ca	−0.29	−0.85	−0.17	0.12	−0.09	0.05
Fe	−0.19	0.16	0.10	−0.85	0.24	−0.02
Mn	−0.23	−0.74	−0.16	0.05	−0.03	0.01
S	−0.16	0.07	0.15	−0.87	−0.19	−0.03
SiO_2	−0.12	0.89	−0.06	0.17	−0.10	−0.10
Al_2O_3	0.85	0.19	0.15	0.06	0.22	0.05
Fe_2O_3	−0.18	0.15	0.09	−0.85	0.23	−0.01
MnO	−0.23	−0.74	−0.16	0.05	−0.05	0.00
CaO	−0.28	−0.85	−0.18	0.11	−0.10	0.04
Na_2O	0.77	0.25	0.16	0.04	0.27	0.03
K_2O	0.82	0.15	0.21	0.04	0.28	0.00
TiO_2	0.22	0.05	0.40	−0.07	0.75	0.10
P_2O_5	0.03	−0.08	0.17	0.04	0.00	0.81
LOI	−0.18	−0.87	0.13	−0.10	−0.17	0.08
B	0.83	0.14	0.27	0.07	0.04	0.06
Cl	0.70	−0.01	−0.27	0.05	−0.15	0.02
V	−0.19	0.05	0.78	0.02	0.10	−0.03
Sm	0.35	−0.03	0.21	0.04	0.08	0.78
Eu	0.10	−0.10	0.25	0.03	0.17	0.83
Gd	0.25	−0.04	0.23	0.04	0.03	0.87
Th	0.84	0.12	0.10	0.03	0.16	0.16
U	0.06	0.14	0.79	0.06	−0.15	0.01
Calcite	−0.28	−0.83	−0.17	0.11	−0.15	0.03
Pyrite	−0.18	−0.04	0.09	−0.86	−0.21	−0.04
Illite + Smectite	0.88	0.21	0.09	0.05	0.09	0.02
OM	−0.21	0.07	0.87	0.00	−0.09	0.12
TS	−0.16	0.08	0.17	−0.87	−0.18	−0.03
TIC	−0.28	−0.85	−0.22	0.10	−0.06	0.02
TOC	−0.22	0.06	0.87	0.00	−0.08	0.11
CC	−0.28	−0.85	−0.22	0.10	−0.06	0.02
S_BET	0.77	0.07	−0.32	0.06	−0.01	−0.02
V_DR	0.75	0.06	−0.37	0.06	−0.02	−0.02
V_BJH	0.76	0.12	−0.32	0.07	−0.05	−0.10
VDFTmic	0.71	0.04	−0.26	0.04	−0.10	−0.02
VDFTmes	0.73	0.11	−0.38	0.05	−0.07	−0.11
V_total	0.73	0.12	−0.36	0.03	−0.01	−0.13
CEC	0.83	0.06	−0.18	0.01	−0.31	0.03

V_BJH, volume of micro-pores and mezzo-pores, respectively, determined according to Quenched Solid Density Functional theory model, VDFTmic and VDFTmes, total volume of pores with diameters less than 350 nm, V_total, S_BET as specific surface according to Brunauer—Emmet—Teller methodology was included, too.

Summarizing the results of factor analysis, similarity in factors construction was visible in well logging and laboratory data sets. Generally, two factors transmitted information on variables related to the most important components of lithology and structure of porous space. The third one transmitted specific characteristics of formations in study which were related to shaliness, volume of kerogene and volume of gas. Comparison of results obtained on well logging and laboratory data sets revealed similarity in factor construction transmitting the variables carried the knowledge on lithology, porous space structure and presence of organic matter. Analysis of relations between number of variables and number of samples in data sets (well logging and laboratory) showed that a variety of variables related to factors are critical in obtaining good results. Lower number of factors really made easier and improved rock formation recognition.
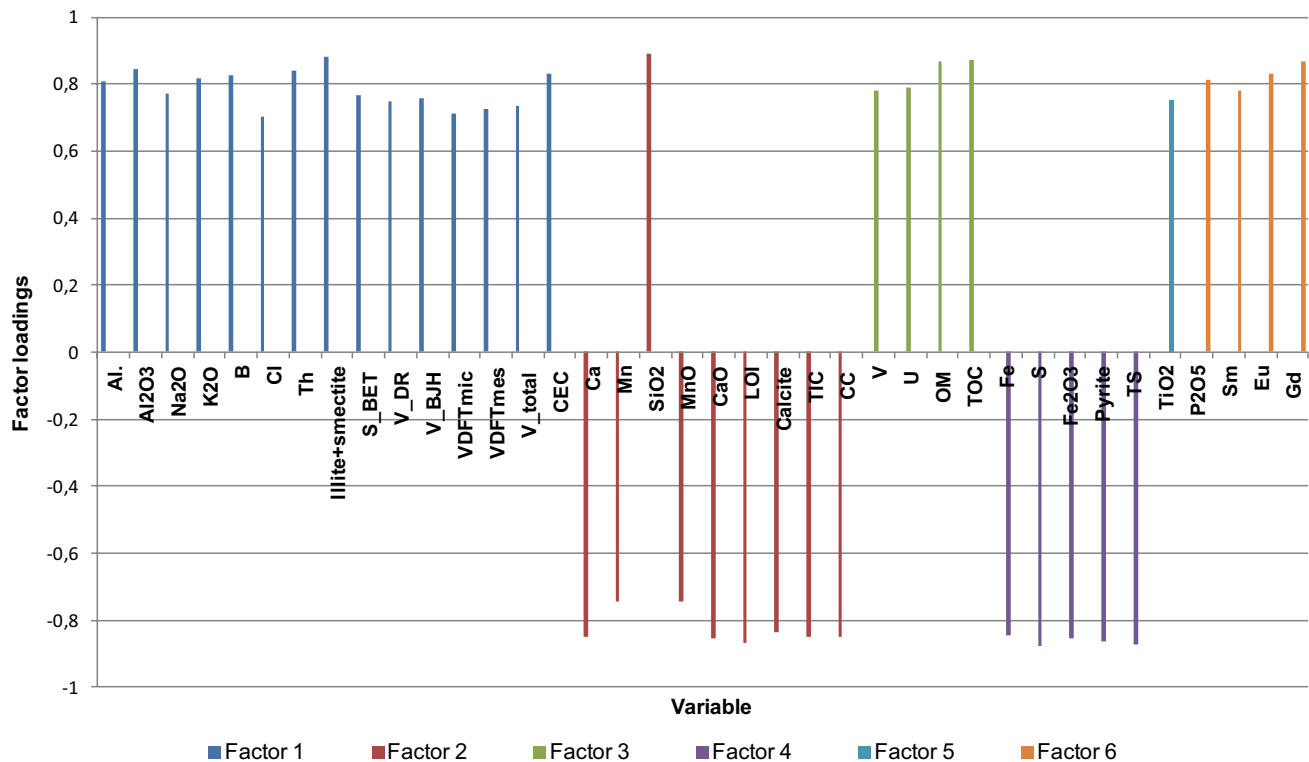
Fig. 6 Factor loadings (only above |0.7|) distribution for three wells, for all formation data set from laboratory experiments

Conclusions

Adopted algorithm comprising standardized data, Maximum Likelihood factoring and normalized Equamax rotation turned out to be useful and enabled petrophysical interpretation of results of factor analysis.

Factor analysis used for lithology discrimination was possible on the basis of standard logs. Even two or three factors transmitted sufficient information on claystone/mudstone and calcareous/carbonate formations.

Description of monotonous lithology of claystone/mudstone formations was a more challenging goal. Useful factors occur in greater number. More variables were visible as highly correlated with the first and second factors in tables with loadings.

To obtain information on volume and position of organic matter, it was necessary including into analyses variables carrying information about porosity and mineral composition due to the knowledge provided by other methods.

Intuitively, closely located wells drilling the same geological formations should provide the same results in FA. Such expectation was confirmed by obtained results, but in some cases, FA provided various results that pointed out heterogeneity of formation (frequent in shale formations).

Comparing obtained results interpreters concluded that the goal of lowering number of variables and adopting data sets consisting of number of samples proportional to laboratory data was successfully obtained in factor analyses using trial-and-error method. Petrophysical experience of interpreters supported intuitive selection of variables and factors.

Mercury Injection Porosimetry results turned out to be not informative in FA. All information about porosity (volume and size of pores) was taken from the nitrogen adsorption/desorption experiments.

Comparison of results obtained on well logging and laboratory data sets revealed similarity in factor construction transmitting the variables carried the knowledge on lithology, porous space structure and presence of organic matter.

Compliance with ethical standards

References

Asfahani J, Aissa M, Al-Hent R (2005) Statistical factor analysis of aerial spectrometric data, Al-Awabed area, Syria: a useful guide for phosphate and uranium exploration. Appl Radiat Isotopes 62:649–661. https://doi.org/10.1016/j.apradiso.2004.08.050

Bała M, Cichy A (2007) Comparison of P- and S-waves velocities estimated from Biot Gassmann and Kuster-Toksöz models with results obtained from acoustic wavetrains interpretation. Acta Geophys 55(2):222–230. https://doi.org/10.2478/s11600-007-0006-6

Costello AB, Osborne J (2005) Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. Practical Assessment Research & Evaluation, 10 (7). Available online: http://pareonline.net/getvn.asp?v=10&n=7, pp 9

Hair JF Jr, Black WC, Babin BJ, Anderson RE, Tatham RL (2006) Multivariate data analysis. Pearson Prentice Hall, New Jersey

Jarzyna J, Wawrzyniak-Guz K (eds) (2017) Adaptation to the Polish conditions of the methodologies of the sweet spots determination on the basis of correlation of well logging with drilled core samples: methodology to determine sweet spots based on geochemical, petrophysical and geomechanical properties in connection with correlation of laboratory test with well logs and generation model 3D. Monography, 500 pp., GOLDRUK Wojciech Golachowski Printing House (**in Polish**)

Jarzyna JA, Bała M, Krakowska PI, Puskarczyk E, Strzępowicz A, Wawrzyniak-Guz K, Więcław D, Ziętek J (2017) Shale gas in Poland. In: Natural Gas. INTECH. ISBN: 978-953-51-4657-5, https://doi.org/10.5772/67301, distributed by Internet

Kaiser HF (1958) The varimax criterion for analytical rotation in factor analysis. Psychometrika 23(3):187–200. https://doi.org/10.1007/BF02289233

Kaźmierczuk M, Jarzyna J (2006) Improvement of lithology and saturation determined from well logging using statistical method. Acta Geophys 54:378–398. https://doi.org/10.2478/s11600-006-0030-y

Puskarczyk E (2018) Applying of the Artificial Neural Networks (ANN) to identify and characterize in shale gas. E3S Web of Conferences; ISSN 2267-1242.—2018 vol 35 art. no. 03008, s. 1–7.—https://www.e3s-conferences.org/articles/e3sconf/pdf/2018/10/e3sconf_polviet2018_03008.pdf

Statistica 13.3 Help, 2019, electronic version

Szabó NP (2011) Shale volume estimation based on the factor analysis of well-logging data. Acta Geophys 59:935. https://doi.org/10.2478/s11600-011-0034-0

Thommes M, Kaneko K, Neimark AV, Olivier JP, Rodriguez–Reinoso F, Rouquerol J, Sing KSW (2015) Physisorption of gases. with special reference to the evaluation of surface area and pore size distribution (IUPAC Technical Report). Pure and Applied Chemistry, 87, 9–10, 1052–1069

Topór T, Derkowski A, Kuila U, Fischer TB, McCarty DK (2016) Dual liquid porosimetry: a porosity measurement technique for oil-and gas-bearing shales. Fuel 183:537–549. https://doi.org/10.1016/j.fuel.2016.06.102

Wawrzyniak-Guz K, Jarzyna JA, Zych M, Bała M, Krakowska PI, Puskarczyk E (2016) Analysis of the Heterogeneity of the Polish Shale Gas Formations by Factor Analysis on the Basis of Well Logs. In: Proceedings of the 78th EAGE conference & exhibition, 2016, Vienna, Austria, 30 May—2 June 2016, Tu SBT3 07