DEPOSITIONAL FACIES ANALYSIS IN CLASTIC SEDIMENTARY ENVIRONMENTS BASED ON NEURAL NETWORK CLUSTERING AND PROBABILISTIC EXTENSION

TWO CASE STUDIES FROM SOUTH-EASTERN HUNGARY AND NORTHERN CROATIA

THESIS OF THE PhD DISSERTATION

by

JANINA HORVÁTH

SUPERVISOR

JÁNOS GEIGER

ASSOCIATE PROFESSOR

UNIVERSITY OF SZEGED

DEPARTMENT OF GEOLOGY AND PALEONTOLOGY

Szeged, 2014

INTRODUCTION AND GOALS OF THE STUDY

TRYON (1939) was the first to use the term of 'clustering'. He defined this method to segregate data into groups (cluster). Cluster analysis developed very quickly over the last 70 years and many diverse techniques developed.

During the past few decades a huge number of papers have introduced different multivariate statistical methods and workflows to identify subsurface facies analysis. Most of them have relied on clustering, but few (if any) have tried to use these classifying methods under the surface combined with lateral extension of cluster members. In fact, this approach can be expected to have significant uncertainty because of the scattered lateral distribution of sample points (wells). This dissertation aimed to contribute this issue by addressing several main points: (1) cluster separation using neural network technique; (2) the lateral estimation of point-like qualitative information of cluster members using indicator kriging (IK); (3) the interpretation of the geometry presented by IK; (4) a comparison of the efficiency of UNN and K-means clustering on the basis of results provided by the previous three analyses.

In multivariate data analysis, clustering is a segmentation process in the basic dataset, but unsupervised and supervised methods are distinguished within that. Data clustering is often confused with the comprehensively applied classification methods. In fact, both are segmentation processes, but the first is an unsupervised, and the second a supervised one. In the supervised method, the objects are assigned to predefined classes, but in clustering, there are no predefined clusters or prototypes. This was the main reason of using an unsupervised neural network method for facies segmentation in this study.

The unsupervised network can solve specific problems such as indirect data mining including clustering, pattern recognition and visualisation. The applied clustering process used the 'Self-Organised Map' (SOM) approach, which is a type of artificial neural network. This process was introduced by KOHONEN (1982, 1984). Like each artificial neural network, this is also an analogy of the manner by which the human brain can logically arrange data, and new information.

As a tool for identification, the Kohonen network has been demonstrated in several publications. In those cases the goals were identifications of lithofacies (CHANG et al., 2002), well-log interpretations for the determination of reservoir facies and fluid contents (AKINYOKUN et al., 2009) and classification of biogenic sedimentation (ULTSCH et al., 1995). The present study also demonstrates an application of the Kohonen network, with the aim of identifying depositional facies supplemented by statistical interpretation and lateral extension of clusters.

Two case studies coming from different clastic sedimentary environments demonstrate the applied methods in a constructed workflow. The first one is aboutSzőreg-1 delta plain rock body of the Algyő Field, South-Eastern Hungary. The second case study deals with a deep water turbidity system in Sava Depression, Northern Croatia.

In the dissertation, beside of the identification of depositional facies, particular emphasize is given on the comparison of the applied UNN and the widely used K-means clustering procedures. This comparison is especially based on the classification results of both case studies.

APPLIED METHODS

The sample points in the vertical column of the wells were characterized by vectors, in which the components were some petrophysical and lithological properties. These sample vectors formed the input side for multivariate approaches of this dissertation.

The basic aim of any genetic oriented multivariate approach is to reveal and identify the genetically homogeneous sub-sets of an inhomogeneous set. To meet this goal, one must group/classify the samples. In this grouping/classification many or all the properties describing the samples should be used simultaneously, or in a weighted forms. In general it can be done by a two-step approach: (1) identification of some “seeds” in the multidimensional space defined by the measured properties and (2) join each sample with the appropriate “seed”. If the “seed” of the multidimensional space hold real genetic meaning, this classification may be correct. Recently two methodological groups have been widely used to this complex problem. They are the clustering and neural network families of numerical methods. Their common background is an easy to understand assumption: the closer the samples in the multidimensional space, the more similar their origins are.

CLUSTERING BY UNN

The applied UNN process was performed using SANN (STATISTICA Automated Neural Networks) algorithm. It includes an inter alia Kohonen-training network, called SOFM (Self

Organising Feature Map) networks. SANN, as a state-of-the-art NN solution, is a comprehensive, powerful, and extremely fast neural network method.

After the normalization of variables, the training set contained 60% of all data points, while size of the validation and test sets were 20-20% of the input data. These three subsets were collected by the network in a random way to avoid bias. The training set was used to build a neural network. The validation set was applied to tune the parameters of a classifier and to determine the end of the learning process. The test set helped to assess the performance of the trained clusters.

Usually it is hard to determine the appropriate number of clusters. This number depends on the user, but must be defined at the beginning of the analysis. In this work, the number of clusters was equal to the number of facies in the reservoir which had been known from previous studies of the rock body. In this way UNN resulted in six clusters in the Szőreg-1, and four in the Sava Field. The constructions were based on several initial parameters as e.g. size of network, neighborhood radius, training rate, learning cycle or early stopping.

STATISTICAL DESCRIPTION AND INTERPRETATION

The statistical interpretation consisted of three main steps: (1) analysing of cluster sizes; and (2) comparison of generated clusters by using non-parametric test with auxiliary graphical technique of exploratory data analysis; and (3) the variance analysis checked the homogeneity or heterogeneity of particular clusters.

The clusters contained different numbers of elements. It made difficult their comparison. To avoid this problem Monte Carlo simulation was used to increase the number of elements of the clusters. In this way the resolution of the corresponding probability distribution also improved. The simulations honoured both the shapes and the statistical properties of the original probability distribution functions. As a result, these improved cluster-sets could be compared with well-log readings by using non-parametric tests.

The Goodman and Kruskal gamma coefficient was applied to reveal the relationship between any two rank-ordered variables.

In the characterisation and comparison of clusters several traditional statistical tests were used. The non-parametric Mann-Whitney test verified the significant differences between cluster-means. These analyses were completed by partly the evaluations of histograms and box-plots and by the calculation of within-group and between-group variances (one-way

ANOVA). MILLER and KHAN (1962) proved the theorem of variance decomposition. It says that, supposing normal distribution, the total variance is the sum of the within group and between group variances (WGV and BGV correspondingly). Clustering seeks to minimize within-group variance (WGV) and maximize between-group variance (BGW). It can rarely reach a substantial difference between them. The within cluster variance (WGV) refers to the spread of objects around the mean and the between cluster variance (BGV) is a measure of how cluster centroids spread out from one another

The difference of WGV and BGV can express the suitability of cluster results. The relatively low WGV and larger BGW mean that cluster analysis results in a number of heterogeneous groups with homogeneous contents. The inverse situation implies that clusters are well separated from each other, and there is a high degree of homogeneity within clusters.

This variance analysis also assisted to compare the applied UNN clustering to the widely used k-means algorithm.

PROBABILISTIC EXTENSION USING INDICATOR KRIGING

The lateral extension and the spatial variability of clusters had an important role in the interpretation of UNN results. The lateral geometry (in the geographical space) of a cluster can be given as the geometry of the convex hull determined by its geographically peripheral points. Determination this convex hull obviously needs interpolation in the geographic space, since clusters are known only at finite number of locations. The geostatistical tool designed to interpolate qualitative property is the indicator kriging. It produces probability ’contours’ expressing how likely the appearance of the studied quality is.

The geometry of the acceptably high probability contour may be compared with some well-known depositional geometry which may be different in different depositional environments (e.g. MOORE 1949; PETTIJOHN et al. 1972). Following this thought, the rock body geometry expressed by probability contours in this work is interpreted in terms of depositional facies.

A classed post map is a ‘traditional’ form of visualisation of the lateral distribution of cluster memberships, where the memberships are represented by points in a map. This solution ignores unsampled locations. The indicator kriging (IK) can offer a reasonable solution for this task. This technique is applied to approximate the conditional cumulative distribution function (ccdf) at each point of grid, based on the correlation structure of indicator transformed data points (JOURNEL, 1983, 1986). The indicator kriging can handle categorical variables, too. In this study the clusters were transformed to binary indicators. It is a regular method to estimate conditional probability of occurrences of classes like rock type, soil etc. (e.g. BIERKENS and BURROUGH, 1993).

The indicator values and the global probability distributions were inputs to ordinary kriging method. This process estimated the probability of all clusters in each grid points. Finally, only that cluster was kept a grid point which the largest probability belonged to. Another alternative interpretation to visualize the estimated probability values of a particular cluster was a contour map. This visualization indicated the probabilities of the appearance of each cluster. In this case it was worthy to display the probability values only over 0.5.

According to the different probability level, the boundary (interpolated) geometry of a particular cluster can change. That is why it is necessary to find a reasonable probability value which (1) can outline the shape of physiographic unites well; (2) is higher than 0.5.

CASE STUDIES

The input side of the numerical analyses provided by INA and MOL consisted of partly petrophysical and partly lithological data. Both of them derived from quantitative petrophysical well log interpretations. Unfortunately, the data sets of Sava Field and Szőred-1 were not homogeneous.

CASE STUDY - I.

Szőreg-1 is one of the largest reservoirs of the Algyő Filed, south-eastern Hungary. It has been regarded to be deposited in a delta plain environment. Some earlier works by GEIGER (2003) and by SEBŐK-SZILÁGYI and GEIGER (2012) have proved that this rock body developed in a quite large interdistributary bay by the processes of distributary channels, mouth bars and crevasse splays. These works have given detailed sedimentological and 3D geometrical descriptions of these facies, too. In the case of Szőreg-1 reservoir, UNN clustering method was applied in two depositional sub-environments: (1) emerging distributary mouth bar and (2) prograded bifurcation channels. Their corresponding stratigraphic positions are 34–35m and 24–27m below the top argillaceous marl of the rock body. Their 3D geometry and geostatistical characters have already been known (e.g. GEIGER, 2003; SEBŐK-SZILÁGYI, GEIGER, 2012).

From the Szőreg-1 reservoir, interpreted quantitative petrophysical data of porosity, hydraulic conductivity and sand content was available from 512 wells.

The petrophysical records were measured at every 0.2m intervals. The original readings were also rescaled by averaging into selected interval (24-27m, 34-35m) starting at the top. The averages of these records between the selected vertical intervals were used as input data in the UNN clustering.

CASE STUDY - II.

The second case study was the Sava Field, Northern Croatia. The entire sedimentary sequence belongs to the Neogene and Quaternary periods. Generally, the Middle and partially Upper Miocene sedimentation was highly influenced by pre-Neogene basement palaeo-relief. The analysed two sequences have been built up by Upper Miocene marls, siltstones and sandstones. The latter two psammitic lithofacies were deposited by periodic turbidity currents over the entire depression (e.g., ŠIMON, 1980). These turbidity flows were active in lacustrine environments during the Pannonian and Pontian ages (e.g., VRBANAC, 1996).

The selected reservoir rocks were transformed into stratigraphic coordinate system. The vertical coordinate was measured from the top of reservoirs. Both reservoir rocks in this new coordinate system were cut by lateral surfaces which are parallel with the top. The vertical distance between each lateral surface was 1m. In this way both reservoirs contained about 20 lateral surfaces.

In case of Sava Field, the analysed data came from 78 wells. Geophysical logs with their quantitative petrophysical interpretations of porosity, water saturation and shale volume were available at 0.2m intervals. The original readings were rescaled by averaging into 1m thick interval starting at the top. In addition, one categorical data (a numerical code) was used to describe the lithology. A code-number between 0-10 (according to the shale alternation in the sandy deposit) characterised the lithology. 0 was thus assigned to ‘clear sandstone’ and 10 to ‘marl’.

THESIS

1. I have introduced a complex workflow of NN, indicator kriging, and classical statistical methods to 3D modelling of clastic depositional environments.

2. I have pointed out, that (1) UNN is able to recognize clusters as facies even in such situation where k-means clustering techniques fail to find any reasonable depositional units; (2) one of the advantage of using NN in facies analysis is that its cluster-forming 'capacity' is self-regulated, that is why it is more efficient than 'classic' clustering.

3. In the case of Sava Field, I have defined four NN-facies (Facies C_1-C_4 in the dissertation): (1) low porosity, massive marls of still water sedimentation; (2) laminated sandstones with siltstones and marls of low-density turbidity currents; (3) thin sandstones and interbedded siltstones of low-density turbidities; (4) massive sandstone. I have proved that the above listed facies are equivalents of F4, F3, F2 and F1 faces defined by VRBANAC's former works (VRBANAC, 1996; VRBANAC et al., 2010).

4. I have shown that (1) Facies C_2 (laminated sandstones with siltstones and marls of low-density turbidity currents) deposited as Bouma Td-Te members within the area developed between the bifurcating channels of a fan system; (2) Facies C_3 developed as a lob-type deposit; Facies C_4 was the result of deposition processes acted at the axes of turbidity channels of a sand rich turbidity fan.

5. I have recognized that the mid-fan area of a sand-rich submarine fan system can be applied as facies model for the Sava Field. The described facieses in the lower reservoir locates in the distal part of this mid-fan and the defined facieses in the upper reservoir lies in the proximal part.

6. According to the migration of facieses, I have concluded that the submarine fan system prograded from NW to SE and the lobated sediments shifted laterally, too. This lateral movement characterized only the lower reservoir. The main accumulation migrated from the central part of the reservoir to the edges.

7. I have characterized two physiographic unites in the lower reservoir in Sava Field. These are bifurcation channel and channelized lob. The bifurcation channel has elongated, dendroid shape in the direction of progradation (NW-SE). Its maximum length is 1200-1300m, its width is 750m. The lob with radial fan pattern is about 700m in the major axis (NW-SE) and in the perpendicular direction it is maximum 500m wide.

8. In the upper reservoir of Sava filed I have defined one physiographic unit which is an elongated channel without distribution. This channel has about 2000m long axis from NW to SE and it is maximum 800m wide.

9. Based on the characterization of physiographic units the heterogeneity of reservoirs is very low. The reservoir continuity and connectivity is very good laterally and in the upper reservoir vertically

PUBLICATION IN PROFESSIONAL JOURNAL

HORVÁTH, J., MALVIĆ, T. (2013): Characterization of clastic sedimentary environments by clustering algorithm and several statistical approaches – case study, Sava Depression in Northern Croatia, Central European Geology, Vol. 56/4, pp. 281-296, DOI: 10.1556/CEuGeol.56.2013.4.1

HORVÁTH, J. (2013): Characterization of clastic sedimentary environments by clustering algorithm and several statistical approaches; two case studies (South-Eastern Hungary, Northern Croatia) - (Selected studies of the 2012 Croatian-Hungarian Geomathematical Convent, Mórahalom), Edited by Geiger, J., Pál-Molnár, E., Malvić, T.; pp. 71-85; GeoLitera Publishing House Institute of Geosciences, University of Szeged, Hungary, Szeged, 2013. ISBN: 978-963-306-235-7

HORVÁTH, J., NOVAK-ZELENIKA K. (2012): Application of clustering methods for identification of environments, case study in one Croatian field, in Sava depression in New Horizons in Cenral European Geomathematics, Geostatistics and Geoinformatics (Selected studies of the 2011 Croatian-Hungarian Geomathematical Convent, Mórahalom), Edited by Geiger, J., Pál-Molnár, E., Tomislav Malvić. - pp.8-99, GeoLitera Publishing House Institute of Geosciences, University of Szeged, Hungary, Szeged, 2012. ISBN: 978-963-306-136-7

HORVÁTH, J., NOVAK-ZELENIKA K. (2011): Identification of Palaeo-environments Using Clustering Methods and Indicator Kriging, Case Study from Late Miocene Sandstones, the Sava Depression, Nafta: exploration, production, processing, petrochemistry, Vol.62(11-12), pp. 364-376, Zagreb, Croatia

HORVÁTH, J. (2011): Define of depositional environment using neural network, Geologia Croatica, Vol. 64/3, No – 2011, pp 251-258. Zagreb, Croatia (doi: 104154/gc.2011.21)

MALVIĆ, T., VELIĆ, J., **HORVÁTH, J.**, CVETKOVIĆ, M. (2010): Neural networks in petroleum geology as interpretation tools - Central European Geology, Vol. 53/1, pp.97-115, Budapest, Hungary (doi: 10.1556/CEuGeol.53.2010.1.6)

PUBLICATION IN CONFERENCE BOOK

HORVÁTH, J. (2014): Depositional facies analysis in clastic sedimentary environments based on neural network clustering and probabilistic extension - Conference Book of 6th

Croatian-Hungarian and 17th Hungarian geomathematical congress, "Geomathematics - from theory to practice", Opatija, Croatia, pp. 77-82 (ISBN: 978-953-95130-8-3)

HORVÁTH, J. (2012): Characterization of Clastic Sedimentary Environments by Self-Organize Clustering Algorithm and Several Statistical Approaches, 2nd Central and Eastern European International Oil and Gas Conference, Sibenik, Croatia

HORVÁTH, J. (2012): Statistical characterization of clastic sedimentary environments derived by clustering method, Conference Book of 4th HU-HR and 15th HU geomathematematical congress - "Geomathematics as Geoscience", Opatija, Croatia, pp. 51-61 (ISBN: 978-953-95130-6-9)

GEIGER, J., MALVIĆ, T., **HORVÁTH, J.**, NOVAK-ZELENIKA K. (2011): The role of stochastic view in reservoir characterization – Conference Book of The First Central and Eastern European International Oil and Gas Conference and Exhibition - Siófok, Hungary

HORVÁTH, J., NOVAK-ZELENIKA K. (2011): Application of clustering methods for identification of environments, case study in one Croatian field, in Sava depression, XIV. Congress of Hungarian Geomathematics and the III. Congress of Croatian and Hungarian Geomathematics – Mórahalom (ISBN: 978-963-8221-45-2)

GEIGER, J., MALVIĆ, T., **HORVÁTH, J.**, NOVAK-ZELENIKA K. (2010): Handling uncertainty in the case of lateral extension of log-porosity values in a turbidity reservoir (Analiza nesigurnosti u slučaju lateralne procjene logaritamskih vrijednosti poroznosti u ležištu turbiditnog podrijetla), 4th. Croatian geological congress, Šibenic-Croatia (ISBN 978-953-6907-23-6)

HORVÁTH, J. (2010): Define of depositional environment using neural network (Odreñivanje taložnog okolišom neuronskom mrežom), 4th. Croatian geological congress, Šibenik-Croatia (ISBN 978-953-6907-23-6)

HORVÁTH, J. (2010): Potential application of neural networks in identification of depositional environments, 14th Annual Conference of the International Association for Mathematical Geosciences-IAMG 2010, Budapest, WECO Travel Ltd. , 2010. 5-5 (ISBN: 978-963-06-9829-0)

HORVÁTH, J. (2010): Define of depositional environment using neural network, Conference of MFT, Szeged, Abstract book-GeoLitera, editor: Pál-Molnár, E. (ISBN: 978-963-306-016-2)

HORVÁTH, J. (2009): Multivariable statistical method and potential application of non-controlled learning neural networks in paleontology - 10th Anniversary Conference of the Czech, Polish and Slovak Paleontologists, Banská Bystrica, Slovak Republic (ISBN: 978-80-8083-807-2)

HORVÁTH, J. (2009): Potential application of non-controlled learning neural networks in geology - XIII. Congress of hungarian geomathematics and the II. Congress of Croatian and Hungarian geomathematics, Mórahalom

HORVÁTH, J. (2008): Biometric Research and Multivariate Statistic Treatment of Viviparus Species in Lake Pannon with a Genetic Approach - XII. Congress of Hungarian geomathematics and the first congress of Croatian and Hungarian geomathematics – Mórahalom

PROJECT REPORT

HORTVÁTH, J., WAGENHOFFER, A., GEIGER, J. (2012): "Analyses of reservoirs in Kisújszállás, Fegyvernek, Nagykörű Fields" in Development of 'knowledge base' of general reservoir properties for EOR and IOR methodologies in Panonian (s.l.) clastic CH-reservoirs.—MOL Plc. Project Number: UX0317.12.69/95, 179, MOL Repertory. Budapest, Hungary

REFERENCES

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, Vol. 10, pp. 507-517

BIERKENS, M. F. P. AND P. A. BURROUGH (1993) The Indicator Approach to Categorical Soil Data. I. Theory. Journal of Soil Science, 44, 361-368

CHANGE, H.C., KOPASKA-MERKEL, D.C., CHEN, H.C. (2002): Identification of lithofacies using Kohonen self-organizing maps, Computers, Geosciences Vol. 28, pp. 223-229

GEIGER, J. (2003): A pannóniai újfalui (törteli) formációban levő Algyő-delta fejlődéstörténete – I.: Az Algyő-delta alkörnyezeteinek 3d modellezése (Depositional history of the Pannonian Algyő-delta (Újfalu Formation). Part one: 3D modeling of the sub-environments of Algyő-delta) - Földtani Közlöny, Vol. 133, no. 1, pp. 91-112

JOURNEL, A. G. (1983): Nonparametric estimation of spatial distributions, Mathematical Geology, Vol.15, no.3, pp.445-468

JOURNEL, A. G. (1986) Constrained Interpolation and Qualitative Information - the Soft Kriging Approach. Mathematical Geology, 18, 269-305

KOHONEN, T. (1982): Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, Vol. 43, pp. 59-69

KOHONEN, T. (1984): Self-Organization and Associative Memory, (3rd edition 1989), New York, Springer-Verlag, p. 312

MOOR, R. C. (1949): Meaning of facies. In “Sedimentary facies in Geologic History”, Geol. Soc. Am. Mem. no. 39, pp. 1-34

SEBŐK-SZILÁGYI, SZ., GEIGER, J. (2012): Sedimentological study of the Szőreg-1 reservoir (Algyő Field, Hungary): a combination of traditional and 3D sedimentological approaches, Geologia Croatica, Vol. 65, no. 1, pp. 77-90

PETTIJOHN, F. J., P. E. POTTER, R. SIEVER (1972): Sand and sandstone, Springer-Verlag, New York, p 618

ŠIMON, J. (1980): Prilog stratigrafiji u taložnom sustavu pješčanih rezervoara Sava-grupe naslaga mlađeg tercijara u Panonskom bazenu sjeverne Hrvatske. (Contribution to stratigraphy of sandstone reservoirs depositional system in the Sava Group sediments in Late Tertiary of Pannonian Basin in the Northern Croatia – in Croatian) PhD Thesis, University of Zagreb, Faculty of Mining, Geology and Petroleum Engineering, Zagreb p. 66

TRYON, R. C. (1939): Cluster analysis. Edwards Brothers, Ann Arbor, Michigan, p. 122

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 ´ Vol. 95, no. I, ed. GI-Fachausschuß 4.6 „Informatik im Umweltschutz“ **7**, Metropolis-Verlag, Marburg, pp. 676-693

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, p. 303

VRBANAC, B., VELIĆ, J., MALVIĆ, T. (2010): Sedimentation of deep-water turbidites in the SW part of the Pannonian Basin. Geologica Carpathica, Vol. 61, no.1, pp. 55-69