Identification of palaeoenvironments using clustering methods and indicator kriging, case study from Late Miocene sandstones, the Sava depression

J. Horváth, K. N. Zelenika

ORIGINAL SCIENTIFIC PAPER

The clustering and Indicator Kriging methods are standard exploration tools in some kind of geological problems and here is presented a case study for the identification of depositional environments using two kinds of clustering processes and the Indicator Kriging. The clustering is based on Kohonen's neural network and hierarchical agglomerative clustering. Through the resulting clusters, the geometry of depositional sub-environment can be visualised by spatial mapping processes and it can be interpreted by the probabilities of clusters through the Indicator Kriging. In this manner the clustering tools can be defined as a spatial pattern recognition method. The research area in the demonstrated study is located in the Sava Depression, Croatia. The field sediments belong to Neogene and Quaternary, and the analysed are Upper Pannonian when sedimentation was accompanied by influence of pre-Neogene basement palaeorelief and the accumulation by turbidity system.

Key words: clustering, palaeoenvironment, Kohonen's neural network, hierarchical agglomerative clustering, Indicator Kriging, Sava Depression, Croatia

INTRODUCTION

The study area is located in the Sava Depression, Croatia, about 35km to the east of Zagreb (Figure 1). The field sediments belong to Neogene and Quaternary. Sedimentation was accompanied by influence of pre-Neogene basement palaeorelief which has heterogeneous lithological composition.

The depositional model of the analysed area is well-known. The lithologically analysed sequence is represented by Late Miocene marls, silts and sandstones, deposited with turbidity current mechanisms. Sedimentation was a combination of hemipelagic-like pelitic deposition and cyclic turbidity flows in lacustrine environments (Figure 2), that area was constantly reducing.[19] Sandstones morphology follows the direction of depositional currents[14], and generally in the central part of the depression appeared thick, bedded, fine-grained sandy deposits with relatively high porosity between 15-25% which is bounded in marginal part with basin plain marl sediment. The main direction of accumulation was NW-SE (locally N-S) and the main material came from Eastern Alps. Detritus was redeposited several times before it was finally accumulated.[9,15]

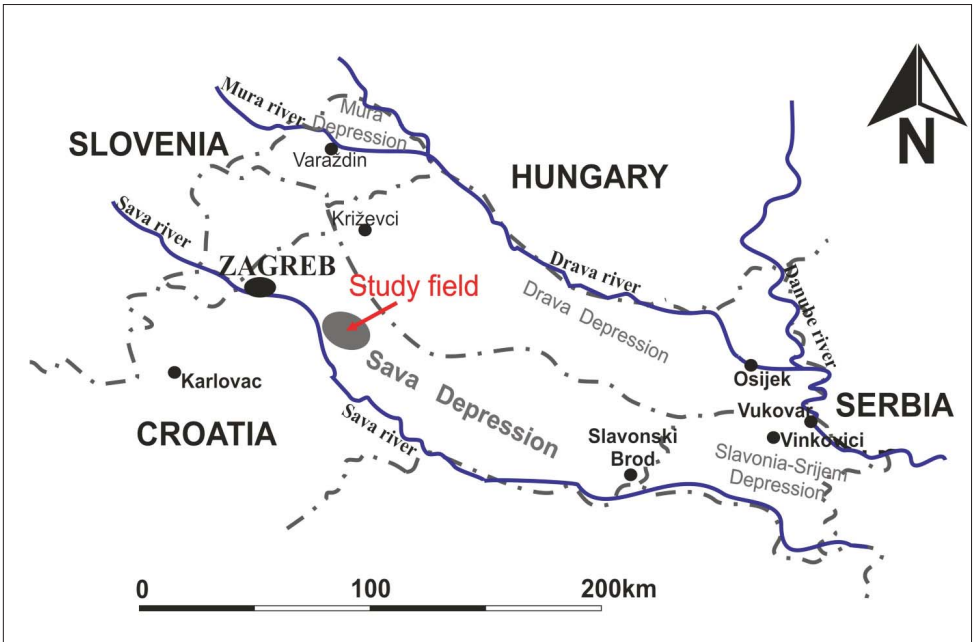

Fig. 1. Location of the study area
Sl. 1. Položaj proučavanog područja

The researching object was the demonstration of a methodology to recognize and define depositional palaeoenvironments and sub-environments (in system sandstone-marl), using clusters based on some petrophysical parameters (porosity, water saturation, shale volume). Two clustering methodologies are presented and both of them can handle simultaneously point data with more attributes or property jointly. The Kohonen's neural network (abr. KNN) method and one kind of multidimensional hierarchical clustering methods were applied in the research. In both clustering processes, the inputs were the basic petrophysical parameters and a binary descriptor of the lithology. This latter one, however, was modified to be a categorical variable for eleven reservoirs according to the shale content of sandy deposit. There are total 88 wells in the study area, but well logging data and analyses were available from 79 locations that are prepared for the clustering model.

In the entire sedimentation sequence, 11 hydrocarbon pools have been defined in Late Miocene sediments. That reservoir "series" consists of fine-grained, quartzite sandstones alternating with marls. The total vertically averaged parameters of the 11units (the total thickness changes between 120-150 m.) would be quite large scaled for exact identification of palaeoenvironments. Because of it only one reservoir of the 11 reservoir pools are modelled. It is the most representative with the largest thicknesses and consequently the most developed e-logs and again consequently the most reliable calculated petrophysics. The selected reservoir pool is labelled as "*study reservoir*" that location according to the chronostratigraphy and depositional environments is shown in Figure 2.

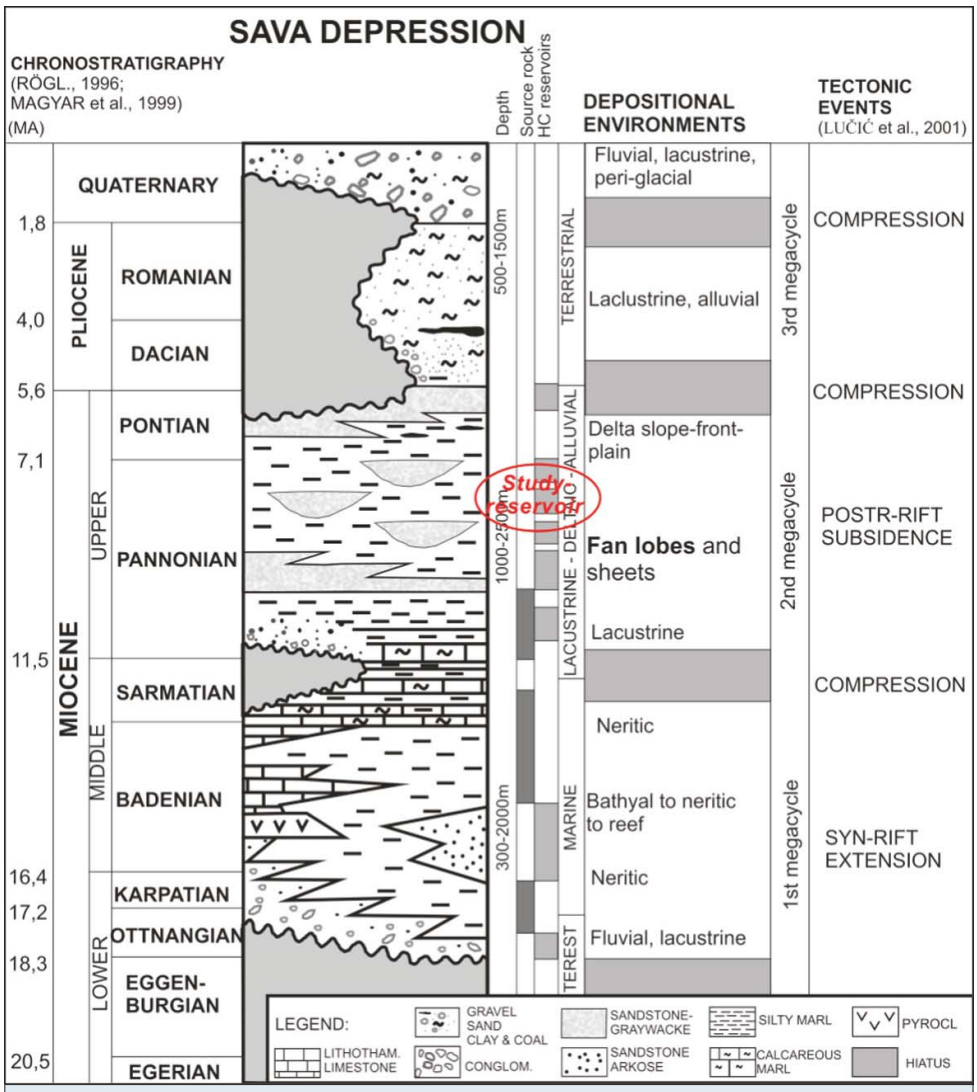

Fig. 2. The location of the study reservoir according to the chronostratigraphy and depositional environments[7,8,13,14]

Sl. 2. Položaj proučavanog ležišta s obzirom na kronostratigrafiju i taložni okoliš[7,8,13,14]

2. APPLIED METHODS

This study applied two robust clustering methods for characterizations of reservoir rock bodies and identification of depositional environment. The clustering processes were Kohonen's neural network[4] and one kind of classical multidimensional analysis method named as hierarchical agglomerative clustering.

The question may arise why the clustering process is used for pattern recognition and identification of environment. Since any clustering method (e.g. Kohonen's network) gives an opportunity to provide solutions for geological problems which can be drawn back to different clustering problems, such as pattern recognition.[18] The clustering processes have the ability to recognize spatial patterns being defined as spatial groups of points bounded by polygon. Shapes are defined as clusters trough data points that are bounded by polygon, it implies the pattern recognition as a clustering problem.[3]

The term clustering was used by the first time by Tryon[17], who defined this method to divide data into groups (clusters). Cluster means the object with similar properties, so cluster analysis simply discovers characteristic structures in data. The cluster analysis is not

based only on the objects of data-groups or their qualitative description, but also on their mathematical relationships.

However, this approach raises an additional question of definition of patterns as statistical groups. Within a pattern as a cluster there are points with the same or similar property that are more or less homogeneous sets. It practically means that all points in a sub-set represent a certain cluster which is bounded by polygons. A polygon represents the shape and reflects the borders of environments, in these study sub-environments of sandstone and marl, and sub-environment of sandy marl and marlitic sandstone (depend on the scale), where the properties change between separated clusters.[3] After the clustering process the resulting clusters are mapped using the original coordinates. The pattern of clusters are visualized on the map and additional analyses, as here applied Indicator Kriging (abbr. IK), could show probability of particular cluster in the individual grid points.

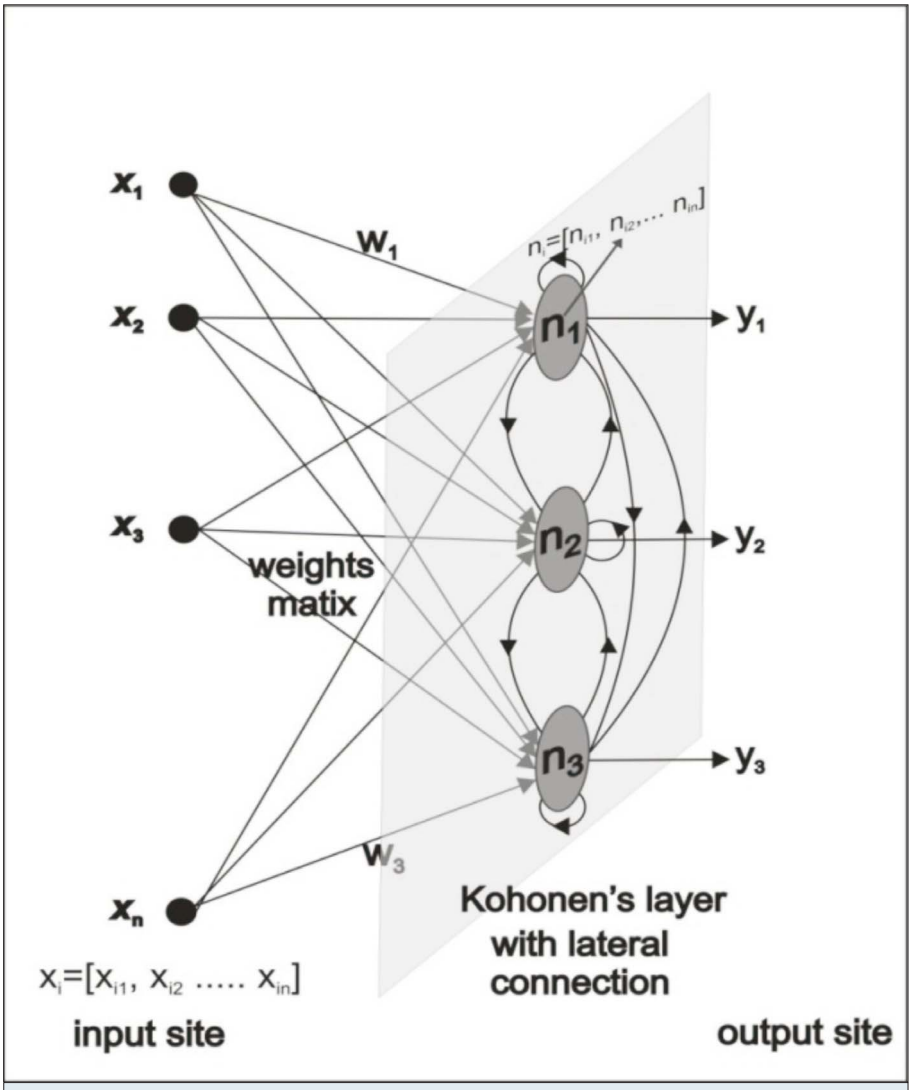

Fig. 3. Framework of the unsupervised neural network of three competing neurons, with n-dimensional neurons in the Kohonen's layer
Sl. 3. Okvir nenadzirane neuronske mreže triju neurona koji se međusobno "natječu", s n-dimenzionalnim neuronima u Kohonenovom sloju

2.1. KOHONEN'S NEURAL NETWORK (KNN)

This method is an unsupervised neural network, i.e. the training procedure using unsupervised learning to arrange the input data. Usually Kohonen's neural network is the so-called Self Organized Map (abbr. SOM) because it has to arrange the original data set on a map.[5] However it is not a real map since it does not assign any spatial coordinates to the samples, so the neurons are organized only in a gird as a 'map'. The other important property of this artificial network is that SOM is able to arrange the input data without dimension loss in the multidimensional space but we often can read it is a clustering method which is based on a dimension reduction way. Actually, it means only quasi reduction since each neuron is n-dimensional weight-vector, where n is equal to the dimension of the input vectors. Also, the framework of SOM reflects the difference from supervised artificial networks and demonstrates the quasi-reduction way in the organizing process. There are not any real output layers, but neurons are arranged into grid dots which is the "output" layer, the so-called Kohonen's layer.

The SOM, through the above mentioned properties, has the ability to recognize the structure of a multidimensional basic data set and use neighborhood function to preserve the topological contacts among the inputs.

Kohonen's network learns the data structure on an iterative way using the modified Hebb's rule (1) as winning neuron algorithm.[6] The modified rule is:

$$w_{t+1} = w_t \eta_t \cdot (x - w_t) \quad (1)$$

where:

x	the training case,
w	weight,
η	the learning rate
t	discrete time iteration.

Inputs are processed by all neurons (Figure 3) shows a case with three neurons on Kohonen layer). Each neuron computes its weighted input, but only the neuron with the largest activation is the winner. These units are adjusted to cluster the training data using the activation values of winning neurons. In Kohonen's layer the other competing neurons are reserved by this active element

through the lateral connections (neighborhood function) during training (Figure 3).

2.2. HIERARCHICAL AGGLOMERATIVE CLUSTERING (HAC)

In hierarchical clustering method, all relate to segmenting collection set of objects (data points) into subsets (clusters). The relation within each cluster is stronger than the relation among clusters. Primary goal of cluster analysis is calculation of degree of similarity (or dissimilarity) between the individual objects in cluster. Hierarchical agglomerative clustering method is used here which is based on cosine distance and centroid clustering. Formed groups are represented by their centroid for each variable that is their mean vector. Inter-group distance is now defined in terms of distance between two such mean vectors. This distance means the reduction rate in the merging procedure. During the group averaging using HAC algorithm, two clusters are merged such that the centroid pair-wise distance, within the newly formed cluster, is minimal.

If a function is based on a distance of the pairs of points, and its value is between 0 and 1, we used similarity to assign a metric, e.g., cosine distance. This is often used to measure cohesion within clusters. Cosine similarity is a measure of similarity between two vectors by measuring the cosine of the angle between them. Calculating the cosine of the angle between two vectors thus determines whether two vectors are pointing in approximately the same direction. The resulting similarity ranges from -1 meaning exactly opposite, to 1 meaning exactly the same. The value 0 usually indicating independence, and in-between values indicating degree of similarity (or dissimilarity).

$$similarity = \cos(\phi) = \frac{A \cdot B}{\|A\| \cdot \|B\|} = \frac{\sum_{i=1}^{n} a_i b_i}{\sqrt{\sum_{i=1}^{n} (a_i)^2} \cdot \sqrt{\sum_{i=1}^{n} (b_i)^2}} \quad (2)$$

where:

A and B	n-dimensional vectors,
a_i and b_i	i^{th} variable of A and B vector,
$\|\cdot\|$	*a vector norm.*

Remark (i):

The correlation coefficient means really the cosine of the angle between two vectors. So, in the case of the results of correlation analysis the largest correlation coefficient means the nearest vectors, i.e. its most similar attributes.

In the case of the centroid clustering, during the group-linking processes the above mentioned cosine similarity (2) of two clusters is defined as the similarity of their centroids[10]:

$$similarity - centroid(C_i, C_j) =$$

$$= \bar{k}(C_i) \cdot \bar{k}(C_j) = \left(\frac{1}{N_i} \sum_{d_m \in C_i} \vec{d}_m \right) \left(\frac{1}{N_j} \sum_{d_m \in C_j} \vec{d}_n \right) \quad (3)$$

$$= \frac{1}{N_i \cdot N_j} \sum_{d_m \in C_i} \sum_{d_n \in C_j} \vec{d}_m \cdot \vec{d}_n \quad (4)$$

Where:

C_i	is the i^{th} cluster,
k_i	is the centroid of the i^{th} cluster,
d_n	is the n^{th} element in one of the clusters,
N_i	is the number of elements in the i^{th} cluster.

Equation (3) is a centroid similarity, and (4) reflects that the centroid similarity is equivalent with the average similarity of all pairs of data set from different clusters. The group average agglomerative clustering algorithm on the contrary to other HAC algorithm is not monotonous and the similarity does not reduce during the process. In the monotonous HAC algorithm the similarity can monotonically decrease step by step in the iteration; however small clusters can be more coherent than large clusters.[10]

Remark (ii):

The differences between the SOM and HAC methods:

- The hierarchical multidimensional algorithm is a quick methodology. Its disadvantage is that the clusters include especially the properties of the first data-vectors because they start to define the centroid of the first clusters with more weight. After that the processes of merging of clusters is based only on the distance of reduction and not the change of weight.
- The hierarchical multidimensional algorithm is not able to detect non-linear relation
- The question is the verifiability of the procedure if it do not use the distance of reduction between centroid, i.e. use a method that is based on the probability of belonging to clusters or SOM model.
- The SOM procedure is a slower multi-step clustering process but also robust and is able to detect non-linear relations.

2.3. MAPPING PROCESS

Figure 4 demonstrates an example of mapping of clusters using the original coordinates. HAC and SOM create Cm clusters (m is the number of clusters), which are based on property values of data points without original spatial coordinates. After this clustering method the mapping processes compose the closed patterns by the arrangement of the points in space through the original coordinates (xi,yi). In this manner clusters with different properties and the pattern are mapped. However, extension of clusters around the known points and defining of closed pattern depend on the mapping processes.

Classical isoline's interpolation cannot be used for mapping of clusters because there are not possible estimate values between points, i.e. cluster areas are in direct contact. The numbers of clusters are not values, only labels. However we can estimate which cluster can appear between the measured points. Even so, we can use one of the interpolation algorithms in the mapping procedure if we follow only assumed anisotropy and Tobler's law[16], which states that the spatial interpolation is based on the assumption that the similarities of value of closely points show higher probability than the points situated far from each other. In the case of extension of clusters it means that the closer situated points belong to the same

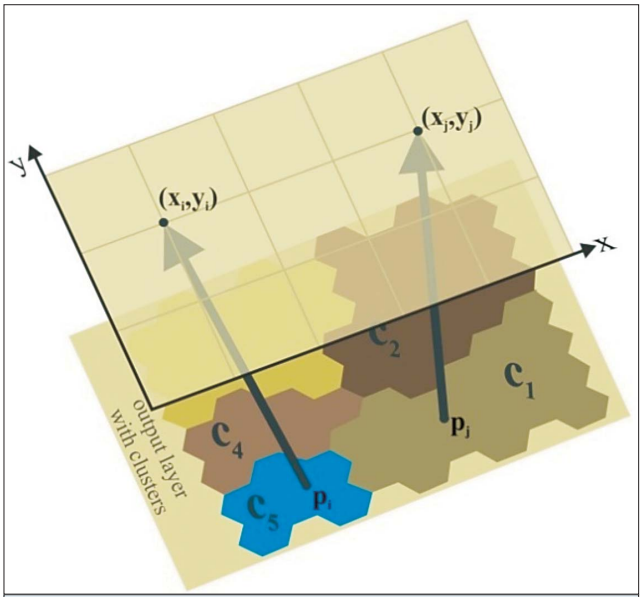

Fig. 4. Arrangement of points on the map
Sl. 4. Uređivanje točaka na karti

or similar clusters with higher probability. On the other hand it is also worthy to consider that the clusters reflect specified geometry which usually also follows orientation (structural, sedimentological etc.) and anisotropy can describes such orientations during the mapping processes. Therefore the applied algorithm was the Radial Basis Function (abbr. RBF) interpolation algorithm with multiquadriatics method[2] was applied in Surfer 8.0. The direction and ratio of orientation are derived from the physical properties of the variables used in clustering processes.

However, the presented approach of the clusters mapping is not an exact method. It is why we applied also IK as method for mapping probability of the clusters. IK uses categorical variable, like discrete number of labels clusters (called categories) where labels are defined as $Z(u) \in \{0,\ldots k-1\}$, where k is the number of clusters and indicator variable by (5):

$$I(u,k) = \begin{cases} 1, \text{ if } Z(u) = k \\ 0, \text{ otherwise} \end{cases} \quad (5)$$

In each grid points in the final cluster are defined according to the highest probabilities. In this manner the resulting probabilities are filtered by a logical function (6).

$$IF(p_k) = \begin{cases} C_i, p_k = MAX(p_1,p_2,\ldots p_n) \\ 0, p_k \neq MAX(p_1,p_2,\ldots p_n) \end{cases} \quad (6)$$

where

IF	is a mathematical logical function (where the function values is fixed values or 0)
C_i	is the i^{th} cluster
p_k	is the probability of k^{th} cluster in a grid point
n	is the number of clusters

3. INTERPRETATION OF RESULTS

The interpretation of the formed clusters is based on some analytical and statistical application such as discriminant analysis and Indicator Kriging. Besides the identification of these clusters were possible by the shapes of the mapped clusters and their averages for each petrophysical descriptors.

3.1. IDENTIFICATION OF CLUSTERS

In both clustering models the inputs were the basic petrophysical parameters (porosity, water saturation, shale volume) and a categorical variable that describes the lithology using a code-number between 0-10 according to the shale content of the sandy deposit. The models SOM and HAC segregated four clusters (C_1-C_4). The interpretation of the results of applied models for the "Study reservoir" is presented in comparison. In the reconcilement and interpretation of the mapped results of the different models, it is worthy to consider the average of the physical parameters of clusters (Tables 1-2). In this way the averages defined the clusters, e.g., C_1 encompasses the lowest, and C_4 the highest porosity value. On the other hand the classification matrices of discriminant analyses (Tables 3-4) are useful to adjudge how current classification functions predict the group membership. In the first column of the classification matrix, the percentage of cases that are correctly classified in each cluster can be seen. The other columns show the percentage of misclassifications.

Table 1. Description of SOM-clusters by average of petrophysical parameters

Cluster	Facies	Water saturation	Shale volume	Porosity
C_1	(F_4) Massive marl	100	95.0206	0.738341
C_2	(F_3) Laminated sandstones, siltstones and marls	84.37476	70.48016	9.333921
C_3	(F_2) Thin sandstones	72.20415	59.63765	18.33316
C_4	(F_1) Massive sandstones	39.65112	24.06714	22.17427

Table 2. Description of HAC-clusters by average of petrophysical parameters

Cluster	Facies	Water saturation	Shale volume	Porosity
C_1	(F_4) Massive marl	100	99.2144	0.37625
C_2	(F_3) Laminated sandstones, siltstones and marls	99.52591	90.92864	2.84875
C_3	(F_2) Thin sandstones	70.38612	58.78594	15.92913
C_4	(F_1) Massive sandstones	45.50047	22.78151	21.182

Table 3. Classification matrix of SOM- clusters

Corrects(%)	Maximum	C_1	C_2	C_3	C_4
C_1	96	96	4	0	0
C_2	53	29	53	14	4
C_3	100	0	0	100	0
C_4	100	0	0	0	100
Total	85				

Table 4. Classification matrix of HAC-clusters

Corrects (%)	Maximum	C_1	C_2	C_3	C_4
C_1	100	100	0	0	0
C_2	75	25	75	0	0
C_3	87	4	0	87	9
C_4	95	0	0	5	95
Total	89				

Remark (iii):

It is important that from the classification matrices alone the goodness of methods cannot be concluded, since HAC arranges the groups into a hierarchical system contrary to SOM. However SOM method also has advantage, it can handle the categorical variables jointly the continuous variables contrary to HAC.

The comparison of results maps (Figures 5-6) reveals that the HAC method extends the clusters C_3 and C_4 more than SOM, especially in the northern part of the field. Thereby, this map on Figure 6 shows a more continued geometry of channels. However, it is worthy to consider the classification matrices of discriminant analyses (Tables 3-4). In the case of the highlighted clusters, according to the classification matrices the cluster C_3 (thin sandstones) and its central part, i.e. C_4 (massive sandstones) are higher (100%) in case of SOM however the probability of cluster C_1 is higher with small difference in the HAC model. In both methods according to the classification matrix, we can detect that the estimation of the cluster C_2 is relatively low than the others. The differences of the physical parameters (Tables 1- and 2) are negligible between cluster C_1 (massive marls) and C_2 (laminated sandstones, siltstones and marls), especially in the case of HAC model because of the distance of these clusters in the hierarchical structure. The dendrogram (Figure 7) can display this strict contact between these two clusters, as well. Since C_2 and C_1 belong to a higher level group (group I. in Figure 7), what means common genetic derivation because we cannot exclude the effects of similar accumulation in the clustering processes. The cluster C_2 is realized as a transitional zone between C_3 and C_4 clusters.

Resulting maps (Figures 5 and 6) demonstrates the main sediment transport directions (NW-SE), outlines the main channel shape and four lithofacies geometry. We can define the four patterns as sub-environments, e.g.

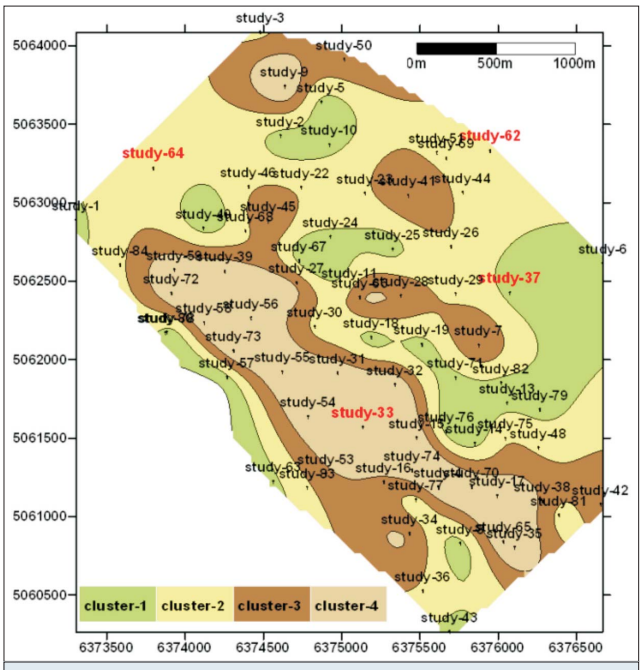

Fig. 5. Result map of SOM clustering
Sl. 5. Karta SOM združivanja

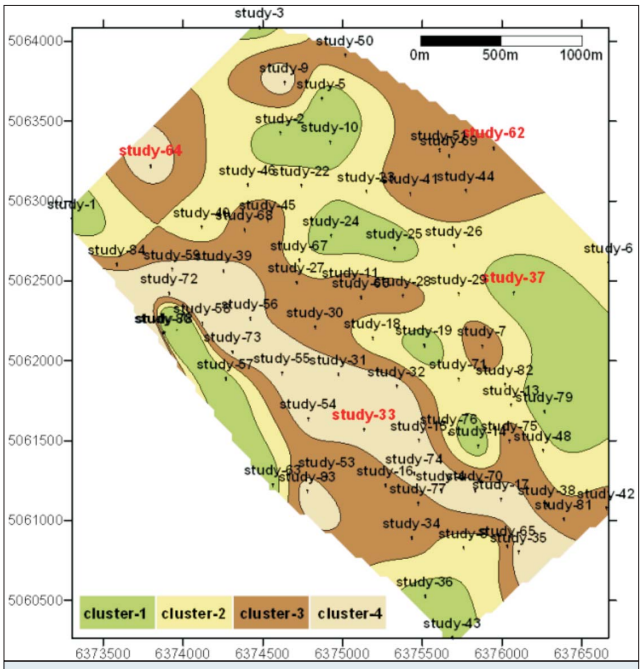

Fig. 6. Result map of HAC clustering
Sl. 6. Karta HAC združivanja

cluster C_3 assigns the geometry of turbidity channel margins (linked to higher porosity) and C_4 means central channel part (highest porosity). Transitional zone also includes cluster C_2, which defines overbank deposits along the channels. The cluster C_1 represents very low porosity, i.e. marl deposited as normal basin pelitic sediment.

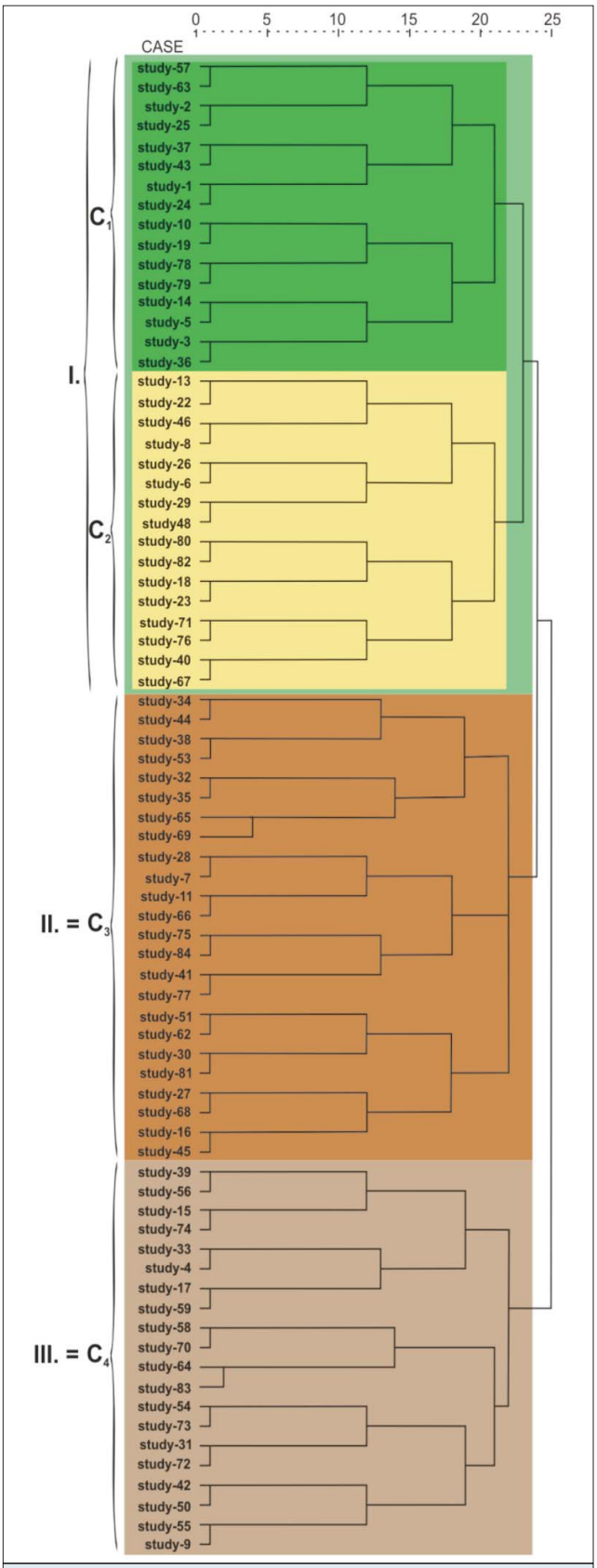

Fig. 7. Dendrogram of HAC method
Sl. 7. Dendogram HAC metode

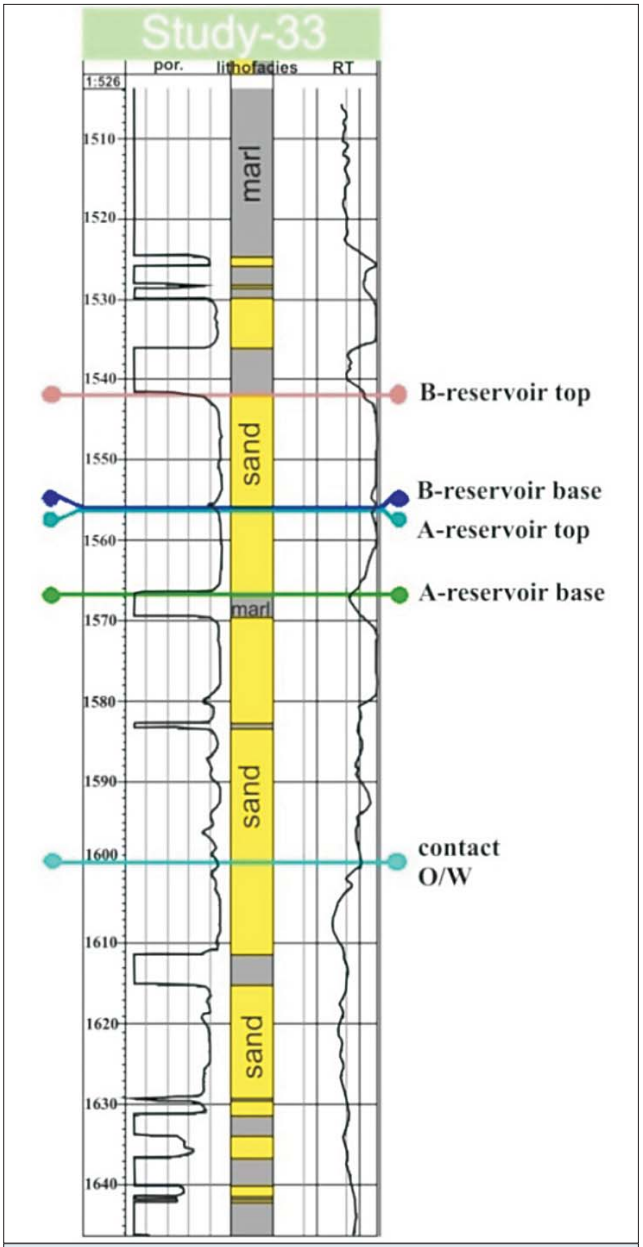

Fig. 8. Resistivity and porosity well-log with lithofacies; it belongs to the C_4 in both of method
Sl. 8. Krivulje otpora i poroznosti na karotažnom dijagramu s naznačenim litofacijesom; sekcija je svrstana u klaster C_4 s objema metodama

In fact, those clusters very well correspond to lithofacies which are defined as typical for Late Miocene sedimentation in the Sava Depression.[20] Previously defined facies of thick-layered massive sandstones (F_1) corresponds to cluster C_4. Facies of thin sandstone layers (F_2) corresponds to C_3, facies of laminated sandstones, siltstones and marls (F_3) to cluster C_2. Eventually, facies of massive marls (F_4) is cluster C_1. The results are in accordance with the published models of the depositional history in the Sava Depression.[15,19]

The result maps (Figures 5 and 6) visualize similar geometry of clusters however there are also some differ-

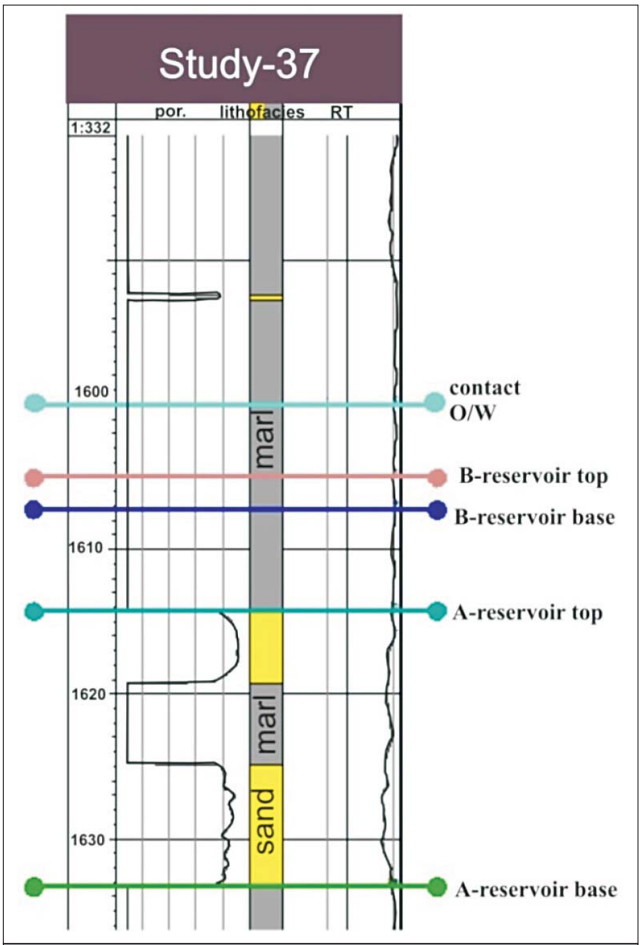

Fig. 9. Resistivity and porosity well-log with lithofacies; it belongs to the C_1 in both of method

Sl. 9. Krivulje otpora i poroznosti na karotažnom dijagramu s naznačenim litofacijesom;sekcija je svrstana u klaster C_1 s objema metodama

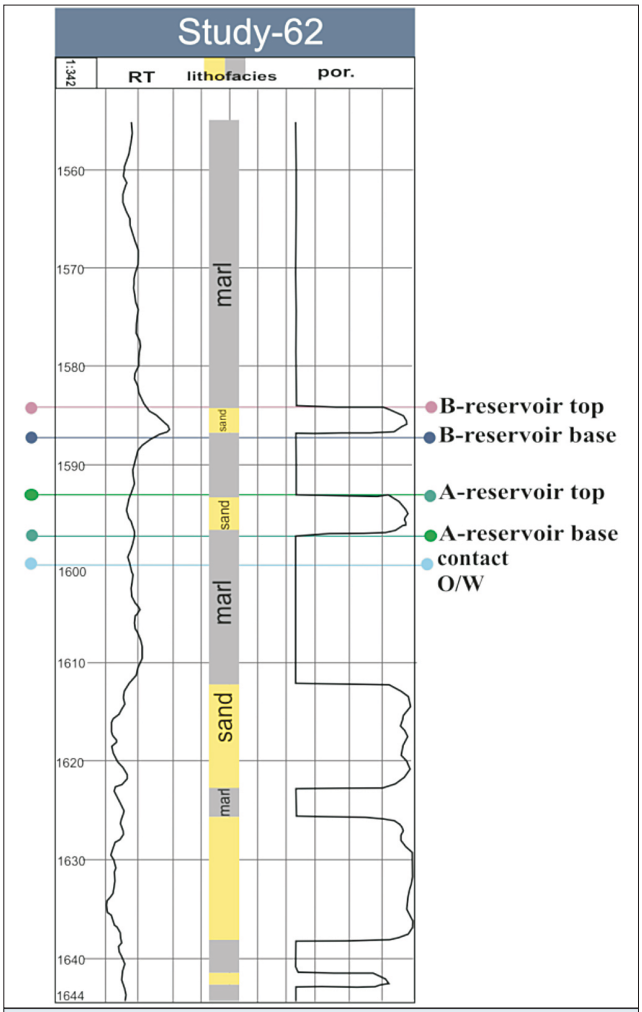

Fig. 10. Resistivity and porosity well- log with lithofacies; it belongs to the C_2 in the SOM method and C_3 in the HAC method

Sl. 10. Krivulje otpora i poroznosti na karotažnom dijagramu s naznačenim litofacijesom; C_2 po SOM metodi i C_3 po HAC metodi

ences, especially in the northern part. Four representative well-log sections (porosity and resistivity logs and lithology column) are also given shown on Figures 8-11. In several cases (e.g. Figures 8 and 9) we can detect that these sections reflect well the defined clusters (Figures 5 and 6). In case of wells no. 62 and 64 (north part of the study area where difference is realized) the clusters by SOM model can reflect accurately the original average from wells.

3.2. IK MAPPING OF CLUSTERS DERIVED FROM HAC AND SOM

Indicator transformation of the clusters was based on equation (5). There were 76 input data divided into 4 classes (Figures 12 and 13)

For each class corresponding variogram was calculated using Variowin 2.21.[12] According to references[2,11], there are some criteria which should be satisfied for variogram models applied in Indicator Kriging (IK) technique:

- The theoretical (model) function must be the same (spherical is used here in both cases);
- The sill must be identical (standardized variogram);
- The nugget effect must be the same (it is 0.6 for HAC and 0 for SOM method);
- Only the range can change for different cut offs.

Theoretical variograms used in IK mapping are shown in Figures 14 and 15. Direction 135° is set for the primary axis. Range of subordinate axis was calculated based on structural map of the reservoir, whose relation was applied in variograms (i.e. according to structural map primary axis is 1.5 times longer than subordinate, so the variogram range for primary axis should also be 1.5 times longer than the one for subordinate).

IK mapping was performed in WinGslib™ package. A 'Full IK" approach was used by 'Simple Kriging' estimation. In 'Grid definition' the number of cells in X direction was 338 and in Y 383. The final model contains 129 454 cells. Results are probability maps for each cluster (Figures 16 and 17).

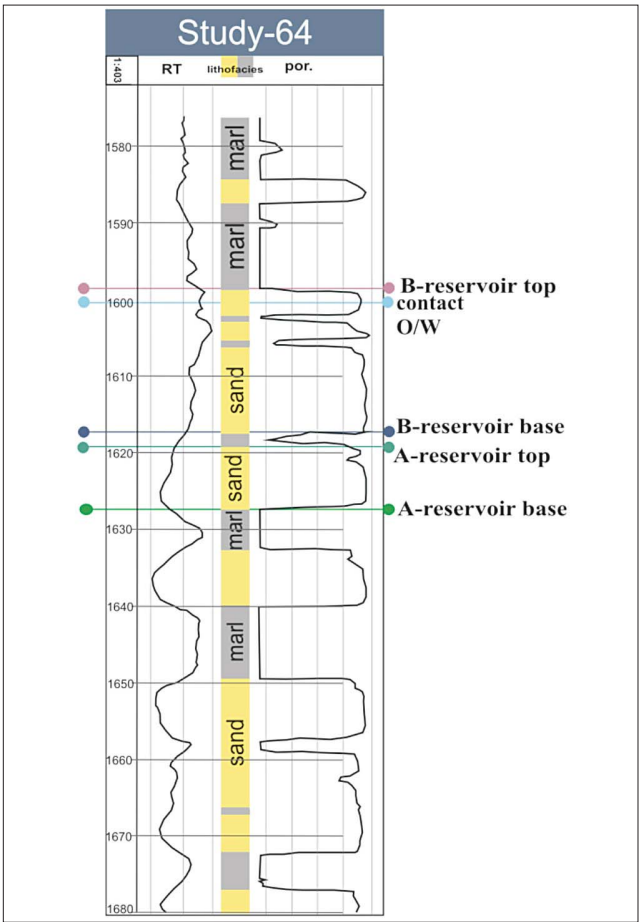

Fig. 11. Resistivity and porosity well- log with lithofacies; it belongs to the C_2 in the SOM method and C_4 in the HAC method
Sl. 11. Krivulje otpora i poroznosti na karotažnom dijagramu s naznačenim litofacijesom; pripada u C_2 po SOM metodi i C_4 po HAC metodi

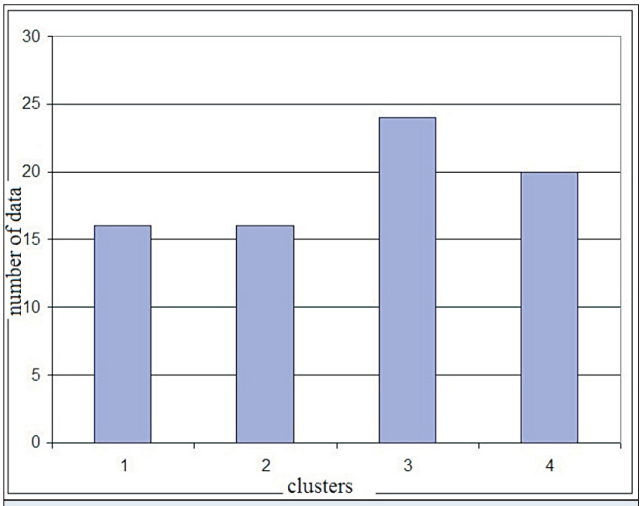

Fig. 12. Histogram of input data based on HAC method
Sl. 12. Histogram ulaznih podataka iz rezultata metode HAC-a

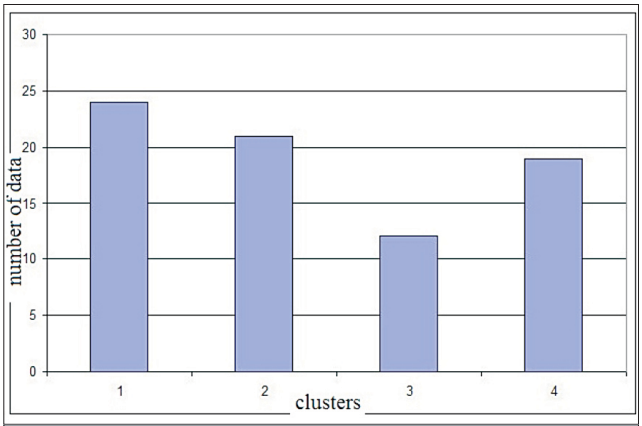

Fig. 13. Histogram of input data based on SOM method
Sl. 13. Histogram ulaznih podataka iz rezultata metode SOM-a

The Figures 16 and 17 show the probabilities of each cluster (cluster 1-4) distribution where the maximum values of probabilities with red colour is 1 and lowest values with blue colour is 0. In the comparison of Figures 16 and 17 clusters derived from HAC method on the maps do not show higher probabilities except in case of cluster 3. Each cluster distribution (from HAC) can be described through the lower probabilities than in case of clusters which derived from SOM method. However in both of cases the cluster 4 has the low probabilities on IK maps, probably due to low number of measurements in basin marls.

According to the equation (6) the probabilities maps are merged. It means that we filtered which clus-

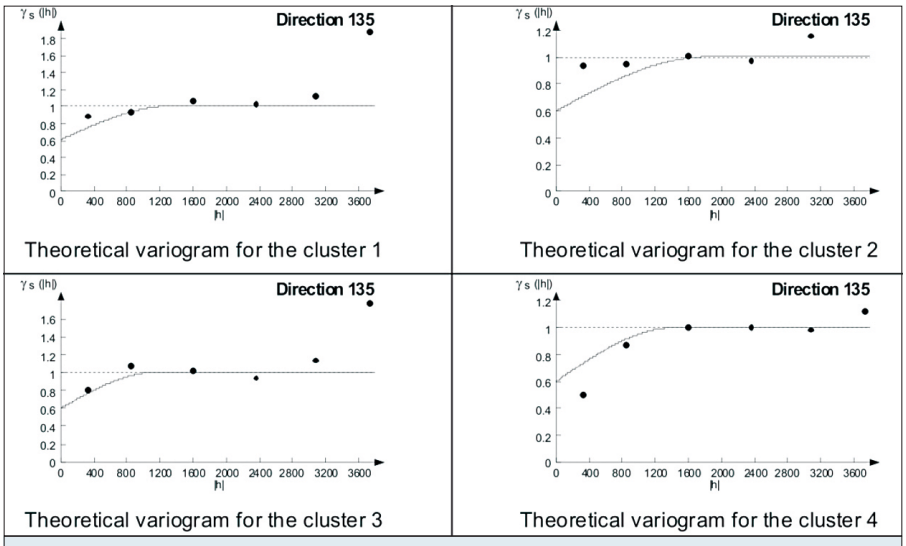

Fig. 14. Experimental points and theoretical variograms for different clusters (from HAC method)
Sl. 14. Eksperimentalne točke i teorijski variogrami za različite klastere (iz HAC metode)

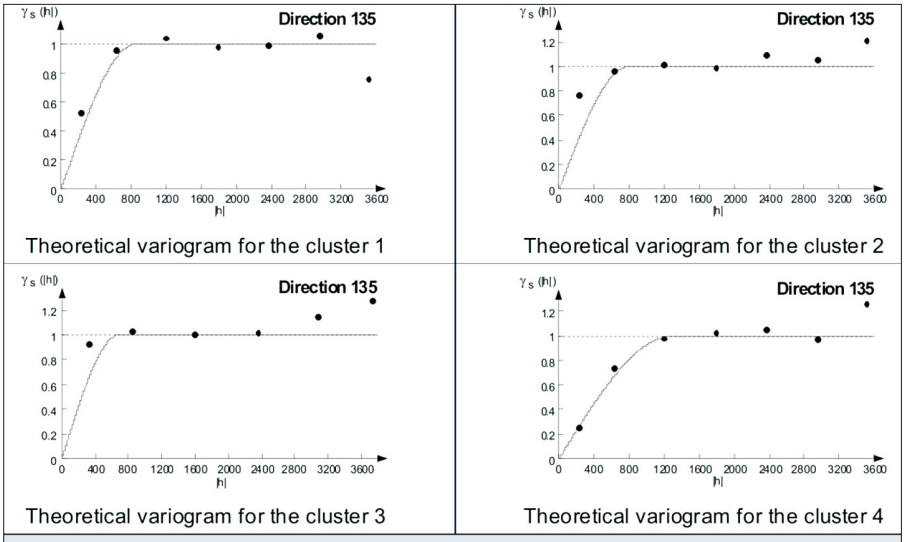

Fig. 15. Experimental points and theoretical variograms for different clusters (from SOM method)
Sl. 15. Eksperimentalne točke i teorijski variogrami za različite klastere (iz SOM metode)

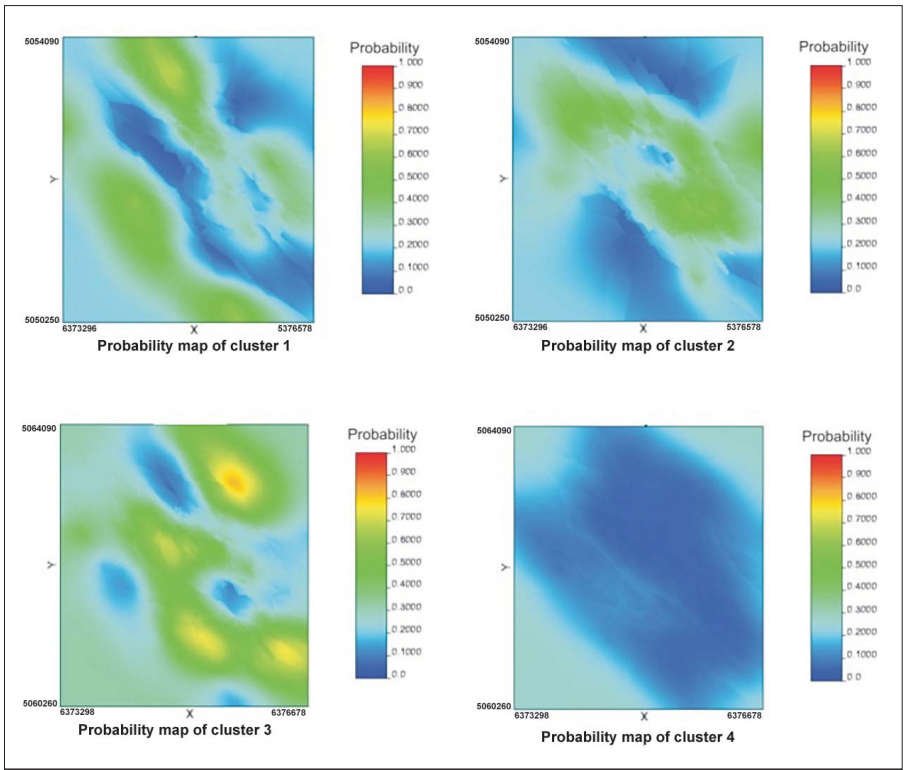

Fig. 16. Clusters probability maps (clusters from HAC method)
Sl. 16. Karte vjerojatnosti klastera (vrijednosti su iz klastera izračunatih metodom HAC-a)

ter can appear with the highest probabilities in each grid dots. These merging resulted in maps in Figures 18 and 19 where the geometries of points (e.g. channel geometries) with similar properties have the highest probabilities are outlined polygons. In this manner the maps show a bit discontinued geometry because of the cluster-extension based on probabilities (the cluster 4 areas, i.e. marls, are not shown on those maps).

4. CONCLUSION

According to the pattern of the mapped clusters we can recognize a turbidity system and its direction that indicates the deposition with direction NW-SE. In both of presented clustering methods the patterns of the clusters in this turbidity system show the deposits of channels, over-bank and basin plain. Usually the sand content can outline the depositional geometry, and in this study we demonstrated that it is also possible to apply clustering methods which are based on three petrophysical parameters jointly (porosity, saturation, shale volume).

Both of models carry uncertainty, what can be seen in the result of the discriminate analyses. However, both of methods are able to recognize the mentioned clusters of the sub-environments where the degree of homogeneity is characterized by properties of clusters. The SOM method can reflect small scale variability, e.g., the thin marl layers in reservoir series. It is important because these layers also can have effects on the average values of used petrophysical parameters. Clustering with HAC method can not reflect such variability.

Presumably this contrast comes from the reason that the character variable for lithology was also built in the SOM model, and on the other hand the SOM neural network can recognize the non-linear contact in the data set. It was explained by the average of the parameters and logs of highlighted wells in the result maps (well no. 62 and 64 on the north part of the field). These averages are closer to the average of clusters in SOM model than in HAC model. The HAC method also has ability to follow the close contact between two clusters on dendrogram, as it is relationship of clusters C_1 and C_2.

5. ACKNOWLEDGEMENTS

The presented research is done through activity of the Sector for Geology and Reservoir Management, INA Plc., Croatia, and Department of Geology and Paleontology, University of Szeged, Hungary. Also, many thanks to mentors, colleagues for their suggestions and comments and to reviewer for the valuable piece of advice.

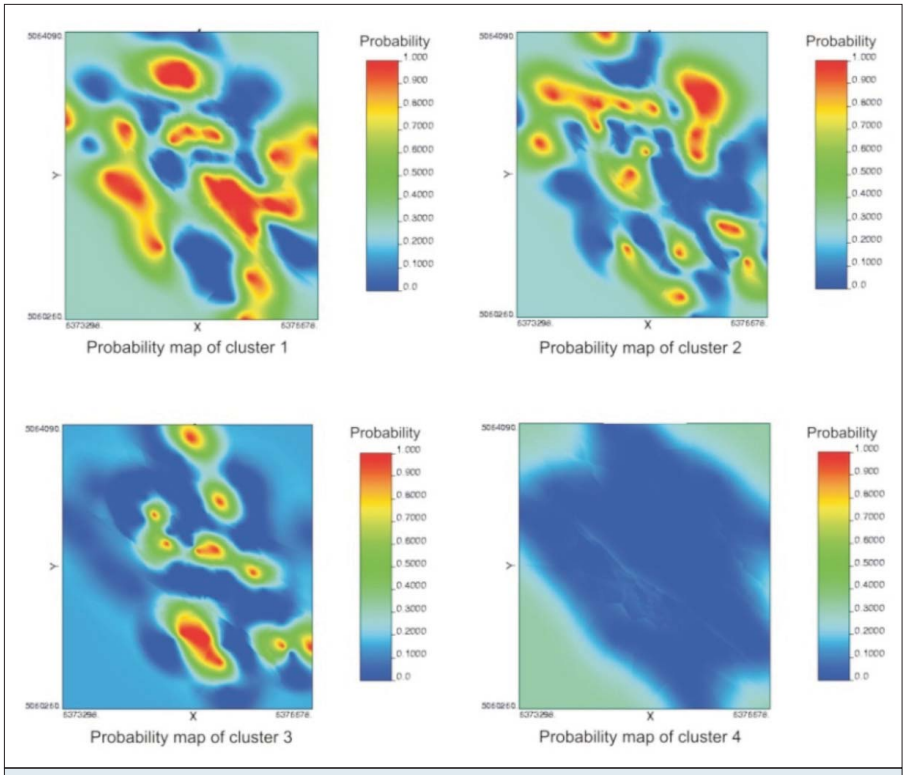

Fig. 17. Clusters probability maps (clusters from SOM method)
Sl. 17. Karte vjerojatnosti klastera (vrijednosti su iz klasteri izračunatih metodom SOM-a)

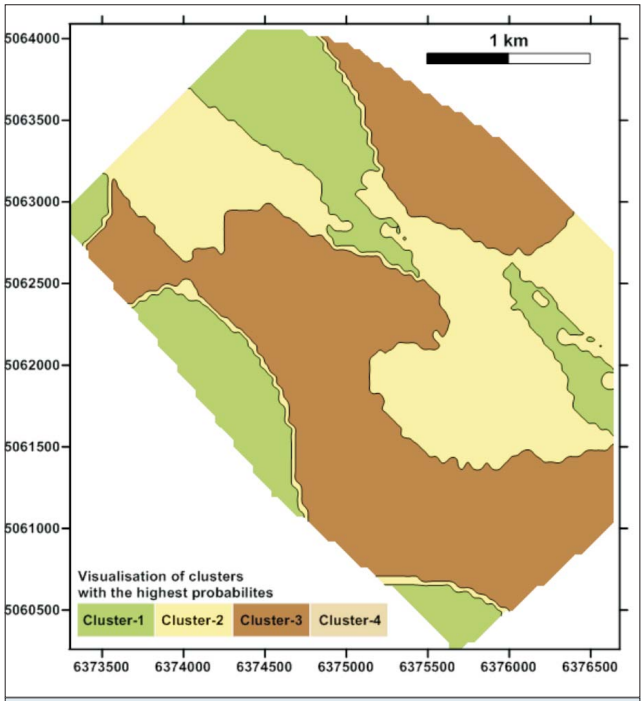

Fig. 18. Visualisation of the points with similar properties and the highest probabilities (cluster from HAC method)
Sl. 18. Vizualizacija točaka sa sličnim karakteristikama i najvećim vjerojatnostima(klasteri po HAC metodi)

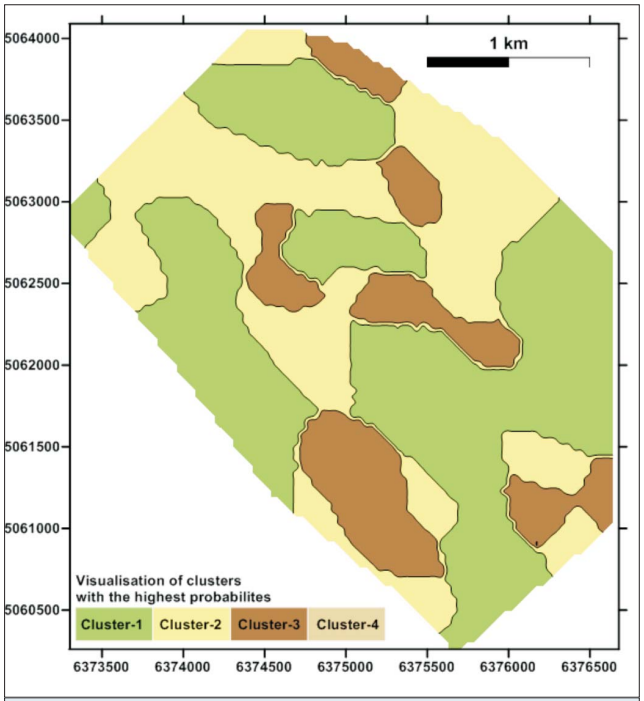

Fig.19. Visualisation of the points with similar properties and the highest probabilities (cluster from SOM method)
Sl. 19. Vizualizacija točaka sa sličnim karakteristikama i najvećim vjerojatnostima(klasteri po SOM metodi)

6. REFERENCES

1. Deutsch, C.V. and Journel, A.G. (1997): GSLIB: Geostatistical Software Library and User's Guide, 2nd edn. Oxford University Press, New York, 369 p.
2. Hardy, R. L. (1990): Theory and Applications of the Multiquadic-Bi-Harmonic Method. Computers Math. Applic, 19, 8/9, 163-208.
3. Horváth, J. (2011-in press): Defining Depositional Environments by Using Neural Networks, Geologia Croatica.
4. Kohonen, T. (1982) Self-organized formation of topologically correct feature maps. Biological Cybernetics, 43, 59-69.
5. Kohonen, T. (2001): Self Organized Maps. Springer, Berlin 511p.
6. Lampinen, T., Laurikkala, M., Koivisto, H. and Honkanen, T. (2005): Profiling Network Applications with Fuzzy C-means and Self-Organizing Maps. Studies in Computational Intelligence "Classification and Clustering for Knowledge Discovery" 4, Springer, Berlin-Heidelberg.
7. Lučić, D., Saftić, B., Krizmanić, K., Prelogović, E., Britvić, V., Mesić, I. and Tadej, J. (2001): The Neogene evolution and hydrocarbon potential of the Pannonian Basin in Croatia. Marine and Petroleum Geology, 18, 133-147.
8. Magyar, I., Geary, D.H., Sütó-Szentai, M., Lantos, M. and Müller, P. (1999): Integrated biostratigraphic, magnetostratigraphic and chronostratigraphic correlations of the Late Miocene Lake Pannon deposits. Acta Geologica Hungarica, 42, 5-31.
9. Malvić, T., Velić, J. and Peh, Z. (2005): Qualitative-Quantitative Analyses of the Influence of Depth and Lithological Composition on Lower Pontian Sandstone Porosity in the Central Part of Bjelovar Sag (Croatia). Geologia Croatica, 58, 1, 73-85.
10. Manning, C.D., Raghavan, P. and Schütze, H. (2008): Introduction to Information Retrieval. Cambridge University Press. 482 p.
11. Novak Zelenika, K., Malvić, T. and Geiger, J. (2010): Mapping of the Late Miocene sandstone facies using Indicator Kriging. Nafta 61, 5, 225-233.
12. Pannatier, Y. (1996): Software for Spatial Data Analysis in 2D. Springer-Verlag, New York, 91 p.
13. Rögl, F. (1996): Stratigraphic correlation of the Paratethys Oligocene and Miocene. Mitt. Ges. Geol. Bergbausstud. Österr., 41, 65-73.
14. Satftić B., Velić J., Sztanó O., Juhász G.Y. and Ivković Z. (2003): Tertiary subsurface facies, source rocks and hydrocarbon reservoirs in the SW part of the Pannonian basin (Northern Croatia and South-West Hungary). Geologia Croatica, 56, 1, 101-122.
15. Šimon, J. (1980): Prilog stratigrafiji u taložnom sustavu pješčanih rezervoara Sava-grupe naslaga mlađeg tercijara u Panonskom bazenu sjeverne Hrvatske [Contribution in stratigraphy and depositional system of the sandstone reservoirs of the Sava group of deposits of Younger Tertiary in the Pannonian basin of the northern Croatia]. PhD dissertation, University of Zagreb, Faculty of Mining, Geology and Petroleum Engineering, 66 p. (in Croatian)
16. Tobler W. (1970): A computer movie simulating urban growth in the Detroit region. Economic Geography, 46, 2, 234-240.
17. Tryon, R. C. (1939): Cluster analysis. Edwards Brothers, Ann Arbor, Michigan 347p.
18. Ultsch, A., Korus, D. and Wehrmann, A. (1995): Neuronal Networks and their Rules for Classification in Marine Geology. In: In Raum und Zeit in Umweltinformationssystemen, 9th International Symposium for Environmental Protection CSEP'95, Vol. I, ed. GI-Fachausschuß 4.6 "Informatik im Umweltschutz" vol. 7, Metropolis-Verlag, Marburg, 676-693.
19. Vrbanac, B. (1996): Paleostrukturne i sedimentološke analize gornjopanonskih naslaga formacije Ivanić Grad u Savskoj depresiji [Palaeostructural and sedimentological analyses of Late Pannonian sediments of Ivanić Grad formation in the Sava Depression]. PhD dissertation, Faculty of Natural Sciences, University of Zagreb, 121 p. (in Croatian)
20. Vrbanac, B., Velić, J. and Malvić, T. (2010): Sedimentation of deep-water turbidites in the SW part of the Pannonian Basin. Geologica Carpathica, 61, 1, 55-69.

❖

Authors:

Janina Horváth, University of Szeged, Department of Geology and Paleontology, Szeged, Hungary, e-mail: th.janina@geo.u-szeged.hu

Kristina Novak Zelenika, INA-Naftaplin, Sector for Geology and Reservoir Management, Zagreb, Croatia, e-mail: kristina.novakzelenika@ina.hr