



Artificial neural networks as a tool for pattern recognition and electrofacies analysis in Polish palaeozoic shale gas formations

Edyta Puskarczyk¹

Received: 11 May 2019 / Accepted: 6 September 2019 / Published online: 20 September 2019
© The Author(s) 2019

Abstract

Unconventional oil and gas reservoirs from the lower Palaeozoic basin at the western slope of the East European Craton were taken into account in this study. The aim was to supply and improve standard well logs interpretation based on machine learning methods, especially ANNs. ANNs were used on standard well logging data, e.g. P-wave velocity, density, resistivity, neutron porosity, radioactivity and photoelectric factor. During the calculations, information about lithology or stratigraphy was not taken into account. We apply different methods of classification: cluster analysis, support vector machine and artificial neural network—Kohonen algorithm. We compare the results and analyse obtained electrofacies. Machine learning method—support vector machine SVM was used for classification. For the same data set, SVM algorithm application results were compared to the results of the Kohonen algorithm. The results were very similar. We obtained very good agreement of results. Kohonen algorithm (ANN) was used for pattern recognition and identification of electrofacies. Kohonen algorithm was also used for geological interpretation of well logs data. As a result of Kohonen algorithm application, groups corresponding to the gas-bearing intervals were found. Analysis showed diversification between gas-bearing formations and surrounding beds. It is also shown that internal diversification in gas-saturated beds is present. It is concluded that ANN appeared to be a useful and quick tool for preliminary classification of members and gas-saturated identification.

Keywords Artificial neural network · Clustering · Electrofacies · Pattern recognition · Palaeozoic shale gas formation

Introduction

In recent years, machine learning methods have been used more and more successfully in petrophysical issues. The use of artificial neural networks for both classification and prediction has become a tool supporting comprehensive interpretation of well logging data (Hair et al. 2006; Szabó 2011; Szabó et al. 2013; Puskarczyk et al. 2015; Puskarczyk 2018).

The term *electrofacies* was introduced by Serra and Abbott in 1980. They defined electrofacies as *the set of log responses which characterizes a bed and permits this to be distinguished from others*. Based on standard well logging

data, like natural gamma ray, bulk density, neutron porosity, resistivity or P-wave velocity log, the electrofacies can be defined and often they correspond to one or more lithofacies. Traditionally lithofacies have been identified manually, based on core description and their correlation to well logs.

The most important step for electrofacies determination is core and log data integration. Electrofacies are based on log responses in the scale according to sampling rate of well logging while facies description based on cores are often in millimetres scale. We have to realize that electrofacies analysis given as general information about rock properties changes and can be used for pattern recognition in geological profiles of wells.

For determining electrofacies, authors like Doveton (1994) and Moss (1997) suggest to use clustering methods. In this paper, we use a few mathematical methods for automatization of the task of facies and electrofacies identification. These methods include support vector machine (SVM), cluster analysis (CA) and algorithms based on artificial neural networks (ANN). We were testing the results of using the mentioned mathematical methods for Polish

Paper was presented at the CAGG 2019 Conference “Challenges in Applied Geology and Geophysics” organized at the AGH University of Science and Technology, Krakow, Poland, 10–13 September 2019.

✉ Edyta Puskarczyk
puskar@agh.edu.pl

¹ AGH University of Science and Technology, al. A. Mickiewicza 30, 30-059 Kraków, Poland

Palaeozoic shale gas formations. The scheme of calculation and analysis is shown in Fig. 1.

Data set

In this paper, we took into account shale formations from lower Palaeozoic basin located at the western slope of the East European Craton. Upper Ordovician and lower Silurian shale deposits are classified both, as a source rock and a reservoir. The formations are characterised by facies changes from depositional environment with biological productivity to the clastic material distribution in the basin causing a dichroism of mud and clay deposits (Poprawa 2010). Because of vertical facies variability, their hydrocarbon potential is varied (Krakowska et al. 2016; Sowizdżał et al. 2016; Wawrzyniak-Guz et al. 2016; Puskarczyk 2017; Jarzyna et al. 2018).

In this study, we analysed data derived from Well-1, located in the northern part of the Baltic Basin. In this well, we choose for analysis depth interval, including deposits belonging to the Silurian Llandovery Formation and Ordovician—Ashgill, Caradoc, Llandeil, Llanvirn, and Arenig Formations. The depth interval cover two potentially unconventional resources formations—Jantar Member (Ja Mb) in Paslek Formation (Pa Fm) and Sasino Formation (Sa Fm). In that depth interval also, two carbonates (partially) formations, the Prabuty Formation (Pr Fm) and Kopalino Formation (Ko Fm) are placed.

The oldest part of Paslek Formation, Jantar Member is built by black bituminous claystones, containing pyrite, with dark-grey calcareous and few marly limestone laminas. It is characterised by high organic matter content. The thickness of the member does not exceed 12 m (Modliński et al. 2006).

The Sasino Formation is built by black, dark-grey and grey-greenish bituminous shales. In some intervals, marls

and limestones intercalations and also bentonite intercalations are visible in place. The thickness of the formation varied from 3.5 on the east to 37 m on the north-west in the land part of the Baltic Basin (Modliński and Szymański 2008).

As an input data in electrofacies analysis, standard well logging data were taken. Those well logs have been selected which are highly sensitive to lithology, porosity, and water or hydrocarbon saturation (Rider 2002; Serra and Serra 2004):

- GR (API)—natural gamma-ray log, total radioactivity of the formation. Gamma-ray log is a first shale indicator. Among the sedimentary rocks, shales have the strongest intensity gamma radiation due to higher concentration of K, U, Th elements. Potassium and thorium tend to be concentrated in clays, whereas uranium often shows high content in source rocks because of adsorption by organic matter.
- LLD (ohmm)—deep resistivity log, measured by laterolog tool. Resistivity is the primary physical property that allows determining reservoir properties, in particular porosity and water and/or hydrocarbon saturation. When organic matter is present in shales, whether in the form of insoluble kerogen or soluble bitumen, it characteristically increases resistivity.
- DTP ($\mu\text{s}/\text{ft}$)—compressional wave slowness log. Principal use of sonic log is porosity evaluation. In shales, lower sonic velocities are observed because of the presence of organic matter.
- NPHI (%)—neutron porosity. NPHI values are expressed in standardized neutron porosity units (p.u.), which are related to hydrogen index, HI—an indication of formation's richness in hydrogen. In geological conditions, hydrogen nuclei are supplied by water and hydrocarbons that are mainly distributed in the pore space, thus NPHI is a good porosity indicator. In shales, NPHI curve has abnormally high values of neutron porosity.

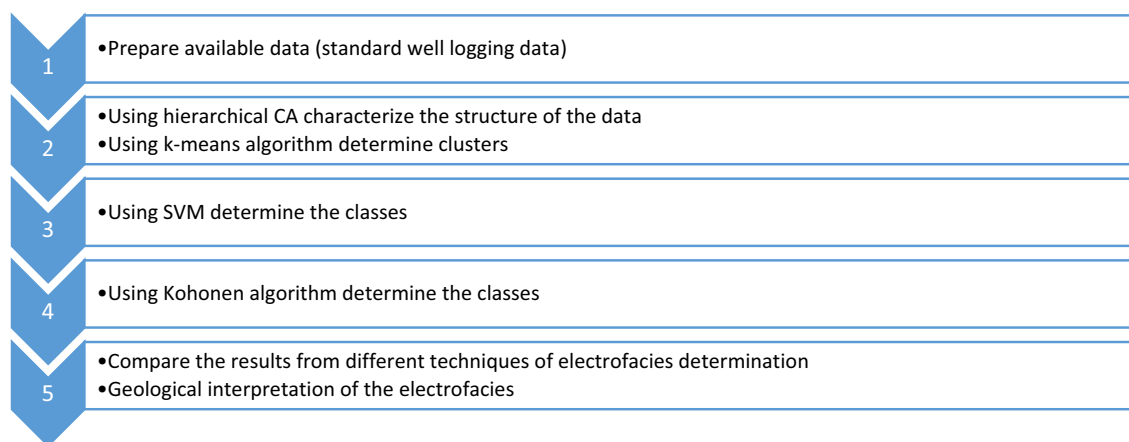


Fig. 1 Flowchart of analyses for electrofacies characterization

This is because clay water consists of free water, clay bound water as well as lattice water, which is part of clay mineral structure. Additionally, organic matter that has characteristic high hydrogen index causes a noticeable increase in NPHI readings in shales and clays.

- RHOB (g/cm^3)—bulk density log. It is overall rock's density that is a function of matrix density (mineral components forming the rock) and total porosity. Thick series of shales demonstrates progressive increase in density due to the compaction, when formation is overpressured there is a break in normal shale compaction trend manifested by a drop of bulk density, organic shales are distinguished by clear low anomalies on RHOB log because organic matter has much lower density in comparison with non-source shales.
- PE (b/e)—photoelectric factor log. PE measurements is an effective tool for matrix identification, because is a function of chemical/mineral composition and is independent from porosity.

Overlay the DT with the LLD is a good qualitative and quantitative source rock indicator and allows to TOC calculation. This technique is known as $\Delta\log R$ method (Passey et al. 1990). Immature source rock has decreased velocity and no changes in resistivity (low resistivity in shales). A mature source rock is marked by low values of velocity and high values of resistivity (increase in resistivity in comparison with immature source rock).

Methodology

Cluster analysis

Clustering methods have been developed in order to combine objects similarity in terms of studied features into homogeneous groups, called clusters. The members of clusters are at once alike and at the same time unlike members of other groups. In the clustering methods, we distinguish:

- the hierarchical methods that show the entire structure of the data set and allow the user to make a decision regarding the number of clusters, e.g. dendrogram method;
- the non-hierarchical methods, where the number of clusters is determined in advance, and then the objects are assigned to groups based on similarity, e.g. *k-means* method.

In clustering methods, the measure of the similarity of objects is the distance between objects.

In the hierarchical methods, in the first step of analyses, a distance matrix is created (e.g. calculated according to the Euclidean or Manhattan metric). Then the most similar

(the nearest) objects are searched and the first cluster is created. In the next step, the distance matrix is reduced and the distances of the objects to the newly created cluster are re-calculated. This process is repeated until all objects are grouped in one cluster. The graphical representation of grouping effects is a dendrogram (Fig. 2).

In the *k-means* method, the number of clusters is assumed in advance. Then, cluster seeds are arbitrarily or randomly chosen, to which the distances of objects are re-calculated. For each cluster, gravity mean of clusters are calculated and in the next step the distances of objects to these centroids are re-calculated. Then, the objects are transferred to the closest cluster. The process is repeated until all objects are assigned to clusters to which centres of gravity have the closest distance.

In practice, hierarchical methods are often used first to determine the number of clusters and the initial centroids, and then, the *k-mean* method is used, to divide large sets into homogeneous groups.

Support vector machine

The objective of the support vector machine algorithm is to find a hyperplane in an N -dimensional space that distinctly classifies the data points. Support vectors are data points that are closer to the hyperplane and influence the position and orientation of the hyperplane. Using these support vectors, we maximize the margin of the classifier. Deleting the support vectors will change the position of the hyperplane. In the SVM algorithm, we are looking to maximize the margin between the data points and the hyperplane. SVM has shown good performance in rock classification tasks (Wong et al. 2005; Sebtosheikh et al. 2015).

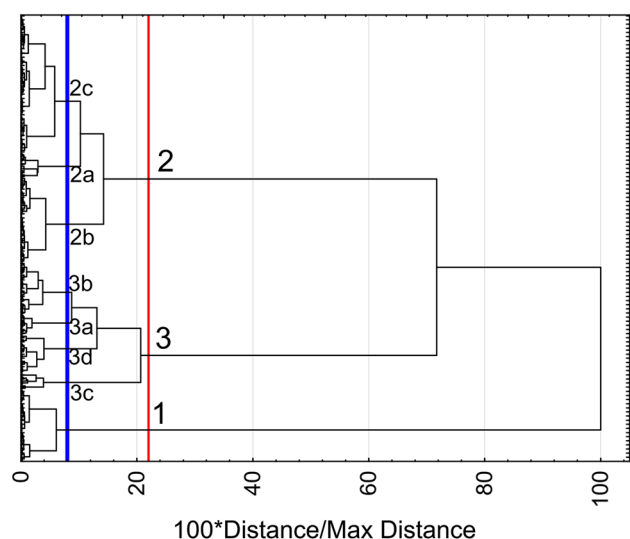


Fig. 2 Hierarchical clustering dendrogram

In this study, we performed two different kernel functions of SVM in the electrofacies prediction:

- linear: $k(x_i, x) = x_i \cdot x$ (1)

- radial basis (RBF) : $k(x_i, x) = e^{(-\gamma|x_i-x|^2)}$ (2)

where x input variables vector, x_i training data points vector and γ gamma parameter.

Artificial neural network

Artificial neural networks proposed by Kohonen (1982) are the most popular model of self-organizing networks.

The Kohonen networks work with the following algorithm:

- at the beginning, we have to select several parameters,
- the n -dimensional weight vectors v_1, v_2, \dots, v_i of the j computing units (randomly selected),
- an initial radius r , a learning constant a and a neighbourhood function α ,
- an input vector ε (selected by using the probability distribution over the input space),
- the unit k with the maximum excitation and for which the distance between v_i and ε is minimal,
- the weight vectors are updated by using the neighbourhood function

$$v_i \leftarrow v_i + a\alpha(i, k)(\varepsilon - v_i) \quad \text{for } i = 1, \dots, j \quad (3)$$

- a and/or α are modified then the recalculation are continued from the beginning (it can be stopped after Z iterations).

By repeating the process several times, arriving at a uniform distribution of weight vectors in input space is expected. During the learning process, both the size of the neighbourhood and the value of neighbourhood function fall gradually so that the influence of each unit upon its neighbours is reduced. From an initial distribution of random weights, and over Z iterations, the networks settles into a map of stable zones. Each zone is effectively a feature classifier.

Results

To create groups characterized by similar physical (and lithofacies) properties, methods of clusters analysis and artificial neural networks were tested.

The following analyses were performed:

- Cluster analysis—hierarchical and *k-means* method,

- Support vector machine,
- Kohonen neural network.

In the first step, we performed standard CA. Based on six input logs (given in “Data set” section), data grouping was performed. Figure 2 depicts the result of hierarchical grouping. Based on the dendrogram analysis, three main clusters were found (red line cut-off on the dendrogram). In the second more detailed division, eight subgroups were found (blue line cut-off on the dendrogram).

It can be seen that from the beginning of division group 1 separates from groups 2 and 3. Group 1 is the most homogeneous group. Group 2 was divided into three subgroups, where subgroups 2a and 2c show greater similarity, in comparison with the group 2b. Among the subgroups of the third group, 3c is the most distant from the others, subgroups 3a and 3b show the greatest similarity.

In the next step, cluster analysis using the *k-means* algorithm was done. Based on the results of the dendrogram analysis (Fig. 2), for that classification, the same number of clusters were chosen (3 and 8). Figures 3 and 4 show the average standardized values of the input logs, in the case of the clusters selected using the *k-means* algorithm. Each of the three clusters is characterized by different input values. It can be concluded that in the most general division, the shales are differentiated from the carbonates and the hydrocarbon saturated were differentiated from the water saturated intervals (Fig. 3). With a more detailed division, a much greater diversification of average values in individual clusters was found (Fig. 4).

In general, division performed based on hierarchical methods and *k-means* algorithm gave very similar results, even in the detailed division. Small differences can be seen in the intervals of the gas-saturated shales.

Analysing the average values of input logs (for better visualization in Figs. 3, 4, we use standardized average values of well logs), we can conclude that:

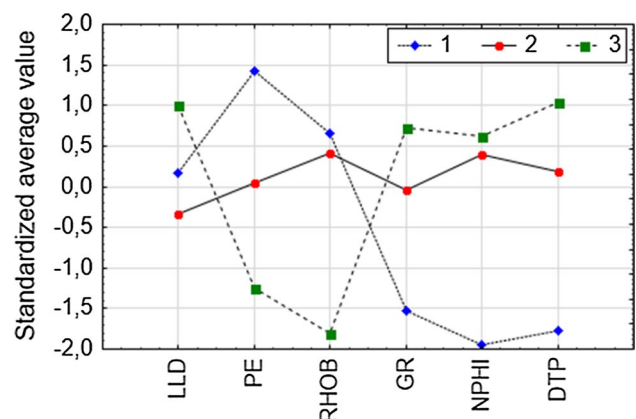


Fig. 3 Average values of standardized well logging data in a division into three clusters. Colours differentiate the number of clusters

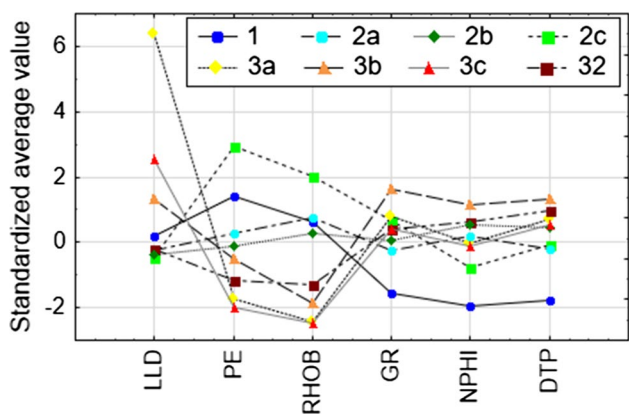


Fig. 4 Average values of standardized well logging data in a division into eight clusters. Colours differentiate the number of clusters

- group 1 (blue colour in Fig. 3) is characterized by the highest photoelectric index values, high density and low values of natural radioactivity and neutron porosity also showing high P-wave velocity. If we analyse the average values on the eight-class division (Fig. 4), we will see that group 1 still shows similar dependencies (it is the most homogeneous group) as in the case of a three-class division. It can be noticed, however, that the higher value of PE is shown by group 2c (higher average value than in group 1).
- group 2 (red colour in Fig. 3) is characterized by medium values (in relation to other groups) of all input logs. Analysing the results of the division of the second class into three subgroups (Fig. 4), it can be visible that the subgroups 2a and 2b show similar average values of the input logs (similar values of well logs in that groups). Among subgroups of the second group, subgroup 2c shows very high PE values (the highest among all subgroups) and high density; however, due to small number of samples (only 5 samples located at the started depth of the Kopalino Fm), it was decided not to analyse this case more detailed.

- group 3 (green colour in Fig. 3) is characterized by the lowest density and photoelectric index, low P-wave velocity and high values of resistivity, natural radioactivity and neutron porosity. Group 3 was divided into three subgroups (Fig. 4). The less numerous group 3a shows the highest resistivity values, low PE and density and high values of GR, NPHI, and DT. A similar value is obtained for most numerous group 3b, which, however, has a lower resistivity than 3a.

It was decided to separate the subgroup 32 (Fig. 4), which partially overlaps with groups 2 and 3 (form three cluster division) and is characterized by intermediate values of the logs.

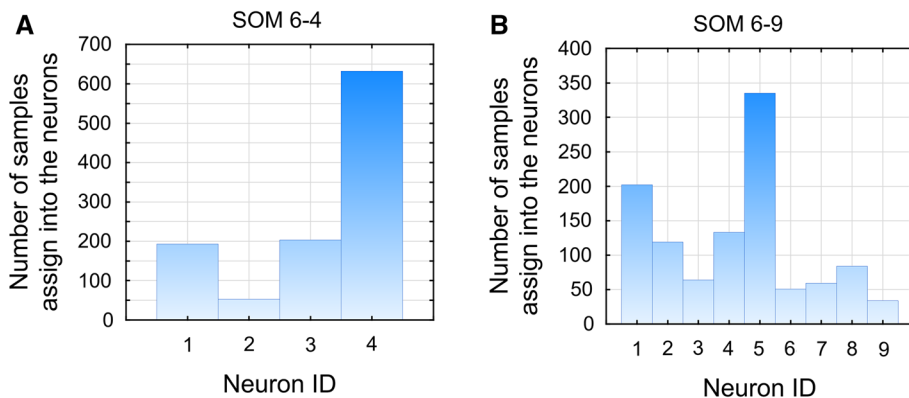
Kohonen algorithm was applied to the unsupervised classification. We tested several networks, with different structure and number of neurons. As the best network, the four (SOM 6-4) and nine (SOM 6-9)—neurons network were chosen. In Fig. 5, where the histograms were shown with the numbers of samples (depth values) assign to the classes. The most numerous class in the general division is group 4 (Figs. 5a, 6a). In a more detailed division, group 5 is the most numerous group (Figs. 5b, 6b).

Results obtained from the Kohonen network have shown greater heterogeneity of the analysed interval than CA. However, many similarities can be noticed, a number of facies in both methods have been assigned to the same (or very similar) depth intervals. We can summarize that both methods gave good results of classification. More detailed results were obtained from ANN that’s why in an interpretation part we focused on it.

The use of the network SOM 6-4 which divided the interval into four electrofacies (Fig. 8, track 16) allowed for the separation of the main levels:

- electrofacies 1—assigned to the depth intervals corresponding to the Ja Mb top interval and throughout the Sa Fm interval (in several sections),

Fig. 5 Histogram of the SOM 6-4 (a) and SOM 6-9 (b) samples assign to the neurons



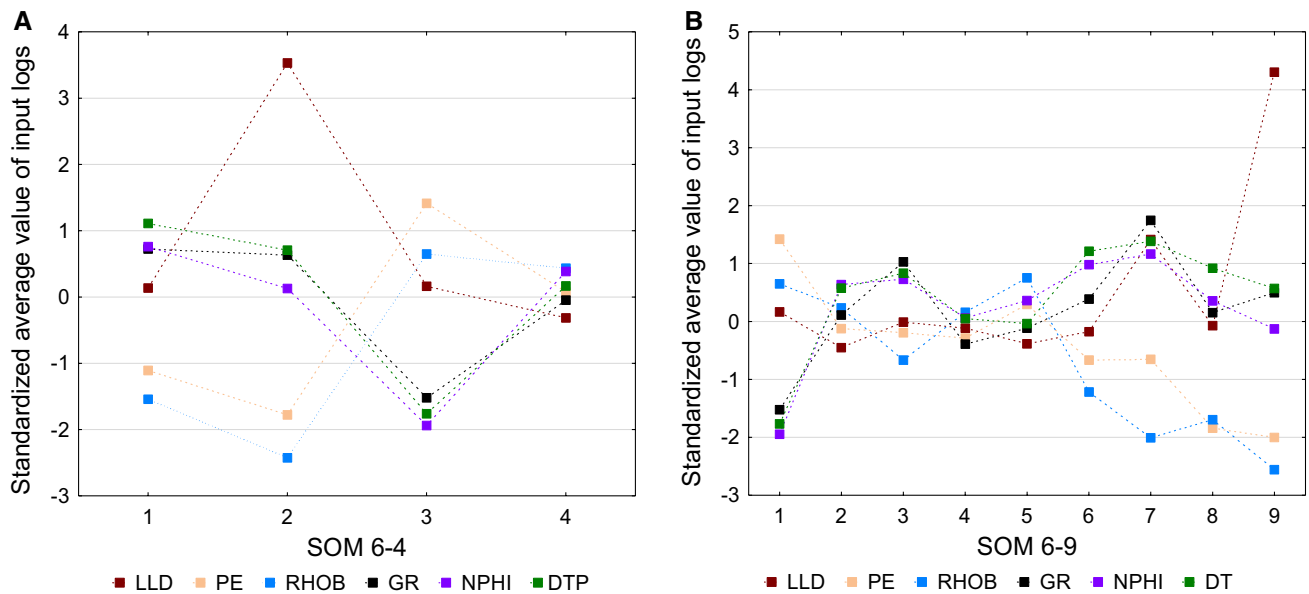


Fig. 6 Average values of standardized well logging data (differentiated by colours) in a division into four (a) and nine (b) clusters. On the x-axis the clusters numbers (ID) are marked

- electrofacies 2—samples assigned to this group were located in the upper parts of the Ja Mb and Sa Fm,
- electrofacies 3—entire Ko Fm and upper part of Pr Fm,
- electrofacies 4—entire whole Pa FM, lower part of Pr Fm interval, and lower parts Ja Mb and Sa Fm.

For more detailed information about electrofacies, we decide to use SOM 6-9 network. As a result, the nine electrofacies were obtained. The relation between SOM 6-4 and SOM 6-9 is as follows:

- from 193 samples from electrofacies 1 (in general division), 75 were assigned to the facies 8, 49 and 48, respectively, for the 6 and 7 facies and 19 samples for the 3 and 2 samples to facies 4,
- among 53 samples belonging to electrofacies 2 (in general division), 34 in the new division was assigned to facies 9, 7 in facies 7 and 9 in facies 8,
- from 203 samples belonging to electrofacies 3 (in general division), almost all samples (202) were assigned to new electrofacies 1, one sample was added to electrofacies 4,
- the most numerous electrofacies 4 (632 samples) was divided for four new electrofacies: facies 5 (335 samples), facies 4 (130 samples), facies 2 (119 samples) and electrofacies 3 (45 samples), 3 samples were assigned to electrofacies 6.

Neural networks were also used to examine the possibilities of machine learning classifications using the support vector machine algorithm. We apply this method for

classification for checking the ability to electrofacies analysis. In that method as a pattern during the learning process (we divide data set in a proportion 80% in a training set and 20% in a testing set), the information from Kohonen algorithm was used. A series of tests were carried out to select the best junction function. The best results were obtained for the linear and RBF (radial based function) kernels. The prediction accuracy was very high, above 98% and 97%, respectively, for linear and RBF kernel (Fig. 7). In results, more than 97% of the data were classified in the same way when we use SVM as when we use Kohonen Network. Such a result testifies to the high effectiveness of the method in the issues of classification.

Figure 8 depicts composition of the input data (well logging data) and the results of the classification using different methods. For a better visualization into the composition were added some information, e.g. gas volume (VGAS), kerogen volume (VKEROGER).

Discussion about the geological interpretation of electrofacies

Based on the artificial neural network, the Silurian/Ordovician interval in Well-1 was divided into 10 classes (electrofacies). The simplest network was built, so as an input only the standard well logging data were used (GR, DTP, RHOB, NPHI, LLD, PE). On the other hand, that logs contain information about lithology, porosity and saturation, so it was a good data set for electrofacies

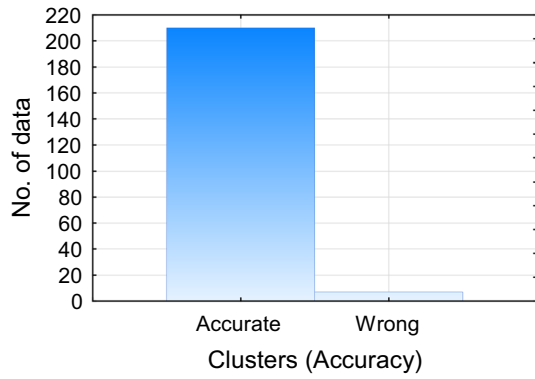


Fig. 7 Accuracy of SVM classification results. No. of data means the number of data points (in this case it is equal the depth points). More than 200 points from the testing data set were classified to the correct clusters (more than 97%)

prediction. During the network building and analysis information about lithostratigraphy, porosity and saturation were not taking into account. In that section, the

classification result will be joint with geological information and the characteristics of the electrofacies will be performed.

Figure 9 shows the relation between several parameters with a division into electrofacies.

Figure 9a, b shows dependencies between variables that were used to create classes. Based on the plots, it can be noticed the exact separation of classes:

- Electrofacies 1 (dark blue) is characterized by distinctly different properties from the other classes: high photoelectric index (around 4.75 average), density (about 2.7 average) and velocity testify to the fact that the samples belonging to this class are carbonates with low porosity.
- Electrofacies 8 and 9 are characterized by low PE (below 3 average) values, low density, and relatively low velocity, it can be a shale interval, saturated with gas.
- Electrofacies 6 and 7 are characterized by low velocity and low density, it can be a shale interval.
- Electrofacies 2, 3, 4 and 5 were divided from one general class (number 4 from SOM 6-4) and are character-

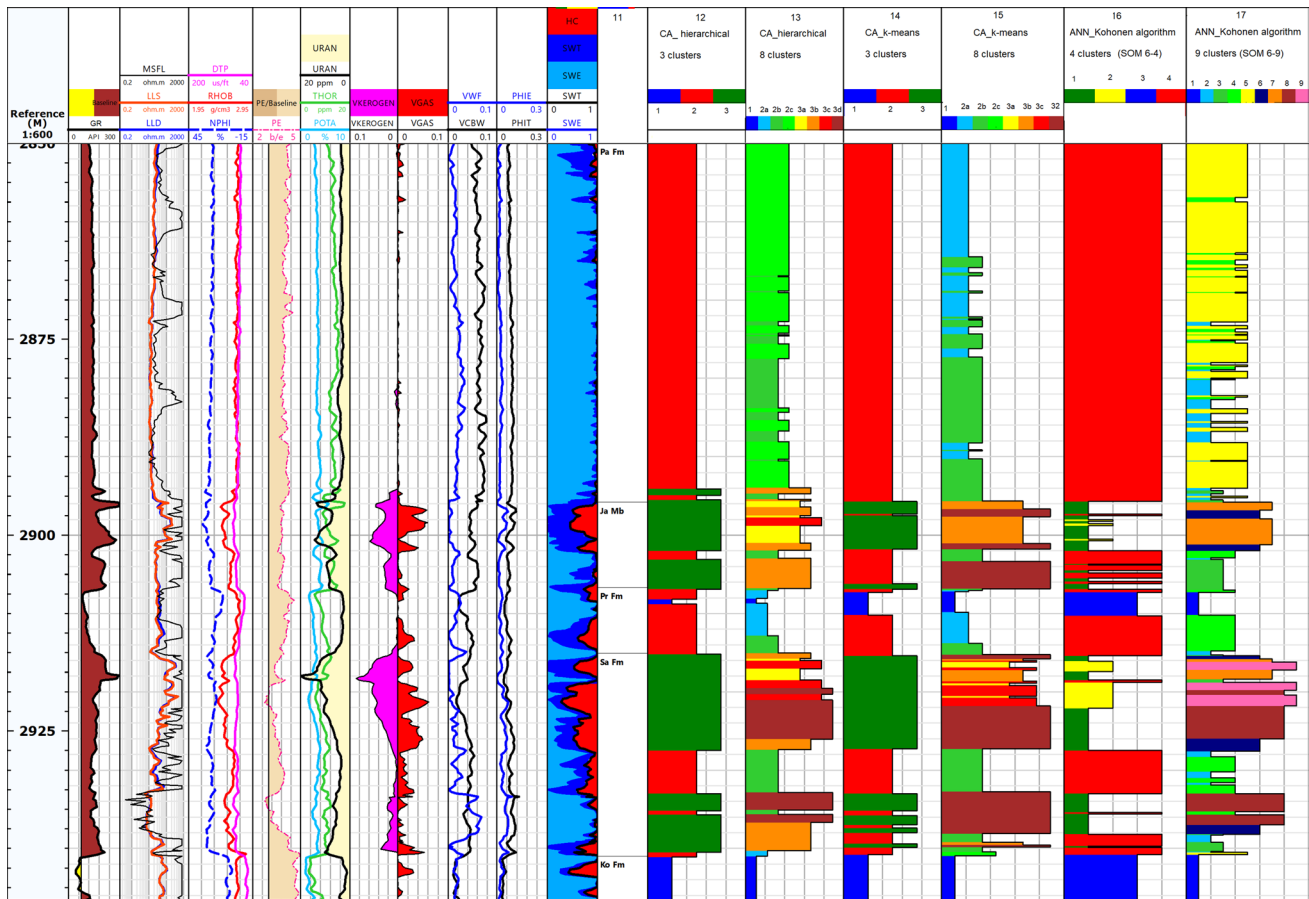


Fig. 8 Composition of input logs, results of interpretation and results of electrofacies prediction

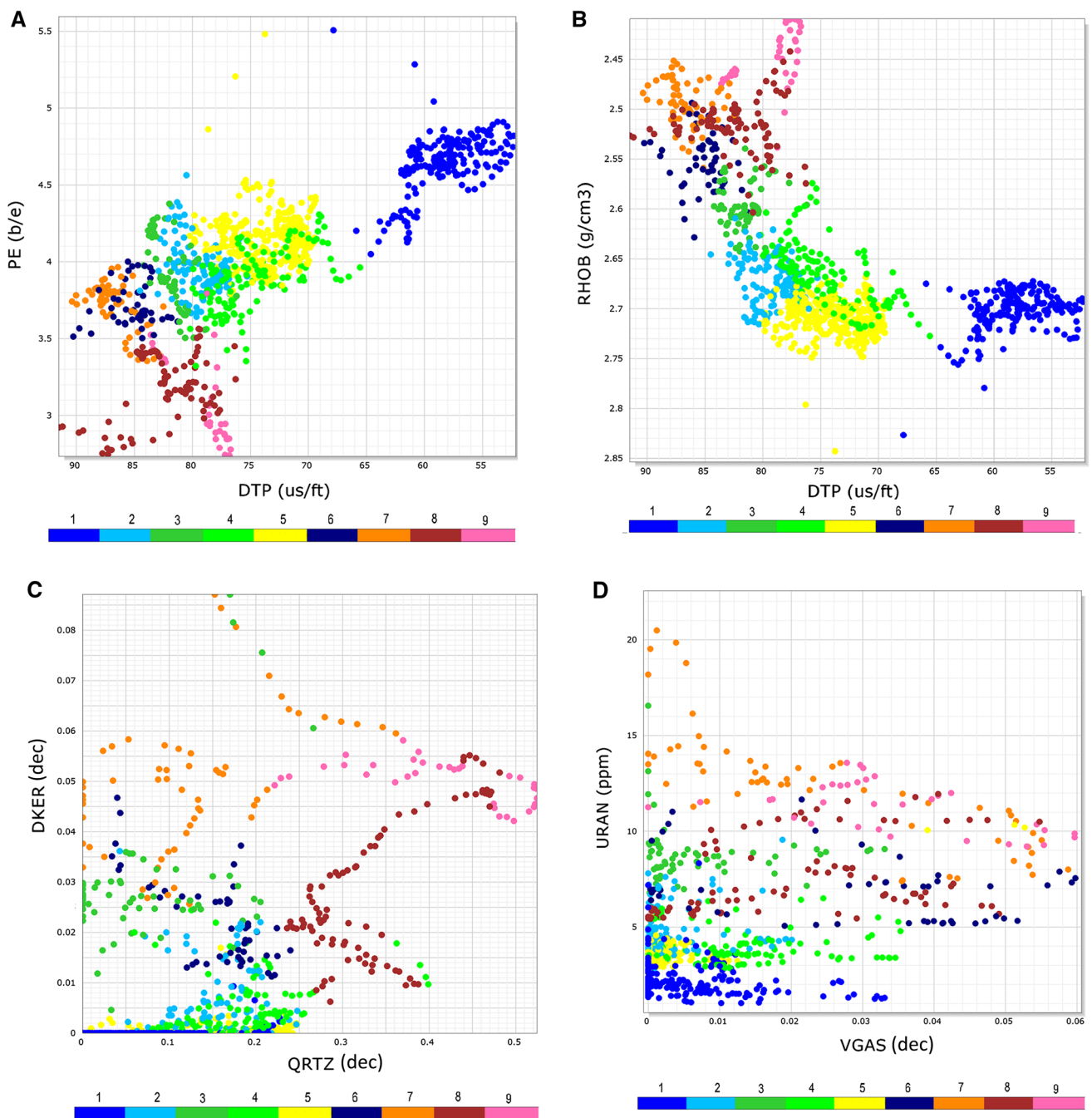


Fig. 9 Cross-plots depict the ANN classification results (colours) in the statement to the relation between input logs (DTP, RHOB, PE) and between other parameters [DKER—volume of the kerogen (dec),

VGAS—volume of gas (dec), URAN—uranium content (ppm), QRTZ—quartz mineral content (dec)]

ized by relatively average values of input logs, they were assigned to the shale classes.

Figure 9c, d depicts the relations between some results of logs interpretation. The results of classification also on that plots (created based on parameters which were not taken into account during network building) show electrofacies

separation. Based on that relations, it can be confirmed that electrofacies 7, 8, 9, which have the highest values of kerogen content and uranium content are the more perspective for hydrocarbon content.

In Fig. 10, the dependence of the gas content on the total organic carbon content was shown. Distinguished groups were differentiated in colors. Figure 10a shows the results

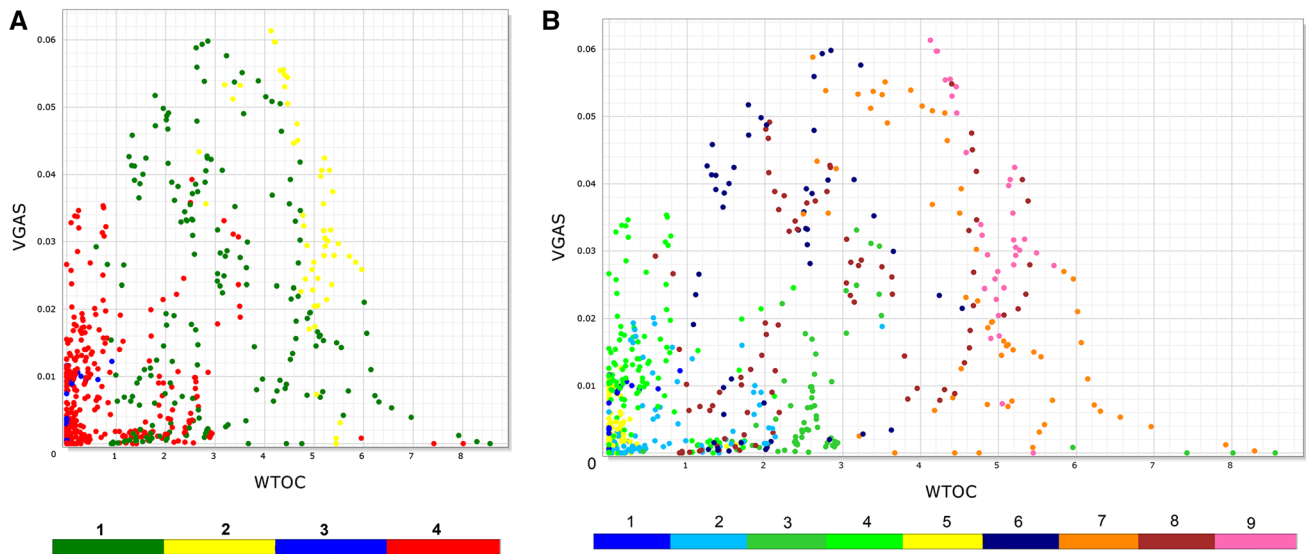


Fig. 10 Relation VGAS versus WTOC for two different classifications (different neurons number) using ANN. Symbols: VGAS—volume of gas (dec), WTOC—total organic carbon content (wt%)

of the SOM 6-4 network, while in Fig. 10b, the results for the SOM 6-9 network are plotted. This correlation confirms the conclusions that group 2 (in 4 classes division SOM 6-4) has been divided into subgroups 7, 8 and 9 in SOM 6-9. It is the most prospective for hydrocarbons presence (TOC above 4% wt and the highest gas volume). From group 1 (SOM 6-4), subgroups 8, 6 and 7 (SOM 6-9) were separated in particular. They have also high TOC values, above 3% wt. Electrofacies 6 has the highest gas content, and electrofacies 7 the highest TOC content.

A box-and-whisker plots (Fig. 11) were created to display graphically the influence of clustering into the analysed interval. If the box is long, log values are very dispersed. The structure of the box plot is the same for each variable (litho-porosity solution based on GEM™ Elemental Analysis Tool results) and is as following:

- the 10th percentile (10% of the log values which corresponds to the outliers), lower whisker,
- the first quantile (25% of the log values), lower box border,
- the median (50% of the log values), line through the box,
- the third quantile (75% of the log values), upper box border,
- the 90th percentile (90% of the log values), upper whisker,
- black dots correspond to the minimum and maximum values.

Based on dependencies summarized in Fig. 11 and previously described results, geological interpretations of electrofacies have been established (note: 1) in the sub-point

stratigraphy, value in bracket is the number of samples/depth; 2) av. as an abbreviation of average):

1. **Electrofacies 1** (202 samples):

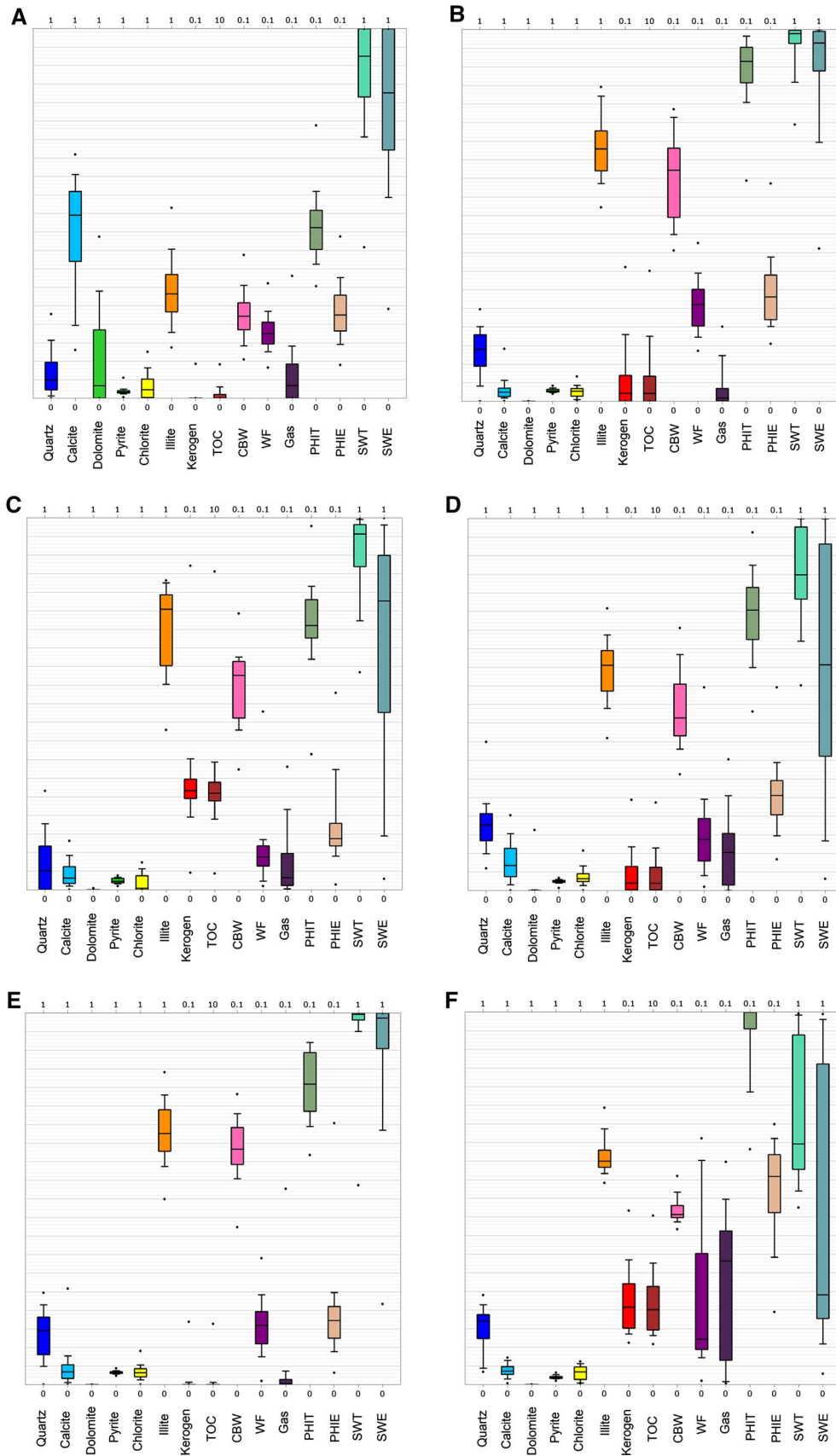
- Lithology consists mainly of calcite, illite and dolomite.
- Stratigraphy: Ko Fm (170), Pr Fm (29), Sa Fm (3).
- Porosity: av. total equal 4.5%, av. effective equal 2.3%.
- Saturation: av. total water saturation equal 88%, av. effective water saturation equal 80%, gas volume 0.5%.
- Organic matter content: negligible quantities.
- Reservoir potential: very low.

2. **Electrofacies 2** (119 samples):

- Lithology consists mainly of illite and quartz.
- Stratigraphy: Pa Fm (84 samples), Sa Fm (30 samples), Pr Fm (3), Ja Mb (2).
- Porosity: av. total equal 9%, av. effective equal 3%.
- Saturation: av. total water saturation equal 96%, av. effective water saturation equal 90%, gas volume 0.3%.
- Organic matter content: av. TOC 0.5% wt, av. kero-gen content 0.5%.
- Reservoir potential: low.

3. **Electrofacies 3** (64 samples):

- Lithology consists mainly of illite.
- Stratigraphy: Ja Mb (37), Sa Fm (18), Pa Fm (9).
- Porosity: av. total equal 7%, av. effective equal 1.6%.



◀ **Fig. 11** Box-and-whiskers plots, describing the distribution of litho-porosity interpretation results for selected electrofacies (11A—electrofacies 1, 11B—electrofacies 2, 11C—electrofacies 3, 11D—electrofacies 4, 11E—electrofacies 5, 11F—electrofacies 6, 11G—electrofacies 7, 11H—electrofacies 8, 11I—electrofacies 9). Symbols: **a** Quartz, calcite, dolomite, pyrite, chlorite, illite—mineral contents in decimal, range from 0 to 1 (dec), **b** kerogen content calculated for dry rock, on the plot values in the range from 0 to 0.1 (dec), **c** TOC—total organic carbon content, on the plot values in the range from 0 to 10 wt%, **d** CBW—clay bound water (dec), WF—free water (dec), Gas—gas volume (dec), PHIT—total porosity (dec), PHIE—effective porosity (dec), on the plot values in the range from 0 to 0.1 (dec), **e** SWT—total water saturation (dec), SWE—effective water saturation (dec), range from 0 to 1 (dec)

- Saturation: av. total water saturation equal 90%, av. effective water saturation equal 65%, gas volume 2%.
- Organic matter content: av. TOC content 2.8% wt (max. 8.5% wt), kerogen content about 3% (max. 9%).
- Reservoir potential: medium.

4. **Electrofacies 4** (133 samples):

- Lithology consists mainly of illite, additionally occurs quartz and calcite.
- Stratigraphy: Pr Fm (48), Sa Fm (40), Pa Fm (36), Ja Mb (9).
- Porosity: av. total equal 7.4%, av. effective equal 2.6%.
- Saturation: av. total water saturation equal 85%, av. effective water saturation equal 60%, gas volume 1%.
- Organic matter content: av. TOC equal 0.4% wt, av. kerogen volume 0.4%.
- Reservoir potential: low.

5. **Electrofacies 5** (335 samples):

- Lithology: mainly illite and quartz.
- Stratigraphy: Pa Fm (332), Sa Fm (3).
- Porosity: av. total equal 8%, av. effective equal 1.7%.
- Saturation: av. total water saturation equal 97%, av. effective water saturation equal 90%, gas volume 0.1%.
- Organic matter content: av. TOC equal 0.03% wt (max. 1.6 wt%)
- Reservoir potential: very low.

6. **Electrofacies 6** (51 samples):

- Lithology: mainly illite and quartz.
- Stratigraphy: Sa Fm (32), Ja Mb (19).
- Porosity: av. total equal 10%, av. effective equal 5%.
- Saturation: av. total water saturation equal 71%, av. effective water saturation equal 45%, gas volume 2%.
- Organic matter content: av. TOC equal 1.2 wt% (max. 4.5%), av. kerogen volume equal 2.2%.

- Reservoir potential: medium/high.

7. **Electrofacies 7** (59 samples):

- Lithology: mainly illite and quartz.
- Stratigraphy: Ja Mb (43), Sa Fm (16).
- Porosity: av. total equal 7.3%, av. effective equal 2.6%.
- Saturation: av. total water saturation equal 70%, av. effective water saturation equal 21%, gas volume 2.3%.
- Organic matter content: av. TOC equal 4.8 wt% (max. 8.6 wt%), kerogen volume equal 5%.
- Reservoir potential: high.

8. **Electrofacies 8** (84 samples):

- Lithology: illite and quartz.
- Stratigraphy: Sa Fm (84).
- Porosity: av. total equal 8.5%, av. effective equal 5%.
- Saturation: av. total water saturation equal 74%, av. effective water saturation equal 56%, gas volume 2%.
- Organic matter content: av. TOC equal 2.7 wt% (max. 5.4 wt%), kerogen volume equal 2.8%.
- Reservoir potential: medium.

9. **Electrofacies 9** (34 samples):

- Lithology: quartz and illite.
- Stratigraphy: Sa Fm (34).
- Porosity: av. total equal 6.3%, av. effective equal 3.8%.
- Saturation: av. total water saturation equal 47%, av. effective water saturation equal 11%, gas volume 3.4%.
- Organic matter content: av. TOC equal 5 wt% (max. 5.7 wt%), kerogen volume equal 5%.
- Reservoir potential: high.

Conclusions

On the basis of Kohonen network and cluster analysis, electrofacies were found and separated from the Silurian and Ordovician interval. Electrofacies as a groups that gave the same or similar response to the input logs used for classification (GR, RHOB, DTP, NPHI, PE, LLD) were analysed in terms of their reservoir potential.

In conclusion, the obtained results were found:

- both cluster analysis and Kohonen networks are a good tool for determining electrofacies,
- the use of basic logs allows for proper electrofacies determination,
- the most detailed classification results were obtained by using the SOM 6-9 network,

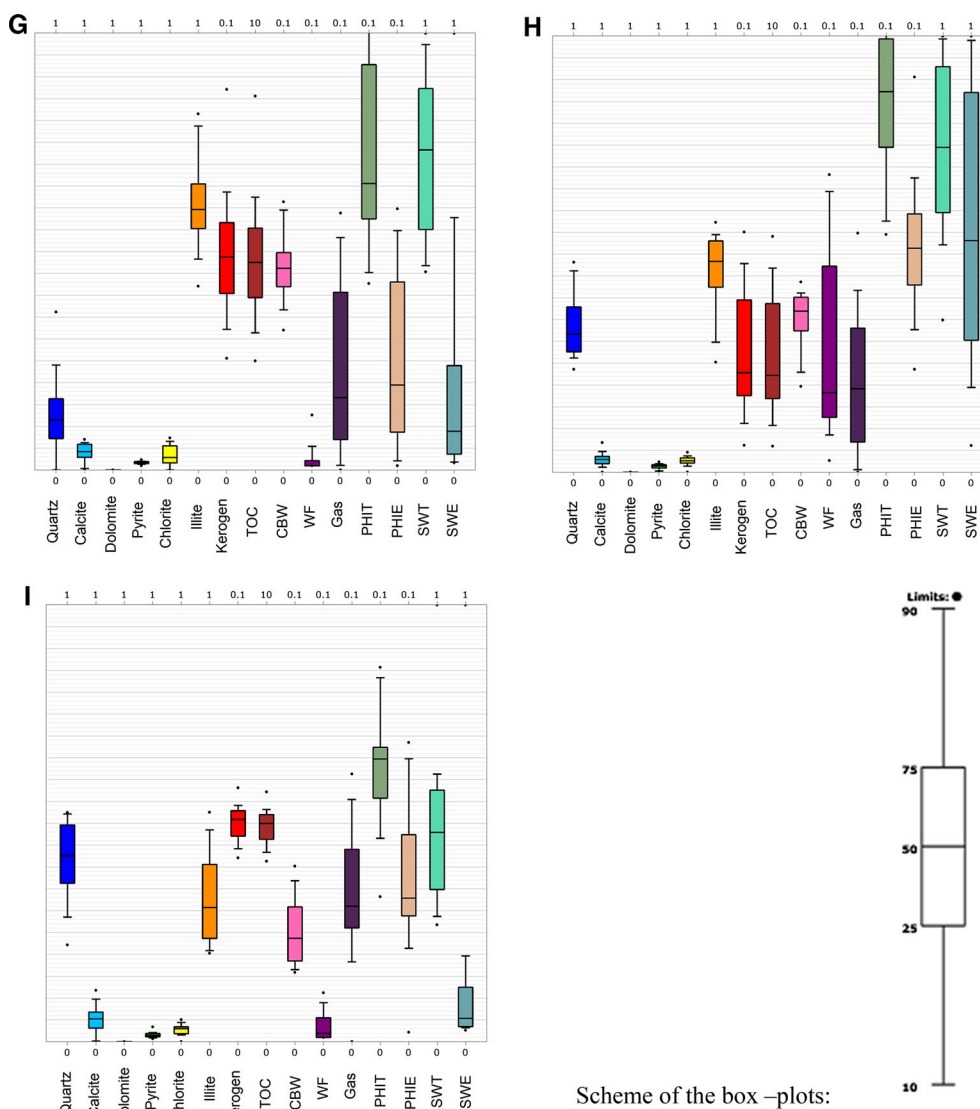


Fig. 11 (continued)

- SMV confirms the correctness of the SOM 6-9 classification,
- nine electrofacies were distinguished, which partially overlap with the lithostratigraphic division,
- the largest internal differentiation was observed in the Ja Mb and Sa Fm (treated as a *sweet spots*),
- electrofacies 7 and 9 were considered the most prospective for the production of hydrocarbons,
- electrofacies 6, 3 and 8 were considered to have medium reservoir potential,
- electrofacies 1, 2, 4, 5 were considered the less prospective.

Acknowledgements Data were allowed by POGC, Warsaw, Poland, for the research conducted during the Blue Gas program project titled:

“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” (MWSSSG) Polskie Technologie dla Gazu Łupkowego (2013–2017). Statistica 13 software was used according to the AGH University grant. Paper was financially supported by the Faculty of Geology Geophysics and Environmental Protection AGH UST funds in 2019 year. Paper was financially supported by the research subsidy No. 16.16.140.315 at the Faculty of Geology Geophysics and Environmental Protection of the AGH University of Science and Technology, Krakow, Poland, 2019.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>)

[mmons.org/licenses/by/4.0/](https://creativecommons.org/licenses/by/4.0/)), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Doveton J (1994) Geologic log analysis using computer methods. AAPG Computer Applications in Geology, vol 2, Tulsa
- Hair JF Jr, Black WC, Babin BJ, Anderson RE, Tatham RL (2006) Multivariate data analysis. Pearson Prentice Hall, New Jersey
- Jarzyna JA, Krakowska PI, Puskarczyk E, Wawrzyniak-Guz K, Zych M (2018) Comprehensive interpretation of the laboratory experiments results to construct model of the polish shale gas rocks. In: POL-VIET 2017: scientific-research cooperation between Vietnam and Poland: Krakow, Poland, November 20–22, 2017; E3S Web of conferences, vol 35, art. no. 03009, pp 1–8. ISSN 2267-1242
- Kohonen T (1982) Biol Cybern 43:1
- Krakowska PI, Jarzyna JA, Wawrzyniak-Guz K, Puskarczyk E, Zych M (2016) Heterogeneity analysis of the Polish shale gas formations based on results of laboratory measurements. In: International Multidisciplinary Scientific GeoConference SGEM, pp 817–823. ISSN 1314-2704; ISBN: 978-619-7105-57-5
- Modliński Z, Szymański B (2008) Lithostratigraphy of the Ordovician in the Podlasie depression and the basement of the Płock-Warsaw trough (Eastern Poland). Biul Państw Inst Geol 430:79–112 (**in Polish**)
- Modliński Z, Szymański B, Teller L (2006) The Silurian lithostratigraphy of the Polish part of the Per-Baltic depression (N Poland). Prz Geol 54(9):787–796 (**in Polish**)
- Moss B (1997) The partitioning of petrophysical data: a review. In: Lovell MA, Harvey PK (eds) Developments in petrophysics, vol 122. Geological Society Special Publication, pp 181–252
- Passey QR, Creaney S, Kulla JB, Moretti FJ, Stroud JD (1990) A practical model for organic richness from porosity and resistivity logs. Bull Am Assoc Pet Geol 74(12):1777–1794
- Poprawa P (2010) Potencjal występowania złóż gazu ziemnego w lupkach dolnego paleozoiku w basenie bałtyckim i lubelsko-podlaskim. Prz Geol 58:226–249
- Puskarczyk E (2017) Shale gas formation heterogeneity analysis using multidimensional statistical methods (PCA, FA, CA, ANN) on the basis of well logging and laboratory investigations. In: Monography R, Jarzyna J, Wawrzyniak-Guz K (eds) 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. GOLDRUK Wojciech Golachowski Printing House (**in Polish**)
- Puskarczyk E (2018) Applying of the artificial neural networks (ANN) to identify and characterize *sweet spots* in shale gas formations. In: E3S Web of conferences, 2018, vol 35, art. no. 03008, pp 1–7. ISSN 2267-1242
- Puskarczyk E, Jarzyna J, Sz Porębski (2015) Application of multivariate statistical methods for characterizing heterolithic reservoirs based on wireline logs—example from the Carpathian Foreland Basin (Middle Miocene, SE Poland). Geol Q 59(1):157–168. <https://doi.org/10.7306/gq.1202>
- Rider M (2002) The geological interpretation of well logs. Rider-French Consulting Ltd., Scotland
- Sebtosheikh MA, Motafakkerfard R, Riahi MA, Moradi S, Sabety N (2015) Support vector machine method, a new technique for lithology prediction in an Iranian heterogeneous carbonate reservoir using petrophysical well logs. Carbonates Evaporites 30:59–68
- Serra O, Abbott HT (1980) The contribution of logging data to sedimentology and stratigraphic. In: SPE 9270, 55th annual fall technical conference and exhibition, Dallas, Texas
- Serra O, Serra L (2004) Well logging data acquisition and application. Serra Log, Méry Corbon
- Sowizdzał K, Stadtmüller M, Lis-Śledziona A, Kaczmarczyk W (2016) Analiza porównawcza formacji łupkowych w wybranych strefach basenu bałtyckiego na podstawie interpretacji danych otworowych i wyników modelowania geologicznego 3D. Nafta-Gaz, no.11, pp 891–900 (**in Polish**)
- 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>
- Szabó NP, Dobroka M, Kavanda R (2013) Cluster analysis assisted float-encoded genetic algorithm for a more automated characterization of hydro-carbon reservoirs. Intell Control Autom 4:362–370
- Wawrzyniak-Guz K, Puskarczyk E, Krakowska PI, Jarzyna JA (2016) Classification of Polish shale gas formations from Baltic Basin, Poland based on well logging data by statistical methods. In: International multidisciplinary scientific GeoConference SGEM, pp 761–768. ISSN 1314-2704; ISBN: 978-619-7105-57-5
- Wong KW, Ong YS, Gedeon TD, Fung Ch. Ch. (2005) Reservoir characterization using support vector machines. In: Proceedings of the 2005 international conference on computational intelligence for modelling, control and automation, and international conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'05)