Defining depositional environments by using neural networks

Janina Horváth

Department of Geology and Palaeontology, University of Szeged, H-6722, Szeged, Egyetem u. 2–6, Hungary; (th.janina@geo.u-szeged.hu)

Geologia Croatica

ABSTRACT

Traditional techniques to identify a depositional body from core data are costly and sometimes difficult to extrapolate to uncored wells. The application of Kohonen's Self Organized Map (SOM) approach may be useful for the interpretation of a depositional rock body through well-log data. SOM is based on a clustering algorithm and this method can be used to discover spatial patterns occurring as clusters in unstructured data sets. An example of the application of SOM is presented whereby clusters through SOM can indicate the contours of well-known depositional patterns such as sub-environments.

Keywords: clustering method, depositional environment, neural network, pattern recognition

1. INTRODUCTION

The neural network approach is a well-known development tool, which became popular within the last couple of decades. Supervised and unsupervised trainable networks are used in many different fields of geology, especially petroleum geology or facies analysis e.g. the unsupervised network as a tool for lithofacies identification (CHANG et al., 2002), application of a supervised neural network for predicting permeability from porosity (ROGERS et al., 1995), using a supervised neural network tool in reservoir characterization (AHMED et al., 1997). However, supervised networks are more frequently used, because the unsupervised learning networks can solve specific problems like clustering or, pattern recognitions problems as a way of defining spatial patterns. Application of a particular unsupervised neural network called Kohonen's network (in other words Self Organized Map, abbreviated SOM) in the identification spatial pattern of some delta-plain sub-environments is demonstrated here. Kohonen's neural network has been successfully applied to studies with different representation methods or as a tool to define clusters e.g. a well log interpretation model for the determination of reservoir facies and fluid contents (AKINYOKUN et al., 2009), lithofacies identification (CHANG et al., 2002), and classification of biogenic sedimentation (ULTSCH at al., 1995).

Traditional techniques to identify depositional environments from core data are difficult as they require a large dataset and/or are less precise than some results gained by using mathematical models for mapping. This study shows a neural network methodology which can handle simultaneously more point data and even join more attributes, or property of data points. Therefore the SOM approach may be regarded as a possible method for pattern recognition by parallel analysis of multiple data sources. In this study we describe a method based mostly on data of electrofacies obtained from e-logs. It is a possible tool for the characterization of depositional environments comprising the rock body. Contrary to traditional geostatistical interpolation methods, neural networks as clustering schemes can provide many possibilities for modeling. Kohonen's algorithm reveals these spatial patterns in the terms of clusters. From this aspect, SOM is quite similar to general clustering algorithms, although ULTSCH (1995) has shown that SOM can even recognize point-clusters in those situations where traditional cluster techniques fail to find any reasonable groups. Here, spatial patterns are defined as spatial groups of points, where points are bounded by a polygon (e.g. contour of sand content). By using this definition, one can conclude, that pattern recognition can be drawn back to a clustering problem. The shape of the polygon, i.e. the contour geometry, reflects the

2D or 3D geometry of the depositional facies identified. This is the sedimentary environment according to PETTIJOHN and POTTER (1972).

In this study, the main emphasis is given to the following issues: (1) Demonstration of how a known depositional geometry outlined by sand-content contours can be honored by using SOM analysis of points of three petrophysical grids. This analysis relies on good average porosity, net pore volume and hydraulic conductivity grids; (2) Description of an example of the application of this robust approach in re-recognizing the known distributary mouth bar shape in a particular Pannonian reservoir.

2. APPLIED METHOD

2.1. Self organized maps (SOM)

Kohonen's neural network (or Self Organized Map (SOM)) means that the neurons are organized in a grid (Fig. 1) as a 'map'. However it is not a real map since it does not assign any spatial coordinates to the samples. So, self organizing gives a 'map' where the nearby locations represent inputs with similar properties.

SOM is a type of competitive and unsupervised network. The competitive and unsupervised learning algorithm implies that the network has to have the ability to recognize the structure of a multidimensional basic data set, by the method of dimension reduction (KOHONEN, 1982, KOHONEN 2001; HAYKIN, 1999; PATTERSON, 1996; FAUSETT, 1994). This reduction is only a "queasy" one, since each neuron is an n-dimensional weight-vector, where n is equal to the dimension of the input vectors (Fig. 1). The main goal of SOM is to represent data points with fewer representatives preserving the original topology (JOHANNSON, 2003).

Even the unsupervised learning character implies that only one data set is available for analysis. So the data structure is explored within the input set without using any reference dataset. In the case of the supervised neural network, the training processes depend on training examples and each example is a pair of an input objects. In contrast with the unsuper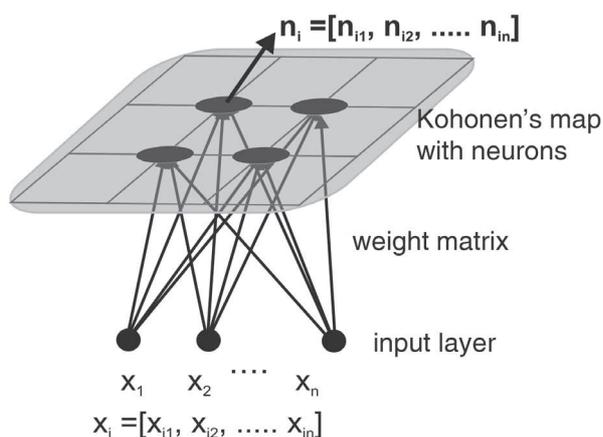vised network, the learning processes are based on un-labelled examples. The framework (Fig. 2) of Kohonen's Neural Network also demonstrates this difference between the supervised and unsupervised neural networks. There are no real output layers, but neurons are arranged into a grid of dots. In other words, the grid is the output" layer. This is the so-called Kohonen's layer. It contains m neurons; each neuron includes one output cluster. Figure 1 shows that in this layer neurons give a two dimensional map (usually we can use one or two dimensional Kohonen' layer), however the neurons saved and represent the original data dimension because each neuron is represented as an n dimensional vector, where n is the dimension of the input, training, vectors.

Weighted x_i inputs are processed by all three neurons (Fig. 2). Each neuron computes its weighted input, but only the neuron with the largest excitation 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 units are reserved by this active element through the lateral connections during training shown in Fig. 2.

An advantage of the SOM network is that it attempts to learn the structure of the original data. It means that the training data set gives the contact between the original data points and identified classes. If new data, unlike previous cases, is encountered, the network fails to recognize it which indicates novelty. Therefore, in the next learning step, the grid is updated based on the novelty. The updating process is quite similar to that of the K-means algorithm, where the arrangements of cluster-centroids change in the regular low-dimensional grid. However, in case of SOM, changing the position of neurons in the data space influences the positions of its neighbors through the lateral connections within a predetermined distance.

Since SOM consists of neurons organized on a regular low-dimensional grid, these neurons control the classification of data using a suitable distance. The input vectors are compared step by step with the reference vectors using a neighboring function and the grid is revised continuously using a winning neuron algorithm. During the comparison
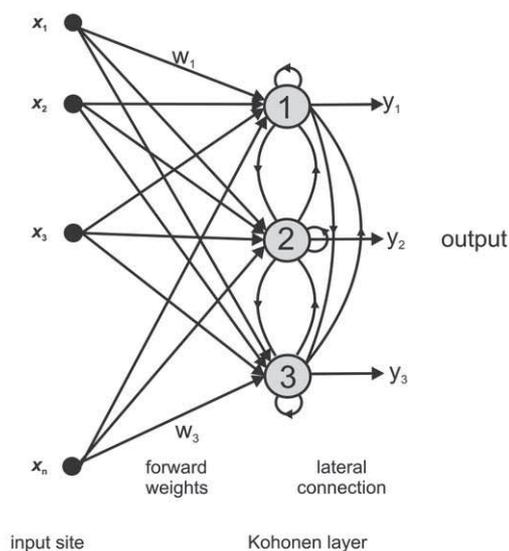

Figure 1: Kohonen's layer with n-dimensional neurons.

Figure 2: Network of three competing units.

process, each sample input vector is randomly drawn from the input data set, and their similarities to the 'codebook' vectors are computed. This process is based on Euclidean distance that measures distance from the data sample to each neuron to obtain the winning neuron (LAMPINEN et al., 2005).owever better selections use either Mahalanobis distance, (developed by MAHALANOBIS in 1936), or cosine distance as alternative measures. The latter is derived from correlations between variables in the property space.

Kohonen's network learns iteratively and uses the modified Hebb-rule as the winning neuron algorithm (LAMP-INEN et al, 2005).

The modified law is w(t+1)=w(t)+ρ(t)×(x–w(t)). Where x is the training case, $w(t)$ is weight, and $\rho(t)$ is the learning rate.

The application of neural networks is relatively simple. An artificial network studies the data structure, which means that it necessarily must have some preliminary empirical knowledge e.g. on appropriate size of network, rate of learning, number of learning steps. In this case, the size of the network corresponds to the number of neurons representing the clusters of which numbers must be given previously. This setting depends on preliminary expertise.

2.2.1. Pattern recognition as a clustering method

SOM is a type of clustering method. It differs from any traditional clustering method in its robustness. However it needs, like any 'traditional k-mean algorithm', to input the number of groups to be revealed.

SOM, (as any clustering method), provides an opportunity for deriving solutions for geological problems which can be related to different clustering problems, such as pattern recognition (ULTSCH, 1995). For implementation, Kohonen's network has been used as it is possible to recognize spatial patterns being defined as spatial groups of points bounded by a polygon. It has potential for applying pattern recognition to problems such as identifying the spatial pattern of a sub-environment on a delta plain, which is resoluble by applying the SOM method based on a clustering algorithm.

Shapes shall be construed as clusters of data points that are bounded by a polygon which implies that pattern recognition is a clustering problem. However, this approach raises the question of the definition of patterns as statistical groups. Within a pattern, there are points with the same or similar properties. In practice, this 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 sub-environments, where the properties change between separate clusters. Figure 3 demonstrates an example for mapping clusters using the original coordinates. SOM creates c_m clusters (m is the number of clusters) based on the property values of data points without original spatial coordinates. After the clustering process we map the resulting clusters as patterns using the original coordinates. In this manner the different properties clusters are mapped and the pattern is visualized on the map. Thus after the clustering process, the resulting clusters are patterns in a map using the original coordinates.
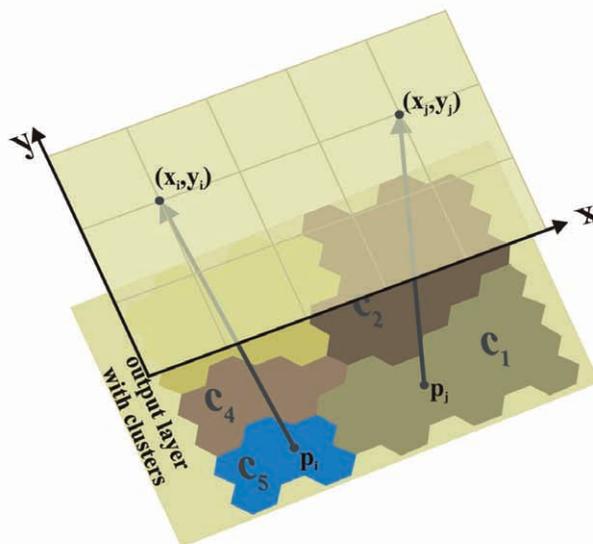

Figure 3: Arrangement points on the map. Where p_i means data points within clusters (c_m); and (x_i, y_i) means the positions of data points with their original coordinates in the map.

2.2.2. Recognisable patterns in the map

The mapping processes compose the closed patterns by the arrangement of the points in space through the original (x_i, y_i) coordinates. In this case it is necessary to give some conditions and definitions for recognizing patterns for mathematical reasons.

(1) A Closed pattern means that there is at least one inner point of the group.

(2) A point is an inner point if all of its neighboring points belong to the same cluster, i.e. there are at least four edge-neighbours to it.

(3) The border of a shape is defined by a point which has neighbours from another cluster. Of course a pattern might also have more neighbours.

(4) Within a shape we can also define another shape if it has inner points. Otherwise we should disregard these points as a pattern. However in this case, the question is how many inner points are enough to define a pattern. The answer depends on the grid dissolution. Since a 100 x 100 m grid size was applied over the study area, we can define the inner pattern which has at least one inner point. (This consideration gives 5 inner points in case of 50 x 50 m grid size and 25 inner points in case of 25 x 25 m grid size and so on.)

(5) The 'inner shape', should be covered by border points of only one other cluster.

3. STUDY AREA

Data originated from the Szőreg-1 reservoir (Algyő Field) that was used as a test area for the application of pattern recognition of palaeoenvironments in a clastic reservoir using the Kohonen's neural network. The Algyő Field is the largest Hungarian hydrocarbon accumulation consisting of several oil and gas bearing reservoirs (Fig. 4). The upper mem-
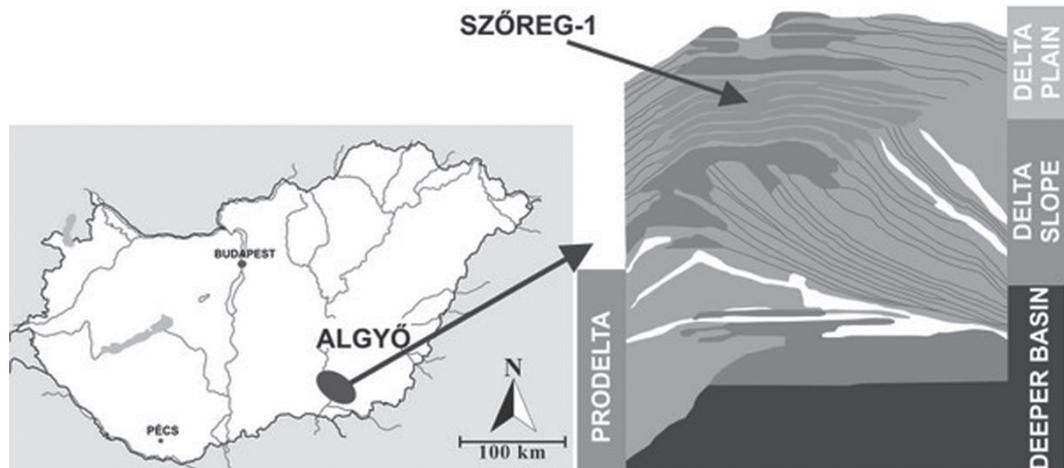
Figure 4: Macro-sedimentological model of the Pannonian series of Algyő (after BÉRCZI, 1988).

bers of these reservoirs developed in a Pannonian delta system as the consequences of complex lateral delta-shifting and prograding phases. They can be subdivided into delta slope and delta plain rock bodies. The sills of the individual reservoirs were formed during delta abandonment phases. Below this series, the lower reservoirs are regarded as turbidity rock bodies of partly pro-delta fans and partly deep basin origin (RÉVÉSZ, 1982).

The Szőreg-1 oil reservoir, with a large gas cap is one of the largest rock bodies within the delta plain record (Fig. 4). Its average gross thickness is about 35 m, but locally it can be up to a maximum of 50 m thick. Earlier works by RÉVÉSZ (1982), GEIGER et al., (1998) and GEIGER (2005) have proved its delta plain origin with a significant amount of fluvial channel sedimentation. Figure 5 shows the original well density, the dataset is from 506 well on Szőreg-1 study area.

4. DATA PRE-PROCESSING AND ANALYSIS BY SOM

The network depends on the quality and quantity of training data, so data pre-processing is very important. This is as follows: 1) definition of the data type (discrete or constant), 2) calculation of statistics: mean, standard deviation, and 3) selection of outlier data. No 1 is critical because the outlier data reduces the efficiency of the network.In the case of category data, it is necessary to encode or scale the data. In this study we used continuous variable data.

It is imperative to set the initial values for the training rate and the neighborhood radius. The Kohonen-learning rate is altered linearly from the first to the last training cycle. We specified 0.3 for the start value and 0.02 for the end value. The neighborhood radius designates the adjacent area centered on the winning unit; in this case the size was 1 and specified a 3X3 square. Normal randomization of weights was used for the training; the mean and variance are specified, and are used to draw the initial weight values.

The second important step is to define the training, test and validation sets for training of SOM. The original data-
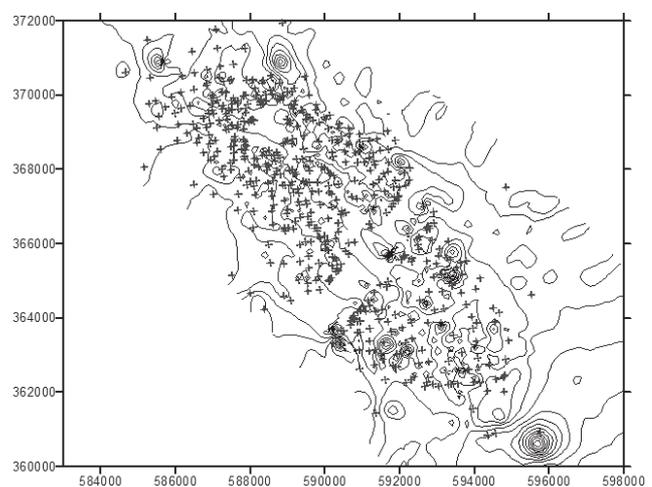

Figure 5: Well density of the study area (Szőreg-1 reservoir).

base comprises 7500 points (grid points). This dataset is deconstructed into three sub-sets (training-, test- and validation set) using a random number generator to avoid bias. 60% of this database is used for building a neural network by SOM as a training set, 20% as a test, and 20% as a validation set.

Training set: this set is important to learn to fit the parameters of the clusters (in this case clusters are the neurons that characterized the clusters). SOM is applied to find the optimal weights to update the Kohonen's layer and to position the neurons of clusters.

Validation set: this set is applied to tune the parameters of a classifier, and determine the stopping point of learning process.

Test set: this set lends itself to assess the performance of the trained clusters.

After pre-processing, the first step only sampled distributary mouth bars developed in Szőreg-1 to analyze how the SOM method can recognize the pattern for this part. The pattern of distributary mouth bars was given by sand contours that can reflect the geometry (Figs. 6a,b). It was an impor-
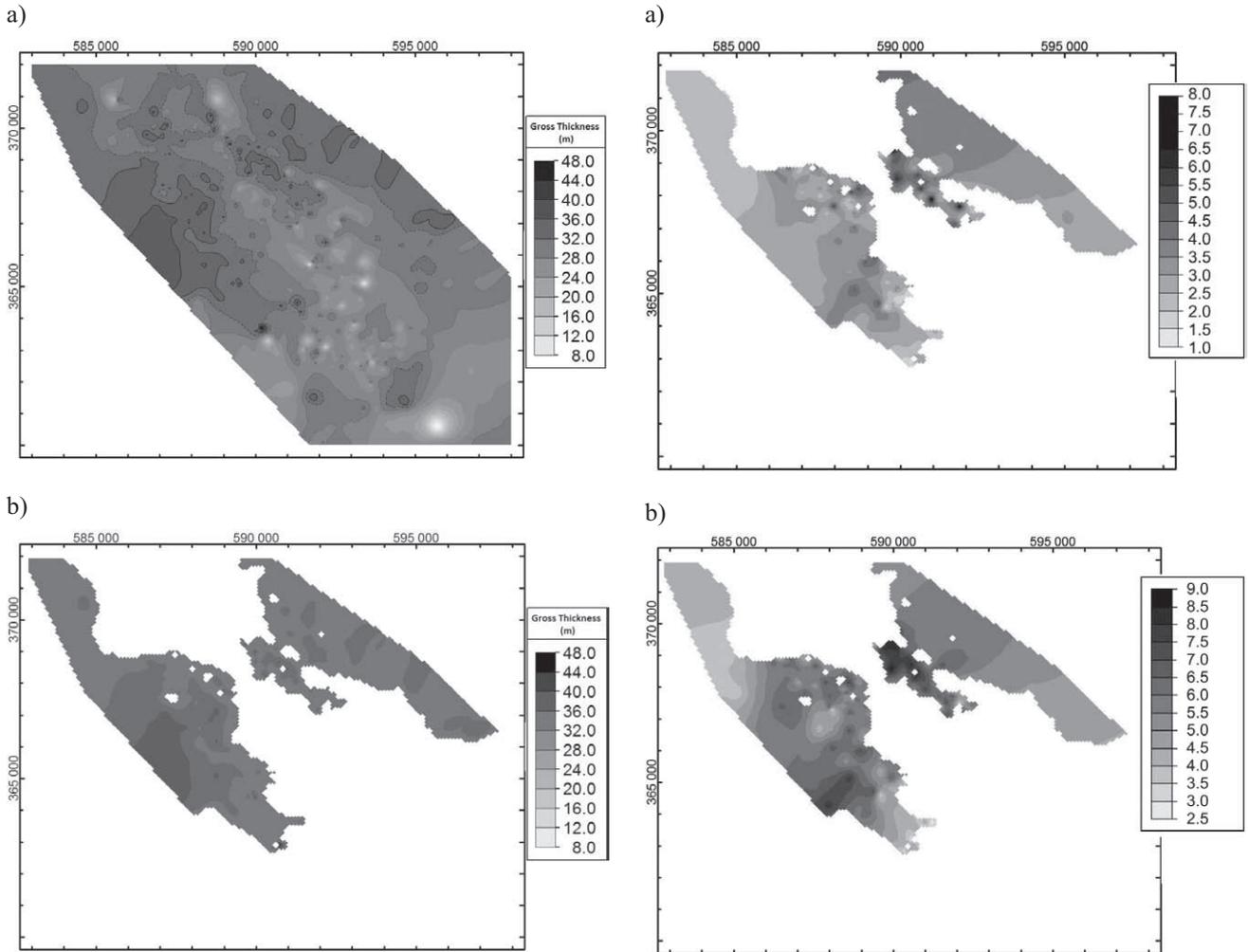
a)

b)

Figure 6: a) Gross thickness map of the Szőreg-1 reservoir (entire area), b) Gross thickness map of the area of the distributary mouth bars in the Szőreg-1 reservoir.

a)

b)

c)

Figure 7: a) Porosity map of distributary mouth bars in the Szőreg-1 reservoir, b) HFI map of distributary mouth bars in the Szőreg-1 reservoir, c) HK map of distributary mouth bars in the Szőreg-1 reservoir.

tant step since it can show how well the SOM-method reflects patterns jointly through three other variables. These variables were porosity, net pore volume and hydraulic conductivity data (Figs. 7a,b). The input data set comprises 2200 grid points (with 100X100 m dissolution), randomly divided into three (training, test and validation) subsets. The predefined number of clusters was tuned into 3, 4 and higher. In the case of more than 4 clusters, the algorithm became too confused for interpretation. Even, after mapping these results it did not constitute disjunctive clusters. In the first run, the minimum number of clusters is easier to tune in 3 because distributary mouth bars can be divided into three sub-environments as channel, outer bar, and bar crest.

After the clustering, we mapped the patterns. In the maps these patterns are the separated clusters of the analyzed area (distributary mouth bars) and this patterns mean the blocks of similar properties. Figure 8a,b represents the mapped result of the first run of SOM. The pattern within this depositional environment can affect the spatial arrangement of reservoir properties. Usually the sand contour used to reflect the spatial patterns but in the case of the clustering method and after that the mapping of clusters, similar patterns are
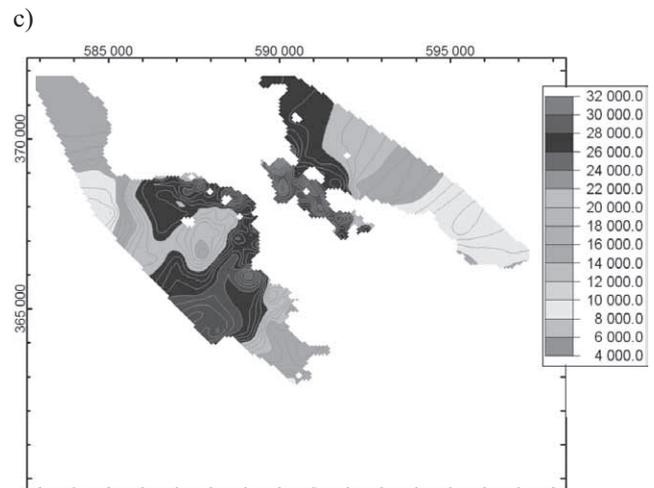
discernible like the maps of three original properties (Figs. 7a,c) or map of sand contour (Fig. 6b). In this case the result-maps can demonstrate that it is also possible to apply clustering methods for outline the depositional geometry which are based on other three petrophysical parameters jointly.
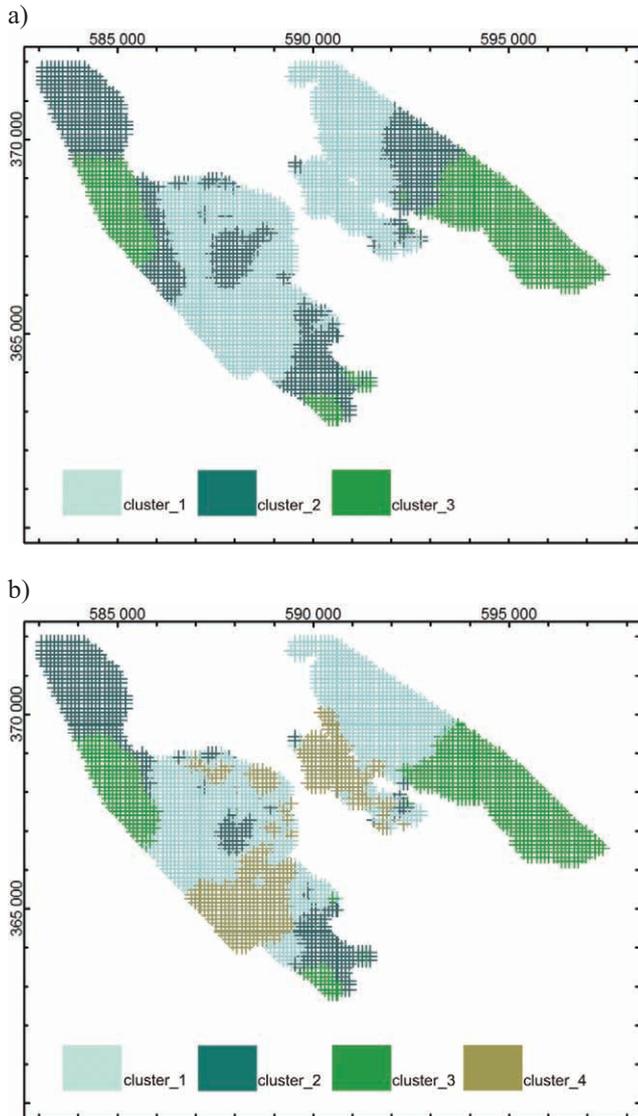
a)

b)

Figure 8: Mapped result for 3 clusters, where clusters contain points with similar properties in the present stage of diagenesis, b) Mapped result for 4 clusters, where clusters contain points with similar properties in the present stage of diagenesis.

a)

b)

Figure 9: a) Mapped result for 4 clusters, where clusters contain points with similar properties in the present stage of diagenesis, b) Mapped result for 6 clusters, where clusters contain points with similar properties in the present stage of diagenesis.

In the second run, we have involved each of the available 7500 grid points of the 'basic' grid system (with 100X100 m resolution, derived from well 506, Fig. 5). In addition to recognizing the patterns again, the goal was also to identify the contours of the distributary bars. In this case, the number of clusters was tuned in 4 and 6, because we assumed at least one, but possibly two additional groups compared to the first run.

Figures 9 a,b represents the mapped result of the second run of SOM. It expresses again, that the boundary of this depositional environment can also affect the spatial arrangement of reservoir properties. However, if we compare the mapped clusters to the original contours of the distributary mouth bars, Figure 10 shows the similarities and differences. It can be seen that joint use of three variables in the SOM method produces a similar pattern within the distributary mouth bars as using one variable (i.e. sand content), but the
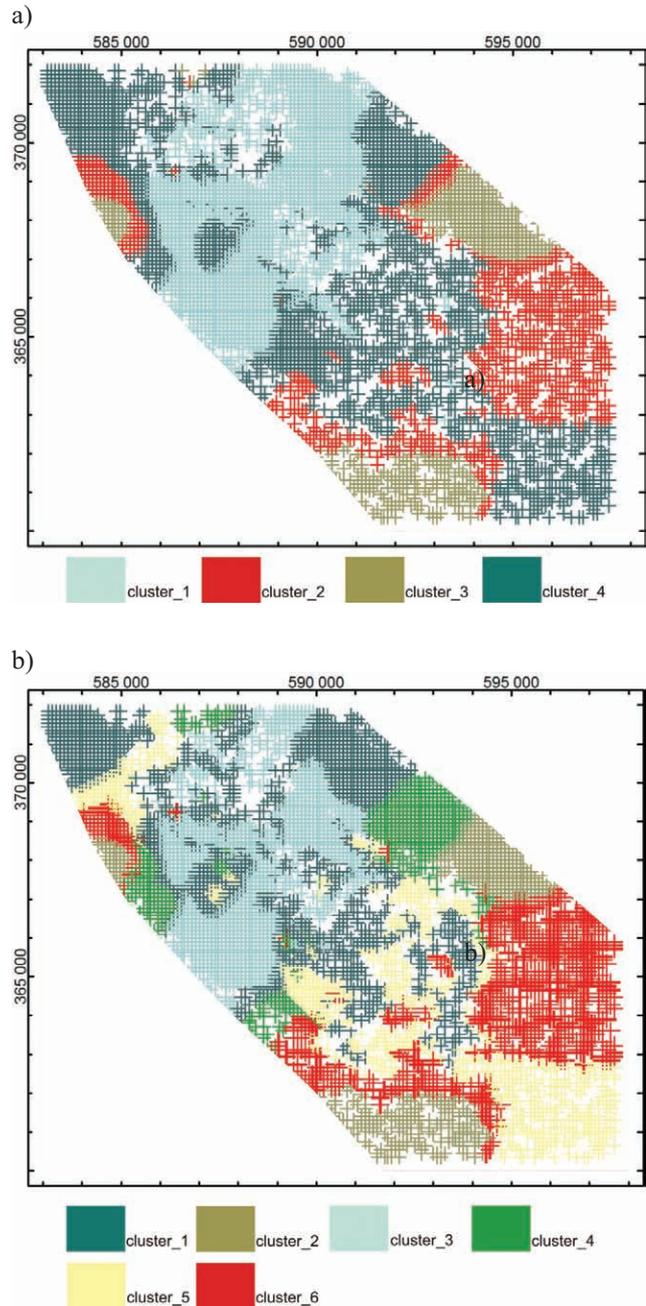
contours of the shapes are different. So, the mapped results show the clusters which contain the points with similar properties in the present stage of diagenesis.

In both cases, the run time of the training algorithm depends primarily on the size of the network and only secondarily on the number of training steps. The presented mapped results (Figs. 8,9) are composed by setting 1000 learning steps in the SOM process. For this setting, (above mentioned numbers of clusters and number of learning steps) the run of SOM gives similar values for the error rates. Table 1 shows the rate of errors for each set.

Table 1: Error rates of running sets in case of running a-b-c-d.

errors rate of sets	case a)	case b)	case c)	case d)
training set	0.02786	0.028644	0.01515	0.01515
test set	0.020765	0.028264	0.031158	0.031158
validation set	0.019769	0.028178	0.0319757	0.031957

Besides the error rates of sets, we can determine the goodness of clustering of SOM. Discriminant analysis can show how well the group a particular case belongs to can be predicted (see Table 2 a-d). Since the size of each cluster is not the same (the clusters contain different amounts of points), the priori probabilities (*p*) are not equal to 1/3 for each subset. The first column of the spreadsheet indicates the percentage of cases that are correctly classified in each cluster. Other columns show the number of misclassifications. According to discriminant analysis, the results of SOM are 95% adequate on average.
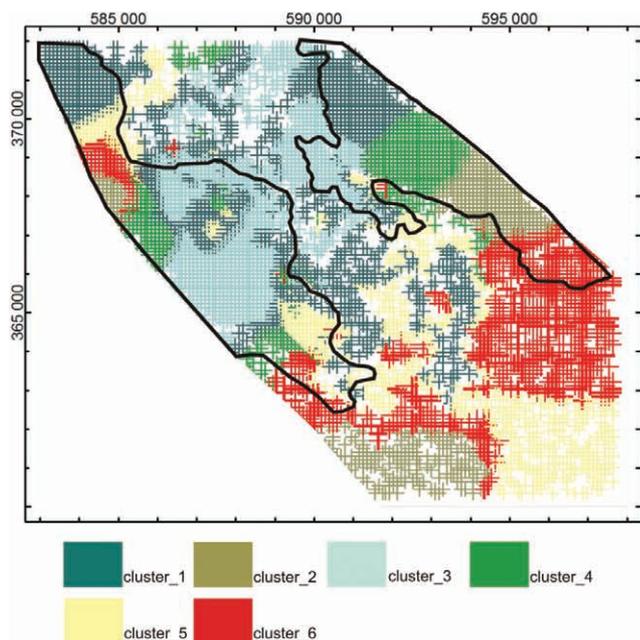

Figure 10: Mapped result for 6 clusters with the original geometry of distributary mouth bars by sand content, (where clusters contain points with similar properties in the present stage of diagenesis).

Table 2a: Classification Matrix (distributary mouth bars – 4 clusters)

C_i i^{th} cluster	Percent Corrects	C_1 p=0.16789	C_2 p=0.39853	C_3 p=0.18582	C_4 p=0.24830
C_1	88.02589	544	56	0	18
C_2	99.38651	0	1458	9	0
C_3	84.31085	0	107	575	0
C_4	92.88840	42	23	0	849
Total	93.07253	586	1644	584	867

Table 2b: Classification Matrix (distributary mouth bars – 3 clusters)

C_I i^{th} cluster	Percent Correct	C_1 p=0.45314	C_2 p=0.31459	C_3 p=0.23227
C_1	98.32134	1640	28	0
C_2	96.80483	30	1121	7
C_3	96.49123	0	30	825
Total	97.41918	1670	1179	832

Table 2c: Classification Matrix (entire dataset – 6 clusters)

C_i i^{th} cluster	Percent Corrects	C_1 p=0.23560	C_2 =0.14747	C_3 p=0.25400	C_4 p=0.13680	C_5 p=0.11200	C_6 p=0.11413
C_1	94.68025	1673	0	72	4	18	0
C_2	99.18626	0	1097	0	9	0	0
C_3	97.42782	43	0	1856	6	0	0
C_4	94.15205	34	0	0	966	0	26
C_5	88.33334	26	0	0	15	742	57
C_6	92.87383	0	18	0	0	43	795
Total	95.05334	1776	1115	1928	1000	803	878

Table 2d: Classification Matrix (entire dataset – 4 clusters)

C_i i^{th} cluster	Percent Correct	C_1 p=0.35520	C_2 p=0.15800	C_3 p=0.32853	C_4 p=0.15827
C_1	94.68025	1673	0	72	4
C_2	99.18626	0	1097	0	9
C_3	97.42782	43	0	1856	6
C_4	94.15205	34	0	0	966
Total	95.05334	1776	1115	1928	1000

5. CONCLUSIONS

Usually Kohonen's network is applied to solve classical pattern recognition problems such as the recognition of a writing character or pictures. Frequently nonetheless, it can be applied to identifying a pattern as a classification problem, since in the background of pattern recognition there are clustering problems. This paper demonstrated the approach of using an unsupervised neural network to distinguish parts of the depositional sub-environment as clusters which contain points with similar properties in the present stage of diagenesis. Thus the SOM is an adequate tool to help identify genetic geological environments, sub-environments supplemented with the mapping process. The first test using the SOM-method reflected considerably well the patterns of the distributary mouth bars using three variables (Figs. 7 a,b). In the case of the second run of SOM, the mapped results cluster does not show exactly the same boundaries of the distributary mouth bars, but a similar pattern within the area can be recognized.

This paper focused on joint handling of more property to characterize parts of the sub-environment; and then mapping these as arranged points in space using some defined rules for geometry.

In fact, the neural network answers do not differ significantly from the original geological model obtained. However, differences are visible, especially at the contour of distributary mouth bars. This raises the question of whether the differences show greater accuracy or greater uncertainty. Nevertheless the discriminant analyses show that the clusters are well-defined at least on average for 95 % of the time.

REFERENCES

AHMED, T., LINK, C.A., PORTER, K.W., WIDEMAN, C.J., HIMMER, P. & BRAUN, J. (1997): Application of Neural Network Parameter Prediction in Reservoir Characterization and Simulation – A Case History: The Rabbit Hills Field – paper SPE 38985 presented at the 1997 SPE Latin American and Caribbean Petroleum Engineering Conference and Exhibition held in Rio de Janeiro, Brazil, 30 August–3 September 1997. doi: 10.2118/38985-MS

AKINYOKUN, O.C., ENIKANSELU, P.A., ADEYEMO, A.B. & ADESIDA, A. (2009): Well Log interpretation model for the determination of lithology and fluid contents.– The Pacific Journal of Science and Technology, Springer.

BÉRCZI, I. (1988): Preliminary sedimentological investigations of a Neogene depression in the great Hungarian Plain.– In: ROYDEN & HORVÁTH (ed.): The Pannonian basin – a study in basin evolution.– AAPG Memoir, 45, 107–116.

CHANG, H.-C., KOPASKA-MERKEL, D.C. & CHEN, H.-C. (2002): Identification of lithofacies using Kohonen self-organizing maps.– Computers & Geosciences, 28, 223–229. doi: 10.1016/S0098-3004-(01)00067-X

FAUSETT, L. (1994): Fundamentals of Neural Networks – Architectures, Algorithms and Applications, Prentice Hall, Englewood Cliffs, NJ.

GEIGER, J. (2005): Sedimentological study of Szőreg-1 reservoir (Algyő Field): a kind of combination of traditional and 3D sedimentological approaches.– In: HUM, L. et al. (eds.): Environmental Historical Studies from the Late Tertiary and Quarternary of Hungary. Department of Geology and Paleontology, University of Szeged, 25–45.

GEIGER, J., KISSNÉ, V.K., VERES, K. & KOMLÓS, J. (1998): A Szőreg-1 telep 3D rezervoár geológiai modellje. (3D sedimentological modelling of Szőreg-1 reservoir), KUMMI Jelentés. p. 216. Tsz. 156–5636., Szeged.

HAYKIN, S. (1999): Neural Networks: a Comprehensive Foundation. Macmillan.– The Knowledge Engineering Review, New York, 13/4, 409–412. doi: 10.1017/S0269888998214044

JOHANSSON, J., JERN, M., TRELOAR, R., JANSSON M. (2003): "Visual Analysis based on Algorithmic Classification" IV, Seventh International Conference on Information Visualization, p. 86.

KOHONEN, T. (1982): Self-organized formation of topologically correct feature maps.– Biological Cybernetics, 43, 59–69. doi: 10.1007/BF00337288

KOHONEN, T. (2001): Self Organized Maps – Springer, Berlin, Germany.

LAMPINEN, T., LAURIKKALA, M., KOIVISTO, H. & HONKANEN, T. (2005): Profiling Network Applications with Fuzzy C-means and Self-Organizing Maps Studies in Computational Intelligence, Volume 4/2005 Classification and Clustering for Knowledge Discovery, Springer Berlin / Heidelberg.

MAHALANOBIS, P.C. (1936): "On the generalized distance in statistics". Proceedings of the National Institute of Sciences of India 2/1: 49–55.

PATTERSON, D. (1996) Artificial Neural Networks, Prentice Hall, Singapore, xy p.

PETTIJOHN, F.J. & POTTER, P.E. (1972): Sand and sandstone.– Springer Verlag Berlin – Heidelberg – New York, 618 p.

RÉVÉSZ, I. (1982): Az algyői Maros-Szőreg szénhidrogéntelepek üledékföldtani modellje – egy fosszilis delta fejlődéstörténete. (Modelling of Maros-Szőreg hydrocarbon reservoir (Algyő Field) – a develop history of the pannonian fossil delta.).– Kőolaj és Földgáz, 115, 176–177.

ROGERS, S.J., CHEN, H.C., KOPASKA-MERKEL, D.C. & FANG, J.H. (1995): Predicting Permeability from Porosity Using Artificial Neural Networks. – AAPG Bulletin, 79/12, 1786–1797.

ULTSCH, A. (1995): Self Organizing Neural Networks perform different from statistical k-means clustering.– In: Proc. GfKl, Basel, Swiss.

ULTSCH, A., KORUS, D. & WEHRMANN, A. (1995): Neural networks and their rules for classification in marine geology, Raum und Zeit in Umweltinformations-systemen, 9th Intl. Symposium on Computer Science for Environmental Protection CSEP '95, Vol. I, ed. GI-Fachausschuß 4.6 "Informatik im Umweltschutz" vol. 7, Metropolis-Verlag, Marburg, 676–693.