Scientific Annals, School of Geology, Aristotle University of Thessaloniki Proceedings of the XIX CBGA Congress, Thessaloniki, Greece	Special volume 99	141-147	Thessaloniki 2010

STATISTICS TO IMPROVE RESULTS OF WELL-LOGGING INTERPRETATION IN RESERVOIR ROCKS: TWO CASES FROM THE CARPATHIAN FOREDEEP

Jarzyna J.[1] and Kazmierczuk M.[2]

[1] *AGH University of Science and Technology, Faculty of Geology Geophysics and Environmental Protection, Department of Geophysics, al. Mickiewicza 30, 30-059 Cracow, Poland, jarzyna@agh.edu.pl*
[2] *Polish Oil and Gas Company in Warsaw, South Department in Sanok, Biuro in Cracow, ul. Lubicz 25, 31-503 Cracow, Poland, monika.kazmierczuk@pgnig.pl*

Abstract: Principal component analysis, cluster analysis, and discriminant analysis were applied to well-logging data from the Miocene clastic formation in the Carpathian Foredeep, Poland. The main goal was to improve the results of interpretation of well logging in terms of determining gas-saturated horizons. The presented examples illustrate how statistical methods help limit the number of log data while preserving sufficient information. In addition, the two cases illustrate the grouping of data into clusters to reveal sets of features attributed to reservoir horizons and sealing layers and construction of discrimination functions to distinguish between gas- or water-saturated beds of sandy-shaly lithology.

Keywords: well logs, thin beds, Miocene gas deposits, Principal Component Analysis, Cluster Analysis, Discriminant Analysis

1. Introduction

The case studies presented in this paper are the results of research performed in the Miocene gas deposits in the Carpathian Foredeep. K and D are small gas deposits formed in the Sarmatian succession of sandy-shaly thin-bedded clastic rocks in which sedimentological conditions cause variability of lithology and reservoir parameters, i.e., porosity and permeability and hydrocarbon saturation (Fig. 1). The two examples provided are intended to show how statistical methods ease interpretation of well-log data in difficult geological conditions.

2. Methods

2.1. Well-logging data

Wireline logging in a borehole provides log analysts with great amounts of data. Usually, a dozen or so logs are simultaneously made in the borehole, with each log based on the different petrophysical parameters. Different techniques can give different results, however; for example, resistivity is measured using several devices that provide information depending on their radius of investigation. Shallow devices penetrate the flushed zone, deeper ones penetrate the invaded zone, and the deepest ones reach the virgin zone. Thus, we cannot freely interchange results from resistivity logs but must select the one that gives consistent and accurate information. In addition, well-log data can be grouped based on the information delivered. For instance, spontaneous potential log (SP) and gamma ray log (GR) address the shaliness in a rock formation, and we can use both or select the one that is more useful in processing and interpretation. Statistical approaches may help in making this decision. Table 1 gives a set of logs that is typical for the study area. The number and type of logs depend on the geological and prospecting goals and

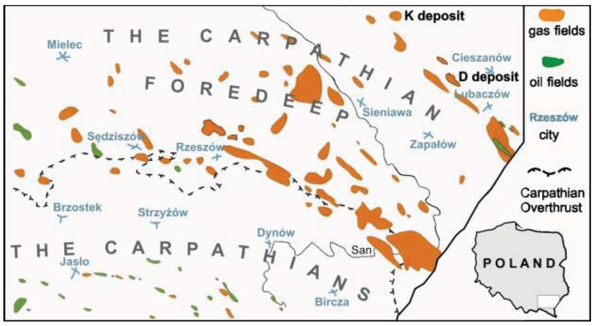

Fig. 1. Location of the K and D deposits on the background of a part of the Carpathian Foredeep (Karnkowski, 1999, modified).

methodology and equipment used. Results of the comprehensive interpretation of well logs, i.e., volumes of mineral components and porosity and water saturation, are also included in the statistical investigation.

Table 1. Well logs and results of the comprehensive interpretation used in the statistical processing.

Well logs	Results of the interpretation
GR*, NPHI, RHOB, SP, DCAL, LL3, EL14, EL28, EN16, EN64, HO01-HO12, DT	PHI, K, SW, VSA, VSH

GR*, gamma ray; NPHI, neutron porosity; RHOB, bulk density; SP, spontaneous potential; DCAL, difference between caliper and bit size; LL3, resistivity from the Russian type laterolog; EL14 and EL28, resistivity from lateral logs of shallow and deep radius of investigation, respectively; EN16 and EN64, resistivity from normal logs of shallow and deep radius of investigation, respectively; HO01-HO12, logs of increasing radii of investigation of High Resolution Array Induction, HRAI, respectively; DT, transit interval time of P wave; PHI, total porosity; K, permeability; SW and SX0, water saturation in the virgin and flushed zones, respectively; VSA, volume of sand; VSH, volume of shale

2.3. Applied statistical methods

The appropriate statistical approach to use for well logging should be a match for the methods that are appropriate to the problem being solved. Efficiency depends on the theoretical precision of methods, the possibility of adapting them to log data, and the accuracy of logging the given data (Ingerman 1995). Three methods were selected for the goals of this work: (1) limiting the available data set without losing information; (2) making easier the processing of data clustered into the selected groups; and (3) analyzing the discrete sets of data after discrimination and classification based on defined criteria (Jolliffe 2002). Principal component analysis (PCA) was used to address the first goal. This mathematical procedure transforms a number of possibly correlated variables into a smaller number of uncorrelated principal components. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. The applicability of PCA is limited by assumptions about data linearity, by the statistical importance of mean and covariance, and by an assumption that large variances have important dynamics. Cluster analysis (CA) is the assignment of a set of observations into subsets so that observations within a cluster are similar in some way. Clustering is a common technique for data mining that builds a hierarchy from individual elements by progressively merging clusters. The first step is to determine which elements to merge into a cluster. Usually, we take the two elements that are closest based on the chosen distance measure. The idea of clustering variables rather than observations is useful and helps make the clustering of PCA results effective. Linear discriminant analysis (LDA) is a method used to identify the linear combination of features that best separate two or more classes of objects. The resulting combination is used as a linear classifier. Canonical discriminant analysis identifies axes that best separate the categories. LDA explicitly attempts to model the difference between the classes of data. PCA, on the other hand, does not take into account any difference in class.

3. Results and Discussion

Two cases are presented here. The first example shows how statistics can help distinguish among thin sandstone/mudstone gas-saturated beds, thin sandstone/mudstone water-saturated beds, or claystones themselves, crucial distinctions for gas prospecting in the Miocene sandy-shaly formation in the Carpathian Foredeep. The effect of masking of gas-saturated low-thickness horizons because of insufficient vertical resolution of logs is observed as low resistivity of gas-bearing layers. In the study area, low resistivity is typical for water-bearing sandstones/mudstones and claystones. A short radius of investigation logs such as bulk density, RHOB, and acoustic transit interval time (DT) also reveal distortion: Because gas is shifted into the far parts of pores due to mud filtration in the flushed zone, these features do not show anomalies caused by gas presence in the porous space.

3.1. Case 1

Case 1 relates to the Miocene sediments in the K gas deposit in the eastern part of the Carpathian Foredeep. K11 and K17 wells are drilled in the Sarmatian formation built of thin sandstone laminas interbedded with thin layers of claystones and mudstones. Differences in sedimentological construction of the beds and in the lithology are apparent, but low thickness of the layers and low vertical resolution of logging probes limit using well logging to obtain proper, detailed petrophysical characteristics of selected beds. Many logs of various vertical resolution are run in both wells and give a great amount of data. Data delivered by single logs do not repeat the information but are instead complementary; however, it is difficult to analyze individual anomalies of many logs separately. Results of the comprehensive interpretation are available together with logs, and the interpreter is overloaded with data. In addition, the interpreter still faces the problem of choice: Which anomaly depicts a gas-bearing horizon and which one is artificial due to limitations of the measured technology and interpretation procedures?

Statistical analysis of data in wells K11 and K17 consists of four steps. First, basic statistics are calculated and means, medians, minima and maxima, and standard deviations show that claystone and mudstone and shaly sandstone dominate in the analyzed rocks in both wells. Histograms of the results of the comprehensive interpretation (Fig. 2) reveal more details. For example, histograms of sand volume (VSA) are similar, but the range of change in VSA is greater in the K17 well. Also, the porosity (PHI) histogram reveals a greater number of data with higher PHI; the histogram of water saturation (SW) is shifted to the right, with a maximum close to 100% but with many samples between 75% to 99% showing gas-bearing beds. The histogram of $\log_{10}K$ (LogK) shows higher permeability than in well K11, i.e., more samples reveal permeability higher than 0.1 mD. Next, PCA reduces the number of primary log data. According to the Kaiser criterion and scree test (Jolliffe, 2002), only three

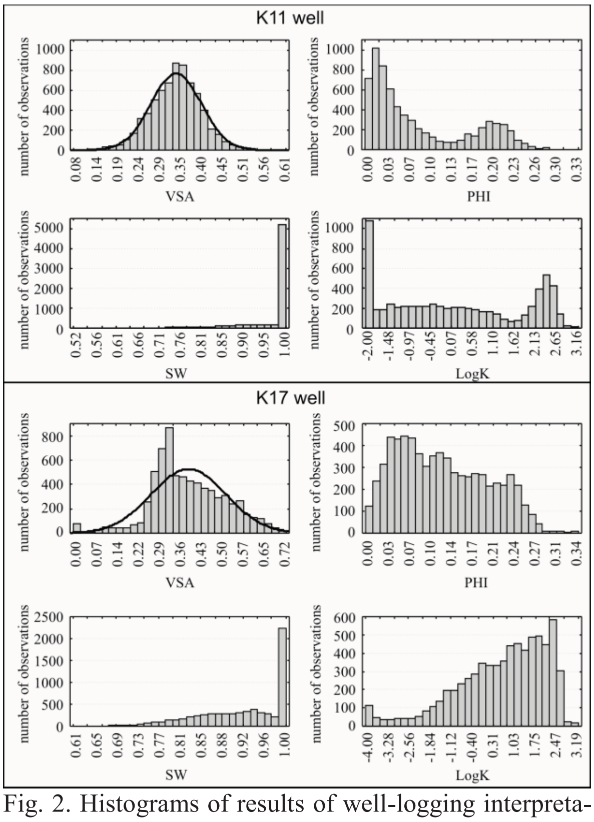

Fig. 2. Histograms of results of well-logging interpretation in the K11 well (upper part) and K17 (lower part).

Table 2. Loading factors for PC and logs in K11 and K17 wells

	K 11			K 17		
	PC1	PC2	PC3	PC1	PC2	PC3
GR	-0.32	0.84	0.23	-0.28	0.83	-0.40
NPHI	0.06	0.74	0.33	-0.15	0.89	0.08
RHOB	-0.52	0.56	-0.10	-0.32	0.32	-0.79
SP	0.15	0.05	0.86	0.56	0.68	0.23
DT	0.35	0.20	0.60	0.37	0.47	0.64
EL14	0.83	0.14	0.23	0.56	0.31	0.60
EL28	0.74	0.20	0.28	0.49	0.34	0.60
EN16	0.93	-0.14	0.13	0.90	-0.00	0.38
EN64	0.94	-0.10	0.10	0.87	0.08	0.41
LL3	0.94	-0.11	0.13	0.91	0.01	0.35
HO01				0.69	-0.15	0.24
HO02				0.91	0.03	0.22
HO03				0.92	0.08	0.24
HO06				0.94	0.10	0.25
HO09				0.93	0.11	0.24
HO12				0.90	0.11	0.23
DCAL	0.27	0.78	-0.09	0.29	0.43	0.20

principal components are selected. They explain 75% of the variability of the data in the K11 well and 82% in the K17 well. Thus, we do not risk losing information about rock formation. Loading factors for the principal components (Tab. 2) show which logs are important in constructing them and explain the variance in the data set. In both wells, resistivity logs construct the most important first principal component, PC1. Resistivity of rocks depends on mineral composition (lithology) and saturation of porous space. All resistivity logs included in PC1 have high loading factors, indicating that information about the invasion zone provided with resistivity from logs of different radial range is crucial for a description of a sandy-shaly gas-saturated formation. Information from GR and NPHI falls into the second position in the PCA. GR gives the information about content of clay minerals, and NPHI gives information about the hydrogen index (HI). A high index is typical for water-bearing formations, but shaliness in the formation can increase it, too. Gas-bearing formations have reduced HI because of lower amounts of hydrogen filling the porous space. The third principal component (PC3) in the study wells is not based on the same logs. In well K17, where gas-bearing sandstones occur more frequently, PC3, based on RHOB, gives information about saturation. In the

K11 well, a high position of SP in PC3 reflects the influence of the clay component of lithology (shaliness in the K11 well rock profile is higher than in the K17 well). In the next step, CA for principal components is performed. Because of the great number of data for the full Sarmatian profile (in K11, 6811 data depth points; in K17, 6941), 334 and 340 data points are drawn for K11 and K17, respectively. The Ward method is used for clustering, and as a result, dendrograms are used for analysis (Fig. 3). Optimal cut-off levels are selected on the basis of agglomeration plots. Five clusters are selected for K11 (A1, A2, B, C1, and C2) and three for K17 (A, B, C). One-way analysis of variance with nonparametric tests for median for results of the comprehensive interpretation was done to confirm the correctness of clustering and prove a real variability of lithology and reservoir parameters in the selected groups. Box and whisker plots for results in wells K11 and K17 show differences between populations (Fig. 4). Clusters A and C are similar in terms of lithology. In K11, all groups (A1, A2, B, C1, and C2) are similar for VSA. In K17, the volume of sand in cluster B is distinctly lower. A similar pattern emerges for PHI and LogK in the groups distinguished in the K11 and K17 wells. Thus, we conclude based on data from the A and C clusters a composition of shaly sandstone beds of good reservoir parameters. Cluster B comprises data from water-saturated horizons of increased shaliness and lower porosity and permeability. In terms of water saturation, cluster A consists of a high-porosity, high-permeability water- and gas-saturated population. Group A2 in K11 and group A in K17 are the most important from a prospecting point of view. Medians, 25% and 75% quartiles, and minimum–maximum ranges provide a good evaluation of the gas-bearing possibility of the formation in groups A2 (K11) and A (K17).

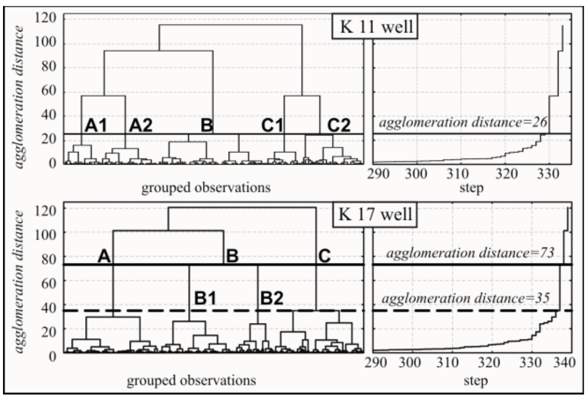

Fig. 3. Results of cluster analysis in K11 and K17 wells.

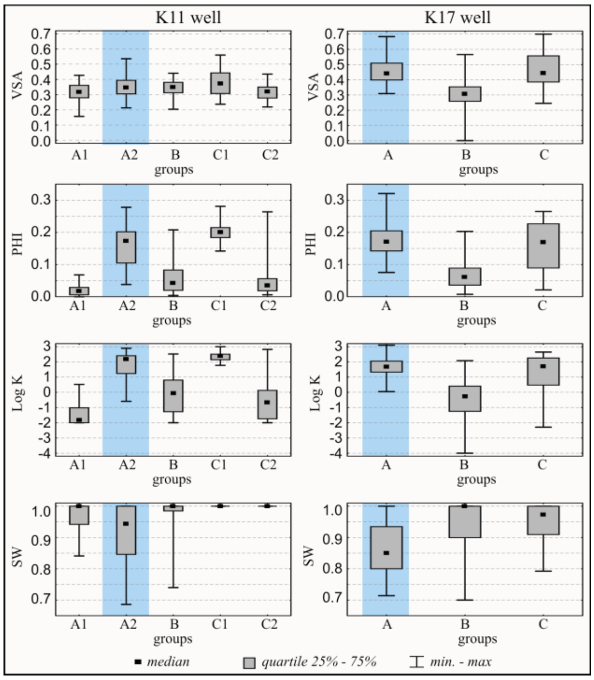

Fig. 4. Box and whisker plots for volume of sand, VSA, porosity, PHI, decimal log of permeability, $\log_{10}K$, and water saturation, SW, for wells K11 and K17.

The third step of data processing is the LDA of principal components. For this approach, the data sets used for clustering are divided into a training set and a testing set in a 3:1 proportion. Training sets are used for determination of the discriminant functions. Correctness of the LDA is checked using the testing sets. Three discriminant functions are defined for data in K11. The principal component PC1 contributes the most to discrimination, a result of the lowest value of partial Wilks lambda and the highest value of F statistics and tolerance (Tab. 3). The first discriminant function, DF1_11, provides 55% of discrimination power. In the DF1_11function, PC1 carries the greatest weight. In the second discriminant function, DF2_11, providing 25% of discrimination power, PC2 is the most important.

Groups of water-saturated shaly-sandy rocks (A1, B, and C2) of lowered porosity and permeability

Table 3. Results of discriminant analysis, standardized coefficients of discriminant functions, K 11 well

	Wilks lambda	Partial Wilks lambda	F	T	DF1_11	DF2_11	DF3_1
PC1	0.13	0.25	187.98	0.89	-0.94	0.23	-0.43
PC2	0.09	0.37	106.22	0.97	0.26	-0.82	-0.54
PC3	0.10	0.31	141.72	0.92	-0.71	-0.52	0.56
Cumulative %					0.55	0.80	1.00

have various values of the first discriminant function, DF1_11, and of the second function, DF2_11 (Fig. 5). Classification functions are calculated based on a classification probability assumed *a priori* in proportion to the number of data in the clusters. For the defined classification functions, correctness of classification is about 93% (Tab. 4). Correctness in the A2 group, the most important from a prospecting point of view, equals 94%. The effectiveness of the method is confirmed in the test data set (Tab. 4). For the full data set, 92% of the data are correctly classified, and the best result is obtained for groups A2 and B. This example shows how the lithology of the horizons with the best reservoir parameters can be distinguished among other, similar lithological types. In the K17 well, the attention is focused on discrimination between gas-bearing and water-saturated beds of favorable (sandy) lithology and good reservoir parameters. Two discriminant functions are considered, and the first one, DF1_17, comprises 70% of discriminant power. In the DF1_17 function, the most important are PC1 and PC3 (lowered RHOB is an indicator of gas-bearing beds). In the second discriminant function with 30% discriminant power, PC2 (based on shaliness indicators) has the greatest weight. The first discriminant function differentiates between sandy beds of good reservoir parameters from A and B clusters and from beds of cluster C of low porosity and permeability. The sandy-shaly, water-saturated group of good reservoir parameters (C) is distinctly separated in Figure 6. The highest values of the first discriminant function, DF1_17, and the vast range of the second discriminant function, DF2_17, are typical for that group. Low values for DF1_17 and high values for DF2_17 are characteristic for the group of good reservoir parameters (as in the previous case) but not for gas-bearing samples. We conclude that the second discriminant function is crucial for selecting gas-bearing beds because of positive values in that group of data.

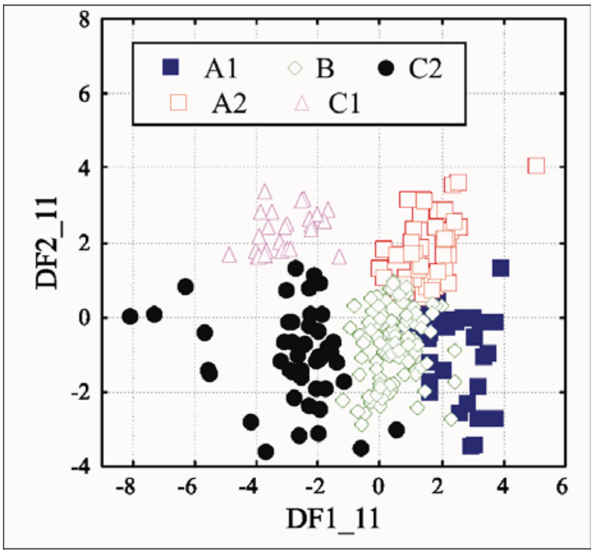

Fig. 5. Results of the discrimination analysis, dispersion plot of the canonical functions of the K11 well. A1, B, C2: water-saturated sandy shales of low porosity and permeability; A2: water- and gas-saturated shaly sandstones of high porosity and permeability; C1: water-saturated shaly sandstones of high porosity and permeability.

The results of discriminant analysis in the K11 and K17 wells are presented together with the lithology–porosity–water saturation solution from the comprehensive interpretation of well logging (Fig. 7). In the upper part of the well profiles, there are water-saturated beds classified according to reservoir properties and the lithology solution obtained from the comprehensive interpretation of well logs. In the lower part of the profiles, more gas-bearing layers are identified and a conformity of the statistical analysis and the comprehensive interpretation is observed.

3.2. Case 2

Case 2 concentrates on the separation of gas-saturated parts of the special sandstone horizon, designated as D sandstones, from other clastic sediments in the Sarmatian sandy-shaly formation in the D gas deposit. The elaborated D sandstones

Table 4 Correctness of classification, K11 well.

Group	%	A1	A2	B	C1	C2	%	A1	A2	B	C1	C2
	Training data set						Testing data set					
A1	82.35	28	3	3	0	0	A1	88.89	8	0	1	0
A2	94.00	0	47	3	0	0	A2	100.00	0	12	0	0
B	100.00	0	0	100	0	0	B	100.00	0	0	19	0
C1	100.00	0	0	0	26	0	C1	93.75	0	0	1	15
C2	82.98	0	0	3	5	39	C2	80.95	0	0	2	2
All	93.39	28	50	109	31	39	All	92.21	8	12	23	17

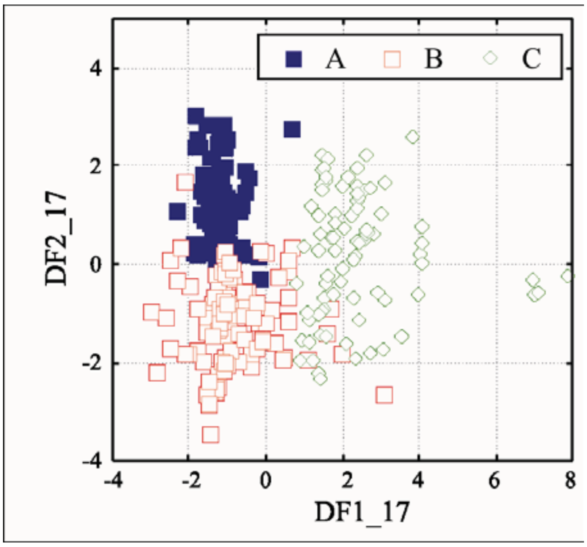

Fig. 6. Results of the discrimination analysis, dispersion plot of the canonical functions, K17 well. A: water- and gas-saturated sandy-shaly rocks of high porosity and permeability; B: water-saturated shaly sandstones of low porosity and permeability; C: water-saturated sandy-shaly rocks of high porosity and permeability.

occurring at the bottom of the Sarmatian succession are porous (15–35%) and permeable (up to a few hundred mD) quartzite arenites of 20–50-m thickness of submarine fans. Their genesis is the result of clastic material transportation from eroded orogene, temporarily uplifted over sea level in the great Miocene sedimentary basin in front of the Carpathians head (Porebski et al. 2000). Typical block-shape anomalies in well logs are observed in the D sandstone (Fig. 8). Statistical analysis is performed using log data from three wells. In the D12 and D13 wells, D sandstones are identified, but in D15 they are not observed. At the first stage, PCA gives the reduction of log data to four principal components that explain more than 80% of the variability of the data. In the D12 and D15 wells, the most important role for PC1 is in resistivity logs (EN16, EN64, and LL3). Resistivity information is crucial in lithology and gas-saturation identification. In the D13 well, PC1 is built by GR, NPHI, SP, and DCAL (the difference between real caliper and bit size). This result indicates that in D13, shaliness plays the greater role in lithology in comparison to other wells, although simple statistics using VSH do not confirm this conclusion. In D12 and D15, PC2s are similar, constructed by GR and SP and confirming the importance of shaliness in the interpretation of the Sarmatian sediments. Contrary to results from the previous wells, in D13 resistivity logs, EN16 and EN64 build PC2. PC3s in all wells are influenced by DT and RHOB, emphasizing the role of gas saturation in well logs and principal components. PC4s are built by lateral resistivity logs, EL14 and EL28. In summary, PCA is an effective tool in the reduction of the number of data. The important information about lithology (especially VSH), porosity, and gas saturation is preserved. PCs do not show special features of D sandstones. In the next step, clustering is done on the basis of PCs, and a distinct A group of D sandstone is separated into D12 and D13. Keeping in mind the ability of PCA to show the internal structure of data, additional grouping of PCs is performed on the selected samples from only D sandstones to check the uniformity of that group

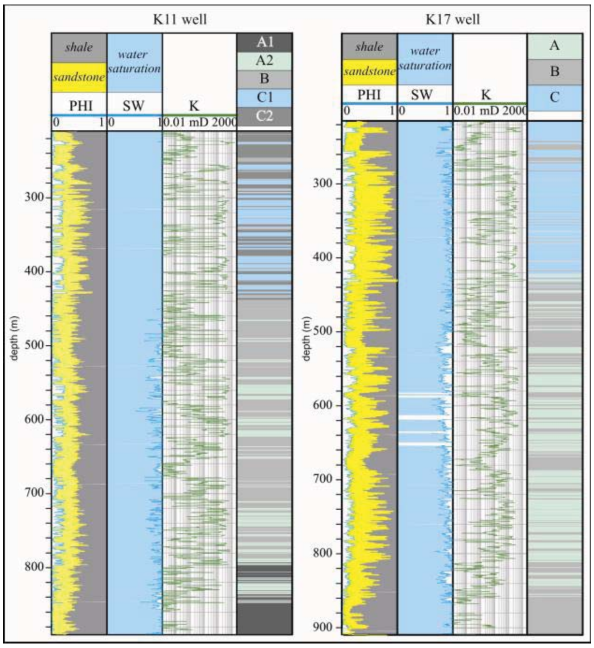

Fig. 7. Lithology–porosity–saturation solution and results of discriminant analysis in wells K11 and K17; explanations as in Figures 5 and 6.

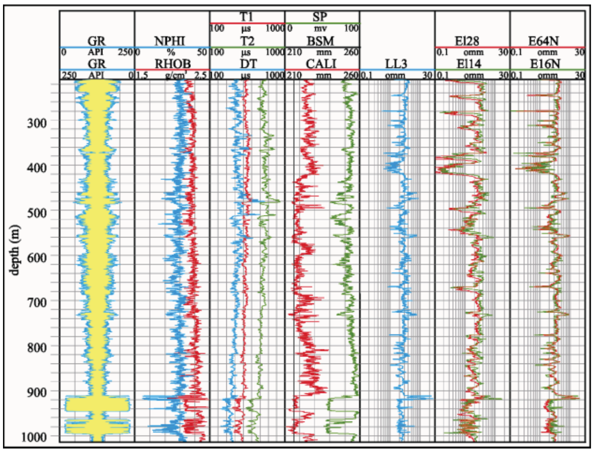

Fig. 8 Well logging in the D12 well in the Sarmatian succession; 915–995 m, D sandstones.

(Fig. 9). Selection of 4 and 5 groups in D sandstone data sets in wells D12 and D13, respectively, is essential to distinguish gas-saturated samples from others. Samples in the A cluster have a relatively high VSA, high values of PHI and K, and low water saturation (Fig. 9). Thus, we can say that D sandstones are not uniform in terms of lithology and water saturation but that the gas-saturated groups—A3 in D12 and A5 in D15—are distinct.

LDA and CA performed on principal components in the whole Sarmatian data set show good results. Training data sets are used to select components that are the best in differentiating the rock formation into classes of the best reservoir parameters and high gas saturation. Principal components built of GR, NPHI, and SP turn out to be the most effective in discrimination in all wells. Results of test data sets confirm classification obtained using training sets. After statistical analysis, we conclude that D sandstones are important in terms of gas saturation, but in gas prospecting, thin sandy-shaly beds cannot be overlooked.

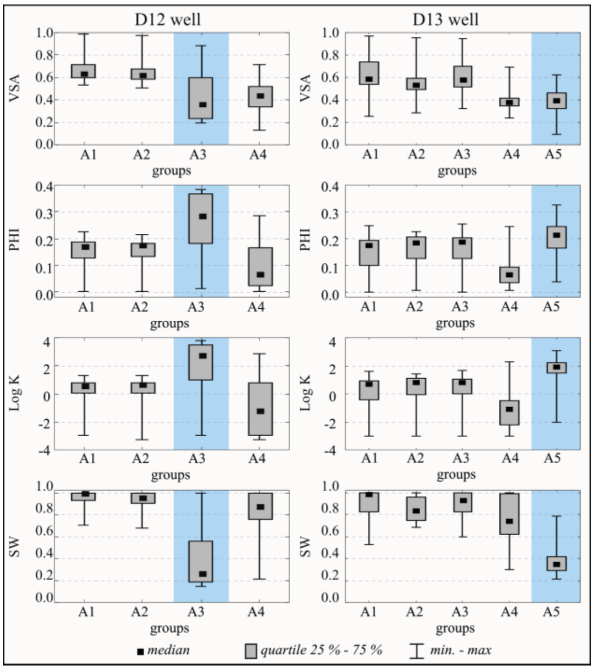

Fig. 9. Box and whisker plots for D sandstones in wells D12 and D13; explanations as in Figure 4.

4. Conclusions

Statistical methods make easier the processing of great amounts of ambiguous well-logging data. Use of principal components instead of full log data sets delivers enough information to classify rock samples in terms of lithology, reservoir parameters, and gas saturation. Thin gas-saturated sandstone layers in the Sarmatian shaly-sandy formation, especially in D sandstones, can be successfully identified.

Acknowledgments

Results were obtained through the project nr N525 023 32/2308 funded by the Polish Ministry of Science and High Education, 2007–2009, and the statutory project of the Department of Geophysics of the Faculty of Geology, Geophysics and Environmental Protection, AGH UST.

References

Ingerman V.G., 1995. Statistical Integration of Log Data for Lithology Determination and Formation Evaluation. The Log Analyst, No 3, p. 38-48

Jolliffe I.T., 2002. Principal Component Analysis. Springer-Verlag, New York Berlin Heidelberg, 487p.

Karnkowski P., 1999. Oil and gas deposits in Poland. Geosynoptics Society Geos, Cracow, 380p.

Porebski S.J., Doktor M., Gumulak K., Laptas A., 2000. Determination of sedimentational systems in region Biszcza-Ksiezpol and their facies and paleoenvironmental interpretation. Archive of POGC, Warsaw, Poland (in Polish).