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

PHD DISSERTATION

JANINA HORVÁTH

SUPERVISOR
JÁNOS GEIGER, ASSOCIATE PROFESSOR

UNIVERSITY OF SZEGED
DOCTORAL SCHOOL OF GEOSCIENCES

DEPARTMENT OF GEOLOGY AND PALEONTOLOGY

Szeged, 2014

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDIX

SYMBOLS AND ABBREVIATIONS

SYMBOLS

D	basic data set
x_i	the i^{th} record,
$x_i = \{x_{i1}, x_{i2}, x_{i3}, \ldots, x_{id}\}$	x_{ij} is the scalar denoting the j^{th} component
d()	distance function
C	discrete constant
$i = (\overline{1\ n})$	i goes from 1 to n each integer
$f(\cdot)$	function
Σ	summa
σ	effective width of the topological neighbourhoods
w	weight vector
w^T	transposed vector (reversal of rows and columns)
arg min $\|x-w\|$	minimizing of the Euclidean distance of x and w vectors
$exp\ (\cdot)$	exponential function
$\rightarrow\infty$	divergence, infinite series
r	vector which shows the position of the excited neuron
$\eta(t)$	Kohonen-learning rate

ABBREVIATIONS

ANN	Artificial Neural Network
CCDF	Conditional Cumulative Distribution Function
DFA	Discriminant Factor Analysis
EDA	Exploratory Data Analyses
GNF	Gaussian Neighbouring Function
HK	Hydraulic conductivity
IK	Indicator Kriging

NLPCA	Non-linear Principal Component Analysis
NN	Neural Network
PCA	Principal Component Analysis
pdf	Probability distribution function
POR	Porosity
P-value	Probability value
SANN	STATISTICA Automated Neural Networks
Sh	Shale volume
SOFM	Self-Organizing Feature Map
SOM	Self-Organized Maps
S.r.	Sand ratio / Sand content
Sw	Water saturation
TLU	Threshold Logic Unit
UNN	Unsupervised Neural Network
WGV	Within Group Variance
BGV	Between Group Variance
TV	Total Variance

1 INTRODUCTION AND GOALS OF THE STUDY

1.1 PREAMBLE

This dissertation is a methodological work, but addresses the definition of sedimentary facies and their characterisation in a specific sedimentary environment. The dissertation uses the term 'depositional facies' as given by Moore (1949): "Sedimentary facies is defined as any areally restricted part of a designated stratigraphic unit which exhibits characters significantly different from those of others parts of the unit". In this wording the main criteria of this facies definition are: lithology, sedimentary structure, sedimentary body geometry, palaeontology, and palaeocurrent pattern (Selley, 1970, p.1).

Clastic sedimentology, and within that facies analysis, has used applied mathematics for a long time. Several mathematical and statistical applications help in facies analysis. At first attention was on analysis based on grain size distribution rather than the separation of particular sedimentary facies and their characterisation. The identification of clastic deposits focused on mathematical and statistical analysis at the beginning of the 1900s. It was Udden (1914) who first wrote about the characterisation of accumulations of clastic sediments using grain size distribution. Many sedimentologists followed this path until the 1950s, supplemented with several different statistical parameters and their correlations (e.g. Folk, 1954; Folk and Ward, 1957; Friedman, 1961, 1962; Passega, 1964, 1972). During this period sedimentologists concentrated on the separation of the different depositional facies using statistical analysis. After the 1960s, and for several decades, the main goal was to demonstrate that in sandy sediments, in the lithification stage of sandstone diagenesis, the lateral distribution of grain size characters could be interpreted in terms of depositional facies. Spencer (1963) and Visher (1967, 1969) presented a method to separate the depositional facies.

The new method was not accepted by everyone. According to Klovan (1966) such a simple statistical analysis for the interpretation of log-normal distribution was not good enough. He claimed that these methods could be used in the identification of depositional facies only if the depositional environment was well-known at the time of analysis. He also noted that diagenesis, which plays an important role in lithification, was not taken into account in this method.

This issue of diagenesis returned again and again to the fore, even in the 1970s and 1980s. For example, in the national scientific community Bérczi (1971) and Geiger (1982) created a comprehensive and very useful method for deposition facies analysis. Geiger (1982) also drew attention to the grain size distribution of clastic sediments modified by diagenesis. He also noted that the lithology parameters are also basically defined by sedimentary environments after the lithification state and maybe also the lithomechanical properties (Geiger, 1982).

After the 1960-1970s the multivariate statistical approaches became widespread in the field of sedimentology. Factor-, principal component-, discriminant- and cluster analysis were the most popular methods at this time. But we can find some multivariate approaches from the 1930s too.

Klovan (1966) demonstrated the application of factor analysis of the weight per cent of grain size classes in the identification of depositional environments. In this methodological work, Klovan used genetically well-known samples published by Krumbein and Aberdeen (1937). The aim was to demonstrate that the applied method was supported or justified by previous knowledge, so he only presented a comparison between arbitrary statistical measurements and the results of the applied factor analysis. Feldhausen and Ali (1974) analysed the same dataset (Krumbein and Aberdeen, 1937) but they used a kind of cluster method completed by Wilk's lambda test and discriminant analysis. The test showed the significance of the defined facies, and discriminant analysis was used to extend the classification to other samples of unknown genetics. Sahu (1964) used discriminant analysis in order to distinguish adjacent mechanisms and environments with similar energy.

The multivariate statistical method, especially the application of cluster analysis, provided a new, appropriate means for the analysis of clastic deposits. Multivariate analysis is able to handle simultaneous observations, several data points, and to analyse more than one outcome variable. Since the aim was a kind of pattern recognition (facies) using high-dimensional properties, the most suitable methods were the classification processes. In these approaches, the basic principle is the following: the more similar the way in which the samples are deposited, the closer their positions are in the property space, and thus they belong to the same group using the clustering method. The goal is to form units which are able to describe sedimentary facies through common characteristics.

One may say that clustering has been the generalised procedure for a wide variety of geological problems since the 1960s. In 1982 and 1984 Geiger applied clustering to textural and morphogenetic analysis of depositional sandstone bodies; Gedeon et al. (2003) used clustering to identify lithofacies from well log data for reservoir characterisation. The recognition of lithofacies or depositional facies is not an easy task in a heterogeneous rock body or reservoir, however. Facies from e-log measurements and the lithofacies of core data are not identical. Sedimentary features are a combination of petrophysical, depositional, and diagenetic properties (Bhatt and Helle, 2002; Gedeon et al., 2003).

Nowadays, principal component analysis, discriminant analysis, K-means analysis and several types of artificial neural network are frequently applied techniques for facies identification based on core data, e-log data or a combination.

As these examples show, many papers relied on clustering the objects (samples) in the sample (or parameter) space to identify lithofacies from well log data. Few (if any), however, have tried to use these classification methods under the surface combined its results with lateral extension of the cluster members. In fact, can be expected that this approach will have significant uncertainty due to the scattered lateral distribution of sample points (wells). This study aims to contribute to solving this issue by addressing a detailed workflow based on neural network clustering to separate the input data set, a lateral estimation of point-like qualitative information of cluster members by indicator kriging (IK); a way of interpreting the geometry presented by IK, and a statistical description and comparison of cluster facies.

1.2 OBJECTIVES OF THE STUDY

Different clustering methods and various neural network techniques are often used to perform the analysis of geological data; this dissertation also incorporates prior knowledge of a problem with the application of a neural network, the interpretation of depositional environments and mechanisms. This work is based on an unsupervised type of artificial neural network because as a segmentation methodology clustering is also an unsupervised process. This distinguishes clustering from classification. Generally the unsupervised neural network and the clustering technique are similar according to the mechanism: the data structure is explored in only one dataset without a control set. This determines which type of neural network is applicable for the clustering in the study. An unsupervised neural network, in this

case the Kohonen network, was therefore used. Despite the fact that a supervised network is more commonly applied in several fields of geology, there are also many examples of the application of unsupervised networks.

An unsupervised network can solve specific problems of indirect data mining, such as, clustering, pattern recognition and visualisation. As a tool for identification, the Kohonen network has been demonstrated in several publications. In those cases the goals were lithofacies identification (Chang et al., 2002), well-log interpretation 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 basis of probability.

The study is based on the term depositional facies as stated by Moore (1949) which includes criteria such as lithology, sedimentary body geometry and palaeocurrent patterns, but ignores sedimentary structure and palaeontology. In case of the selected study area there are core samples available but these aspects of the facies were not analysed. Based on the available data, however, all other criteria were involved in the identification.

According to Moore's definition, depositional facies have characteristics which are significantly different from features of other parts of the unit. Various facies interfinger with one other, however, which implies that they are not laterally disjunctive groups. In contrast, clustering methods separate disjunctive groups. This results in the following questions. How or why is it possible to define the separated clusters as laterally interfinger facies? The goal is not to define the clusters directly as facies. Using the second criteria of facies terminology, the visualised geometry is analysed, which was displayed by probabilistic method. This step can remove the contradiction between the disjunctive clusters and the interfinger depositional facies. The mathematically separated clusters are only point-like results, which are extended laterally using probability distribution.

The spatial appearance of the particular clusters may be different depending on the probability which was used to delineate the geometry. They can also interfinger with each other, so the spatial border of clusters can be not defined directly. The lateral extension is the bridge between the spatial features of depositional facies and the mathematically strict sense of the cluster (disjunction). In this lateral extension indicator Kriging was used because it can estimate the probability of the appearance of clusters at each grid node on the map.

Probability based visualisation can assist not only in the visualisation of lateral distribution of clusters, but the probability contours of clusters can also be useful as blank polygons. Within these polygons the spatial variance and lateral continuity can be analysed. In this manner the palaeocurrent pattern (as the last criteria in terms of depositional facies) also can be characterised by lateral continuity and spatial variance based on the variogram analysis.

The whole workflow contains several applied techniques, from the pre-processing to the interpretation. Figure 1-1 shows the main steps of this procedure as a flow chart.

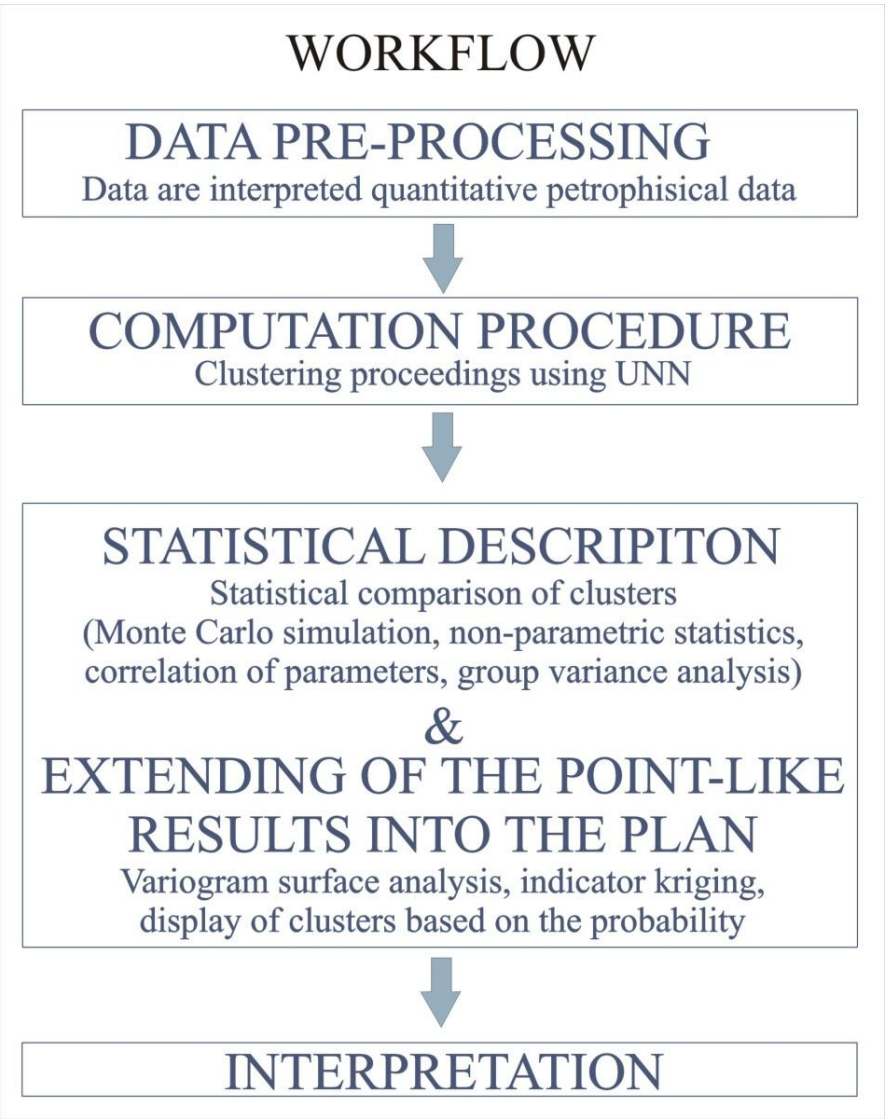


Figure 1-1: Workflow of the applied methods in identification of depositional facies

1.3 THE INTERPRETED DEPOSITIONAL ENVIRONMENTS

This dissertation presents two case studies, from different clastic sedimentary environments. The first involves the Szőreg-1 reservoir of the Algyő field, south east Hungary, which is a delta plain environment. This study area was selected because it was the subject of a research and development (R&D) project several years ago. This R&D project was run by the University of Szeged, Department of Geology and Palaeontology, and the Hungarian Oil and Gas Plc (MOL). This area could be described as one of the most investigated areas in the world, considering the number of boreholes. In this field a huge number of wells have penetrated the selected rock body, and several papers and reports deal with the Szőreg-1 reservoir. This is why it seems an appropriate study area in which to test the framework of the applied method.

The second case study involves a deep water turbidity system in the Sava Depression, north Croatia. This study focuses on two reservoirs located in an oil and gas field in the Sava Depression. On one hand, this study area was the subject of internship during my Ph.D. studies, but on the other hand, using the developed workflow for depositional facies identification had great potential. These fields represent completely different sedimentary environments, and there were also other types of data available. The previously small scale analyses of these selected reservoirs have not yet been published. Several papers have described the reservoirs in these areas, but without vertical decomposition.

The present dissertation analyses the depositional facies using high resolution stratigraphy and palinspastic and vertical decomposition in both case studies.

The components of the input vectors are from basic petrophysical data (corrected and interpreted well-logging data). In the first study, porosity, hydraulic conductivity (as a multiplication of the thickness of the flow unit and the corresponding permeability), and sand content were available for the clustering method. In the other case study, in addition to the measured e-logs, water saturation, shale content and a categorical variable which described the lithology were applied.

The dissertation is organised as follows. Chapter 2 outlines the classification methods which have been successfully applied to facies analysis since the 1970s. It highlights several disadvantages of the commonly used clustering processes.

Chapter 3 introduces neural network application and focuses on the Kohonen neural network, the so-called Self-Organised-Maps (SOM).

Chapter 4 presents an overview of the workflow of the applied methods.

Chapter 5 presents two case studies based on the workflow. This section contains a description of the study areas, the basic datasets, and the interpretation of results using lateral extension, and variography analyses. At the end of each case study several statements are made about the particular area.

Chapter 6 deals with differences between the applied neural network method and the widely applied K-means clustering, in light of the results of the case studies.

The chapters in the final section of the dissertation discuss and draw conclusions based on both study areas.

2 OVERVIEW OF CLASSIFICATION PROCESSES AND THEIR APPLICATION IN FACIES IDENTIFICATION

Tryon (1939) was the first to use the term ‘clustering’. He defined a method to segregate data into groups (clusters). Cluster analysis developed very quickly over the last 70 years and many diverse techniques developed.

In multivariate data analysis clustering is a segmentation process in the basic dataset; 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 process. In the supervised method the objects are assigned to predefined classes, but in clustering there are no predefined clusters or prototypes.

In view of the purpose of the analysis, the definition of clustering is as follows: cluster analysis divides data into groups when the main information in the groups is not the description of the linked objects, but rather their relationship. In other words, cluster analysis is an exploratory data analysis tool which aims to arrange the different objects into groups in a way that the degree of association between two objects is maximal if they belong to the same group and minimal otherwise (Gun et al., 2007).

Mathematically, the definition is the following: suppose that there are n objects (records, data points) in the basic dataset, and each object is d-dimensional (they are described by d-different variables), $D=\{x_1, x_2, x_3, \ldots, x_n\}$, where the x_i is the i^{th} record and D is the symbol of basic dataset. Each x_i data point can be written in the following form: $x_i = \{x_{i1}, x_{i2}, x_{i3}, \ldots, x_{id}\}^T$, where x_{ij} is the scalar denoting the j^{th} component of x_i. Two records (e.g. x_i and x_j) belong to the same cluster if $d(x_i, x_j) < C$, where d(.) is the distance function, in other words it is the measure of similarity, and C is a constant distance value.

The literature of cluster analysis uses a number of different terms for “cluster” such as ‘group’ and ‘sub-set’. Hartigan (1975) described this approach as ‘cases-by-variables data structure’. In this paper ‘group’, ‘class’ and ‘subset’ are used as synonyms for ‘cluster’ to avoid repetition of the word.

As the literature uses different terms for ‘cluster’, in cluster analysis these terms are used without a unified definition (Everitt, 1993). Bock (1989) summarised some criteria and

requirements that should be fulfilled for all objects in a particular cluster, however. Data points in a cluster have to:

(i) have the same or closely related properties;

(ii) represent small distances or dissimilarities;

(iii) have a connection with at least one other object in the particular cluster;

(iv) be clearly distinguishable from other clusters (the complement).

If criteria (iv) is expected, then it is so-called ‘crisp’ or ‘hard’ clustering. If disjunctive subsets (i.e. criteria (iv)) are not required, then it is ‘fuzzy’ clustering. In this case a record may belong to more than one cluster with different probabilities.

Hard clustering is also divided into several types as non-hierarchical or hierarchical clustering. The latter has two main segmentation categories according to the algorithm: agglomerative and divisive algorithms. Figure 2-1 indicates the system of cluster analysis types.

This study will focus on crisp clustering.

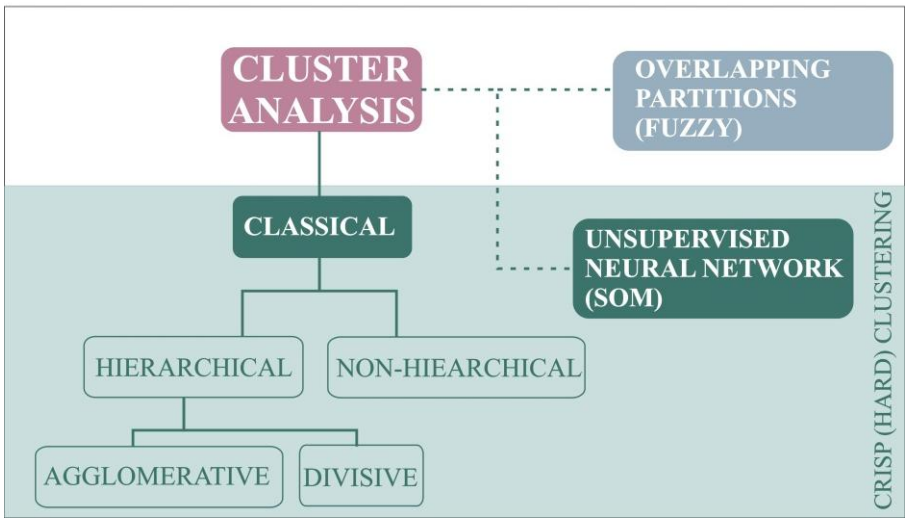


Figure 2-1: System of cluster analysis methods

The goal of hard clustering is to separate groups where the within group variance is less than the variance between the partitioned clusters. These clusters contain similar records, which means that they are close to each other in the d-dimensional features-space. Of course

this similarity, or distance between two data points, is based on the dissimilarity measure or similarity coefficient which is an important parameter in cluster analysis (Anderberg, 1973; Jain and Dubes, 1988).

It should be emphasised, that unlike classical statistical procedures, cluster analysis is generally used when there is no prior hypothesis about groups. This is a typical problem in the exploratory phase of research.

Multivariate statistical approaches, such as classification techniques (cluster analysis, discriminant analysis, principal component analysis, etc.), can provide good opportunities to explore sedimentological characteristics, and the morphological features of depositional rock bodies. There are several clustering techniques that use either multidimensional statistics or one of the artificial neural network methods.

As a particular method of data analyses, cluster analysis is able to handle multidimensional data and to produce primary knowledge through partitioning datasets (Duda et al., 2000; Hastie et al., 2001; Ding and He, 2004). In this way the data space is transformed into a feature space when the abstract space is defined by a feature extraction procedure. It transforms the raw data into sample vectors and represents a reduced number of effective features (Haykin, 1999). Clustering is a common procedure, and the first scientists, Wolff and Pelissier-Combescure (1982), and Serra (1986) applied principal component analysis (PCA) to clusters. Their goals were to separate facies using well readings. Delfiner et al. (1987) and Bush et al. (1987) also tried to identify lithological units with discretion of log values based on discriminant factor analysis (DFA). As Haykin (1999) highlighted, in both approaches (PCA, DFA) the projection inflicts a kind of distortion on the original data dimension. These projections are simple linear maps of the multidimensional dataset or singular value decompositions (Jolliffe, 2002). What is the problem with these multivariate approaches? Duda and Hart (1973) and later Grimm (1987) emphasised that in these approaches the main problem is the "discrete optimization that projects the cluster centres as far apart as possible, while the points of the same cluster are close to each other" (in Bhatt and Helle, 2002).

K-means clustering method is a classical and widely used clustering process which was first introduced by Macqueen (1967) Later this method was also described by Hartigan (1975) and Hartigan and Wong (1979). This method is a centre-based algorithm, and it is considered very efficient for clustering large and high-dimensional datasets. The K-means method applies centroids as prototypes to represent a subset or group in the overall dataset.

Ultsch (1995), however, has suggested that this method was not able to recognise clusters in any situation and sometimes failed to find any reasonable groups.

Unsupervised Neural Network (UNN) clustering is generally compared to K-means because the process is most similar to centroid methods. This is why it is important to deal with this similarity in more detail theoretically in the dissertation, and with respect to the case studies.

There are numerous types of neural networks. Two main types of artificial neural networks are supervised and unsupervised networks. These terms are equivalent to supervised and unsupervised types in segregation processes, in determining when there is a predefined control group or not. Clustering is one kind of unsupervised segmentation process. This implies that if we do not have control sample the unsupervised network alone might be a solution.

Recently, artificial neural network has been a commonly used procedure in log-correlation, facies identification and lithological units separation. Originally this computer-aided approach was introduced in log correlation by Fang et al. (1992) and Gill et al. (1993). Rogers et al. (1992) used a kind of supervised neural network technique to identify lithology types from e-logs, but their solution is the equivalent of a classification and not a clustering separation because supervised neural network is based on control samples.

Without a control sample the Self Organising Map (SOM) is able to group the objects and segregate the database into subsets. SOM belongs to the family of unsupervised artificial neural network methods. These procedures can subdivide datasets without requiring a reference set. SOM was originally introduced by Kohonen (1982, 2001). It is widely used for clustering and can be considered a discrete version of non-linear principal component analysis (NLPCA) (Cherkassky and Mulier, 1998). The unsupervised network can solve specific problems such as indirect data mining, including clustering, pattern recognition and visualisation. The Kohonen network was demonstrated in several publications as a tool for identification including in lithofacies identification (Chang at al., 2000), well log interpretation for the determination of reservoir facies and fluid contents (Akinyokun at al., 2009) and the classification of biogene sedimentation (Ultsch at al., 1995).

Kohonen neural network method (SOM process) will be introduced in detail in the following section of the paper.

As mentioned above, the K-means method is very similar to the unsupervised neural network algorithm. According to Ultsch (1995), Varfis and Versino (1992) or Murtagh and Hernandez (1995) this unsupervised network is the Kohonen neural network. Despite the similarity they also claim that sometimes K-means fails to find any reasonable groups when SOM is able to recognise clusters. Warren (1994) and Ultsch (1995) dealt with this issue and performed tests for comparison. They noted only that these two methods do not have a statistical analogy in spite of the fact that SOM and K-means have similar algorithms in cluster separation. Ultsch (1995) demonstrated that if the dataset contains chainlink clusters, the results of UNN and K-means will be different. He shows a test, however, and not a theoretical proof. Nonetheless, it is also an acceptable proof, since the outcomes of the linear (K-means) and non-linear (SOM) algorithms were compared. The test by Ultsch (1995) was a geometrical experiment. The input dataset contained two so-called chained clusters or non-globular datasets. Figure 2-2 shows the two linked data groups. It is clear that there is no linear projection which can completely divide these into two or more sub-clusters (Fig. 2-2). This is important, as Lorr (1983) suggested distinguishing between at least two different cluster types: the compact cluster and the chained cluster.

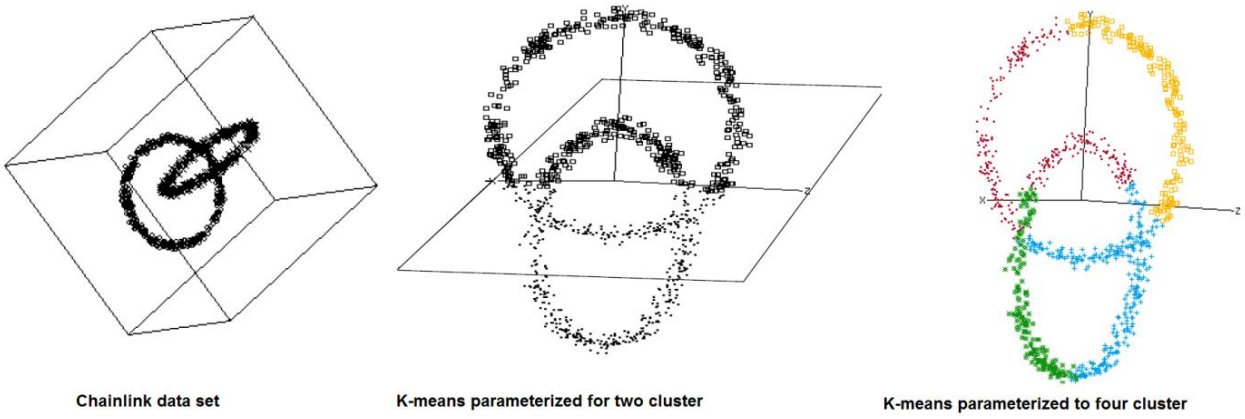


Figure 2-2: A chainlink dataset and its separation using the K-means method

According to the definition of Gan et al. (2007) "The chained cluster is a set of data points in which every member is more like other members in the cluster than other data points not in the cluster. More intuitively, any two data points in a chained cluster are reachable through a path i.e. there is a path that connects the two data points in the cluster." This means that there are cases when this path cannot be given by linear projection, so the clusters cannot

be separated by any hyperplane. Conversely, compact clusters can be represented by centres (Gan et al. 2007). Consequently, it can be accepted that in some cases the NN method can better separate datasets than the similar K-means algorithm.

3 KOHONEN NEURAL NETWORK AS ONE TYPE OF CLUSTER ALGORITHM

3.1 INTRODUCTION TO THE WORLD OF ARTIFICIAL NEURAL NETWORKS

3.1.1 BASIC THEORY

Altrichter et al. (2006), Haykin (1999) and Rojas (1996) summarised the history and development of artificial neural network. The neural network approach has been a well-known development tool over the last couple of decades. Generally, neural networks are considered modern interpretation tools with several purposes. The wide application of neural network has been prompted by the development of computer technology in recent decades (e.g. an increasing number of different software applications contain additional neural network options as built-in modules).

The neural network method is derived from biological systems. The study of the nervous system and neuronal structure launched the development of this area of science. In natural neural systems, a large number of connected neurons are able to solve very different complex tasks. This principle led to the invention of an adaptive device which is able to perform computational tasks through a system of artificial neurons (processing elements). What is the difference between the traditional computation tools and the neural networks? The artificial neural network has a parallel structure, as in the natural neural systems, and ANN is able to learn during application. This ability ensures that the complex problem solving capabilities of these systems are more effective. This property has ensured the rapid development of neural networks over the last 60-70 years.

In the early days of artificial intelligence, Rosenblatt (1957, 1958) developed a machine called the perceptron, based on memorizing the pattern of the human mind. In 1958, he proposed the perceptron as a more general computational model than McCulloch-Pitts units (McCulloch and Pitts developed the first artificial neuron, the Threshold Logic Unit (TLU) in 1943, which was later improved by Rosenblatt). The essential innovation was the introduction of numerical weights and a special interconnection pattern. In the original Rosenblatt model the computing units were threshold elements and connectivity was determined stochastically.

In their study of the perceptron Minsky and Papert (1969) retarded the development of neural networks for a short time, and their work showed that the perceptron was only able to resolve linearly separable classification tasks. Some twenty years later the development of neural networks was in the spotlight again when Hopfield (1982) denied Minsky's statement about the perceptron. Rumelhart et al. (1986) published the Hopfield-algorithm and demonstrated its efficiency.

In the 1980-1990s research into artificial neural networks and their application grew quickly.

3.1.2 IMPORTANT DEFINITIONS AND FUNCTIONS

This chapter summarises the most important definitions and properties characterising an artificial neural network model. The theoretical introduction is based on Altrichter (2005), Haykin (1999) and Rojas (1996). The following chapters will use these terms and definitions.

An artificial neural network is a parallel distributed process which constitutes simple processing units. These are able to order information and knowledge. The ANN is able to learn and use the acquired information. This process is directed by a learning and recall algorithm. The simple processing units that build up the UNN are the neurons. A neuron is a processing element with multiple inputs and only one output. The connections between the input and output are usually described by a non-linear function (transfer function). These connections are illustrated in Figure 3-1, where x_i is the input ($i=\overline{1\ n}$), w_i shows the weights ($i=\overline{1\ m}$), s is the symbol of the linear combination of weighted inputs, $f(\cdot)$ is the non-linear projection, and y is the output value (in other word: activation).

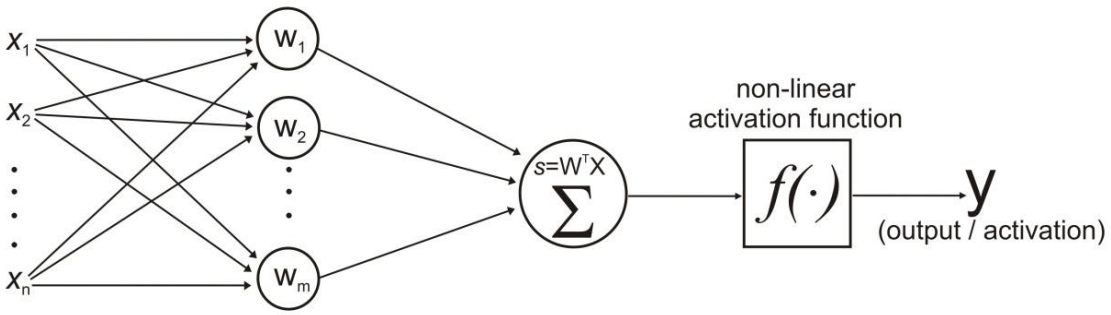


Figure 3-1: Model of non-linear neuron

Figure 3-2 represents only general cases, when the processing element is a non-linear unit, but there are artificial neural networks (e.g. unsupervised neural networks) which make up linear processing units. Figure 3-2 demonstrates the construction of one neuron.

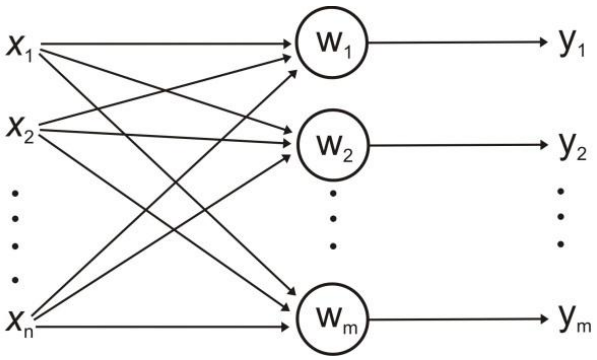


Figure 3-2: Model of a linear neuron

If ANN is based on linear neurons the sum of weights is $s = \sum w_i \, x_i = w^T x$ and the output y is based on this sum ($y = s = w^T x$), but a special learning process such as Kohonen learning can support the development of weights, and may also be a non-linear method (it is the $f(\cdot)$ non-linear function in Figure 3-1). The Kohonen neural network is also a special case.

In a structural approach, the framework of ANN is a weighted spanning tree. ANN consists of points and edges (neurons and connections) and these connections are oriented. This process, as a signal-flow, is a well-defined set of rules.

There are two types of algorithms based on the structure of ANN: feedback and feed forward systems. The unsupervised neural network belongs to the feed forward systems, whose topology is a directed acyclic graph (with no directed cycles). In the other words, there are no nodes (vertices in the graph) which could be connected in a closed chain. The other system is the feedback (recurrent) neural network which comprises a structure cycle. Contrary to feed forward NN, recurrent NN applies its internal memory to processing arbitrary sequences of inputs.

The unsupervised network has a feed forward structure. This means that ANN represents the input functions. There are no other internal conditions, only the weights. Conversely, in the feedback structure, ANN back propagates the output values for input. In this way, ANN represents a dynamic system with short-memory.

The feed forward ANN is ordered in layers and each layer receives signals only from the previous layer, but if ANN is constructed with only one layer, it is so-called single layer system. The single layer neural network is the perceptron (developed by Rosenblatt (1957, 1958)). The perceptron is based on a kind of threshold function during the learning process.

A more complex network system, which has more layers, is the so-called ADALINE network (Adaptive Linear Element), developed by Widrow and Hoff in 1960. ADALINE is based on the McCulloch–Pitts neuron.

The most interesting property in ANN is the adaptation, which is the learning ability. This implies that a neuron is able to change the input-output behaviour as a response to environment change. Why does the input-output behaviour change? Since the activation rules are fixed, the input and output vectors are not modifiable, therefore the weights are changeable. Several such learning rules are available for neural network models, but the learning method must adapt to the fixed network parameters. This adaptation is based on previous experience until a solution is found, if it exists (Rojas, 1996).

Let's look at the classes of the most important learning algorithm comprehensively. There are three main learning methods: supervised, unsupervised and the analytic learning methods (note, the last one is not a real learning process).

In supervised learning there is also training data. This consists of a set of training pairs made up of an input object and a supervisory signal. The training is based on the expected responses for some input and the actual response according to the training pair. The differences between the actual and expected responses indicate the modification required to reduce this variance through optimising the weights of the input. This difference needs to be lower in the next learning cycle until it equal to the minimum learning rate.

Contrary to supervised learning, in the unsupervised method there is no reaction that can control the learning. ANN needs to develop an 'attitude' to find hidden structure in the input data which is based on the similarity of input signals without labelled responses. In the Kohonen neural network the unsupervised learning technique means that the system learns the topology and distribution of the input data. The algorithm tries to explore every correlation among the input, and segregates the input into categories or clusters. During this method the system has the ability to modify itself, which makes the categorisation improve. This may also be referred to as a self-organising network.

For this dissertation, the most important way of learning is the unsupervised method, and therefore the next part deals only with those briefly. The most important learning methods are (i) Hebbian learning, which is sometimes supplemented by (ii) a competitive learning mechanism.

In general these two learning types are as follows:

(i) *Hebbian learning mechanism*: This learning process has biological roots (Hebb, 1949). His assumption was as follows: if two "areas" in the brain are often activated simultaneously, direct contact develops between these "areas" and they are able to activate each other later. During learning process the connection between neurons, the synaptic weights, changes. This procedure is the so-called long-term potentiation. The changing of weights is based on Hebb's rule. According to the rule, the strength of connection between two neurons (weight coefficient) changes in proportion to the multiplication of activation of these neurons. Several alternatives to Hebb's learning rule have been developed, however. These modified Hebb's rules have a very important property: a normalisation process is incorporated in the rule. The reason for this modification is that when using the original Hebb's rule the weights may increase beyond all thresholds. There are several methods for the normalisation procedure. The one actually applied will be discussed in the next subsection.

(ii) *Competitive learning mechanism*: this learning process is one of the most important algorithms. Often, it completes the Hebbian learning in the unsupervised ANN. The neurons in the same layer compete with each other, or increase the activation level through the lateral connections (only one is the "winner" with the highest activation value). The output of the winning neuron, and only this one, will be active (the "winner takes all"). In contrast, the neurons in a cooperative learning mechanism support the activation level of the others (common winning).

During the competitive learning, ANN separates the sample space into regions. The response for input from a particular region can activate only one processing element, so, the procedure results are a kind of clustering, or partitioning.

Generally, competitive learning involves two tasks. In the first step, the activation value of each neuron is determined by the actual weight, and after this the winner will be selected. The change of weights is the second step. It is also based on Hebb's rule. It can

signify the modification of only the weight of the winning neuron or it also can mean changing the weight of the winner and also the neighbouring neurons (according to the previous setting of the particular ANN). So the competitive process supports learning, but during the competition there is no learning procedure. In this step only the selection of the "engine" of learning (the winning neuron) is selected. After that, it can influence the neighbouring neurons in the weight-modification procedure. This implies the real learning mechanism. The choice of the active neuron can be made in numerous ways. In Kohonen NN, for example, it is directed by the non-linear activation function and the lateral connection between the processing elements. This non-linear function ensures that the output of the active element will be one, and that the other case is zero. This will be described in more detail in the next subchapter.

3.2 KOHONEN NEURAL NETWORK IN THE GLOBAL SYSTEM OF ARTIFICIAL NEURAL NETWORKS

The applied clustering process used the 'Self-Organised Map' (SOM), which is a type of artificial neural network. This process was introduced by Kohonen (1982, 1984, 1990, 2001). 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. This is a kind of associative memory, which supports the systematic organisation process without any external help. This implies an unsupervised neural network (UNN) method belongs to the feed forward, non-linear methods such as ANN (Fig. 3-3).

In the unsupervised learning method, the network tries to learn the data structure in order to separate the data into clusters without any help (like a reference set). In other words, the network has to assume that cluster membership is broadly defined by the input patterns sharing common features, and that the network will be able to identify those features across the range of input patterns ([1]Bullinari). This is a really simple means of cluster organisation because during the self-organising process the relationship of the arranged input in the feature-space is maintained and refined in the iteration. The name Self-Organised Map originates from the position of the neurons that are arranged in a grid like (metaphorically) a

[1] Bullinari, J. A.: http://www.cs.bham.ac.uk/~jxb/inn.html, 09.10.2013.
[2] Unpublished report about the study field in Sava Depression - Tertiary CO2 Injection (2003): INA, Zagreb

'map', which is the Kohonen layer. This is a computation layer where complex data structures are mapped using inner spatial non-linear projection.

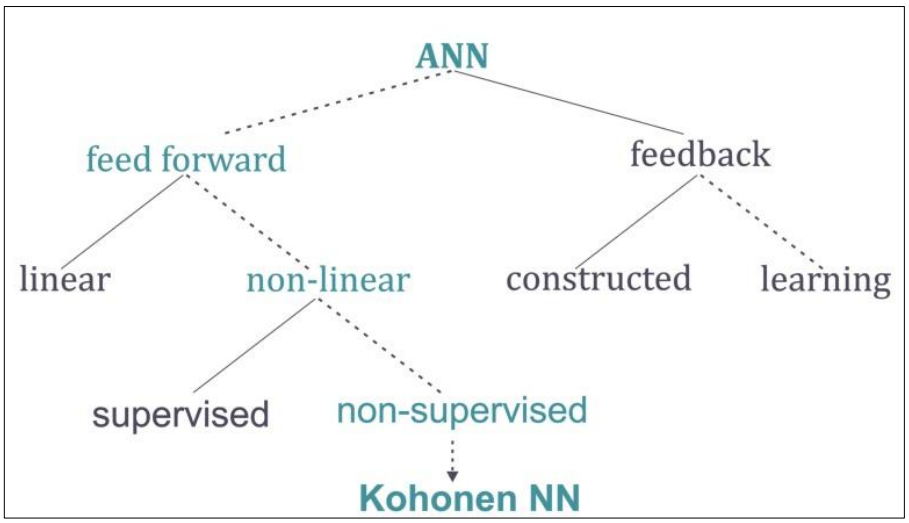


Figure 3-3: Position of Kohonen neural network in ANN system

SOM can be characterised by three essential properties. The algorithm proceeds are (i) competitive, (ii) cooperative, and (iii) adaptive processes (Haykin, 1999).

Let's briefly consider these features; later it will be important in the description of the process.

(i) The *competitive learning* method implies that the network has to have the ability to recognise the structure of the multidimensional basic dataset using the method of dimension reduction (Kohonen, 1982; Haykin, 1994; Fausett, 1994; Patterson, 1996). This reduction is only a "queasy" one, however, since each neuron is an n-dimensional weight-vector, where *n* is equal to the dimension of the input vectors.

In the learning processes, the competitive learning feature helps select the "winning neuron". This neuron is active only at a time of iterative learning. It is also called the 'winner-takes-all' neuron. The method of inducing winning neuron competition among the output elements is to apply lateral connections between them. This lateral network is called a negative feedback path. Application of the lateral connection was established first by Rosenblatt in 1958 (Haykin, 1994, 1999).

(ii) The cooperative process means that the winner neuron determines the spatial location within the topological connection of neighbourhood neurons, so the winner neuron is able to cooperate with the neighbouring neurons. The weights of the winner and its closest neighbours will also change in the next iterative step. The rate of this weight modification relies on the applied neighbourhood function and on the initialised radius of neighbourhood relationships.

(iii) Based on the *adaptive property*, ANN is able to adequately respond to the changing environment: the learning, training, self-organisation and generalisation and all parts of this property. During the learning process, NN receives positive feedback as reinforcement or discrepancies. ANN can thus respond to the recurring problem or even to new but "familiar" questions. It is ‘self-organisation’ when ANN modifies the weights of neurons according to a kind of learning-rule and this way adapts to the changing environment.

The structure of the applied neural network and the Kohonen layer are presented in Figure 3-4. This figure shows that the processing elements organise in only one layer (Kohonen-layer). Usually it is a plane in 2D, or line in 1D, but 3D or more-dimensional Kohonen layers are also applied. In any case, it is a discrete low-dimensional output space. If the Kohonen layer is a two-dimensional map, the neurons compose a lattice or if it is one-dimensional, the neurons organise nodes in a line. These neurons represent linear projection and each neuron is connected to all input. The neurons compute the weighted input values. In SOM there are feed forward connections, but there is “feedback”. This means only lateral contact among neurons. In Figure 3-4 the structure shows that SOM does not contain any specific output layer. At the same time, each neuron also acts as an output node.

According to the computation of weighted input, the neuron with the largest output will be the ‘winner’. So the neurons compete with each other. The goal of the competition is that in the subsequent step of learning, only the weight of the winning neuron and its neighbours will change. The weights of neighbouring neurons are controlled by the lateral non-linear connection. The winner - the most active neuron - may give positive reinforcement to its neighbours and block the others. The output values of these blocked neurons will be zero.

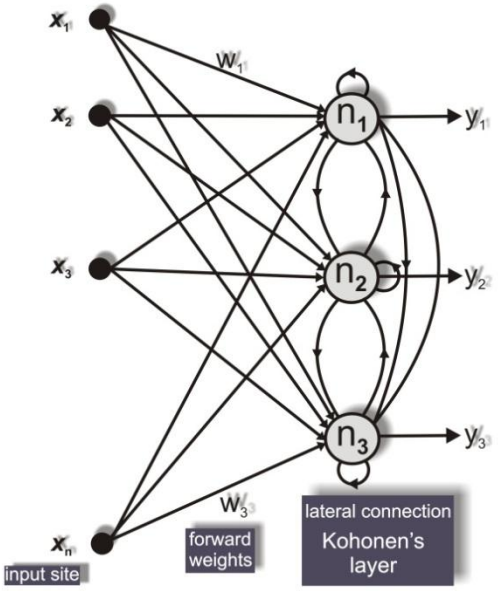


Figure 3-4: Structure of Kohonen neural network with three competing units

This competition-feature supports the rule of the learning. The learning is based on the commonly used Hebb's rule, or more accurately its modified rule.

Learning the data structure takes place iteratively using the *modified Hebb's rule* presented below as the "winning neuron algorithm" (Lampinen et al., 2005).

The modified Hebb's rule in the case of SOM is normalised: it keeps weights within a specified range. The application of the modified Hebb's rule represents this adaptation.

Mathematically, the (i) competition, (ii) cooperation, and (iii) adaptation processes which built up the whole Kohonen methodology are the following:

(i) In the *competitive process* the neurons compute the weighted input (Eq. 3.1):

$$y_i = \sum w_{ij} \cdot x_j = w^T x \qquad \textit{(Eq. 3.1)}$$

$x_j = \{x_{j1}, x_{j2}, x_{j3}, \ldots, x_{jn}\}^T$ input vector which is selected randomly from the data space.

y_i is the computation, value of output for i^{th} neuron.

$w_j = \{w_{j1}, w_{j2}, w_{j3}, \ldots, w_{jn}\}^T$ is j^{th} synaptic weight vector of each neuron with the same dimension as the input data space; ($j=\overline{1\ m}$) where m is the total number of neurons. In other

words, they are the connection weights between the input units i and the neurons j in the computation.

If in a case, $l\ {w_{i*}}^T x > {w_l}^T x$, then the learning rule modifies only the w_{i*}. So UNN selects the neuron with the largest inner product (${w_{i*}}^T x$) through with the winner is determined. The maximisation of the inner products ($w^T x$) is equivalent to minimising the Euclidean distance (Eq. 3.2):

$$i^*(\mathrm{x}) = \arg\min_j \left\| x - w_j \right\|, j = \overline{1\ m}), \qquad \textit{(Eq. 3.2)}$$

$\arg\min_j$ is a distance function; i^* is the index of the neuron "that we want to identify", which is the so-called best-matching or winning neuron for the input vectors x. The weight vector of this winning neuron comes closest to the input vector.

(ii) The update of weights during the learning process depends on the cooperative process. So, in the learning process not only the weight of winner may change. Using the application of neighbouring relationships with neurons the modification of the winner's environment is also possible. This modification is made through the lateral connection of processing elements. This is based on the neighbourhood function and the degree of neighbourhood. During learning, all neurons that are located close to the winner will be activated. This closeness means a pre-defined distance in the Kohonen map.

The development of a lateral connection as inhibition (as used by Kohonen in 1982, and 1984 for UNN) is of biological origin. It has been demonstrated that there are also excitatory and inhibitory lateral connections among the neighbouring neurons in the mammalian brain. In this way, the "parts" of the brain can cooperate with each other to support learning (Kohonen, 1984).

The excitatory and inhibitory lateral connections show that the value of the neuron (which has the largest output value the very first time) increases while the others decrease. This is acceptable if the winner's output is not divergent. It is therefore required that the winner should continuously converge to one, while the

other tends to zeros the binary output is guaranteed using a non-linear projection after the linear combination (Eq. 3.1).

The weights of lateral connections are defined by the distances among the processing elements in a kind of topology. It is first necessary to define this distance and the topological neighbourhood among the activated neurons.

Let h_{ji*} denote the topological neighbourhood (at the centre of the winner) and d_{i*j} mark the lateral distance between the winning element and the neuron being activated by the winner. In this case, the activation is made by Gaussian neighbourhood function (GNF) (Eq. 3.3; Fig. 3-5).

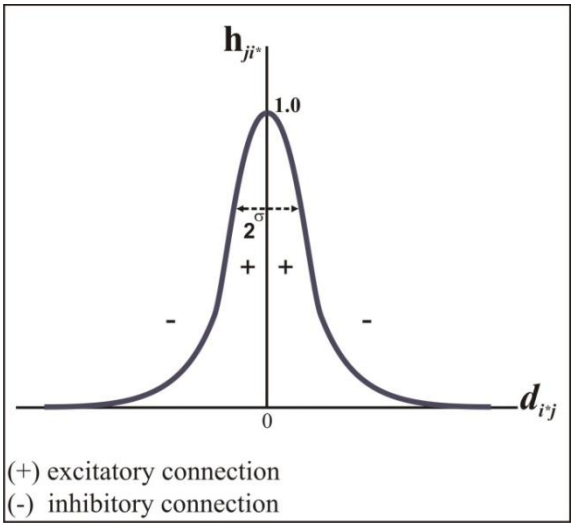


Figure 3-5: Gaussian neighbourhood function

$$h_{ji*}(x) = exp(-\frac{d_{i*j}^2}{2\sigma^2}) \qquad \textit{(Eq. 3.3)}$$

σ parameter is the effective width of the topological neighbourhood (Fig. 3-5).

The role of σ is to define the degree the neighbourhood around the winner. The participation of the excited neuron depends on it in the learning process; this parameter is a continuously decreasing value during the progress of learning.

This GNF must satisfy two conditions:

(1) h_{ji} must be symmetric about the maximum point, where $d_{i*j} = 0$

(2) a necessary requirement for the convergence is that the amplitude of the GNF decreases monotonically with increasing lateral distance d_{i*j} decaying to zero for $d_{i*j} \rightarrow \infty$

If the neurons are organised on a one-dimensional lattice, d_{i*j} equals $|j\text{-}i|$, and in the case of a two-dimensional lattice it is defined by Equation 3.4,

$$d_{i*j}^2 = \|r_j\text{-}r_i\|, \qquad \textit{(Eq. 3.4)}$$

where r is a discrete vector. It denotes the position of the excited neuron in case of index j and the position of winner in the case of index i. It is measured in the output space.

The modification of weights is not applicable when simply using Hebb's rule (Eq. 3.5), because in this way the weights may grow beyond all limits. For the "correct" competition of neurons, UNN must determine the winner using the direction of the weights and not their absolute value. This requires the normalisation of weight and application of the modified Hebb's rule (Kohonen, 1982) (Eq. 3.6).

$$w_{i*}(t+1) = \eta(t) \cdot (x - w_{i*}(t)), \qquad \textit{(Eq. 3.5)}$$

$$w_{i*}(t+1) = w_{i*}(t) + \eta(t) \cdot (x - w_{i*}(t)), \qquad \textit{(Eq. 3.6)}$$

(iii) The updating of weights which supports the learning of UNN is represented by the *updated Hebb's rule* and using GNF. Equation 3.7 shows the weight-updating of close enough neurons (Kohonen, 1982; Ritter et al., 1992; Kohonen, 2001).

$$w_{i*}(t+1) = w_{i*}(t) + \eta(t) \cdot h_{ji*} \cdot (x - w_{i*}(t)), \qquad \textit{(Eq. 3.7)}$$

$w(t)$ - is the weighting and $i*$ index is the mark of the winner processing element;

$\eta(t)$ - is the Kohonen-learning rate.

This learning rate is pre-defined by the start value and the (minimum) end value in the initialisation process. The rate during the training cycle converges from the start value to the end value in a monotonically decreasing way.

4 WORKFLOW OF APPLIED METHODS

4.1 DATA PRE-PROCESSING

A network depends on the quality and quantity of training data. Consequently, pre-processing of data is very important after the definition of the data type (categorical and continuous). This is as follows: (i) selection of outlier data (which can reduce the efficiency of the network); and (ii) attribute normalisation (data pre-processing: data preparation for cluster analysis requires some sort of transformation, such as standardisation or normalisation).

4.2 COMPUTATION PROCEDURE

The applied UNN process was performed using SANN (STATISTICA Automated Neural Networks). SANN includes an inter alia Kohonen-training network, called SOFM (Self Organising Feature Map) networks. SANN is a comprehensive, powerful, and extremely fast neural network data analysis package which is known as a state-of-the-art NN package. The package is "user-friendly" since it is capable of both integrated pre- and post-processing. These processes include data selection, nominal-value encoding, scaling, normalisation and missing value substitution with interpretation for classification problems. The computation procedure is presented in Figure 4-1.

4.2.1 SELECTION OF INPUT DATA (TRAINING, VALIDATION, TEST SETS)

When applied UNN randomly creates three sets: training, validation and test sets. The sizes of the sets are pre-defined. They can be set by percentages of the whole dataset. UNN fits the network model into the training set and selects the model using a validation set, after it determines the stopping point of the learning process. Finally, UNN assesses prediction error using the test set. This test set lends itself to assessing the performance of the trained clusters.

UNN can run without a validation set, but sometimes the test data alone may not be a sufficient condition for a good generalisation. The test error is not applied in the training, but could be used to compare different UNN models. Since this is entirely possible, an extra check of performance is necessary at the end of the training. The model thus applied the validation set to calculate the error. If each error (training, test, and validation error)

converges to minimum and they are similar values, it is reasonable to assume that the network effectively generalises according to unseen information.

Commonly used percentages (size of sets):

- training set: 60-70 per cent of the whole dataset
- test and validation set: 20-20 or 15-15 per cent of the whole dataset

UNN can randomly select samples for the three sets. Another option is that users can randomly prepare pre-processing subsets for UNN, in order to avoid bias.

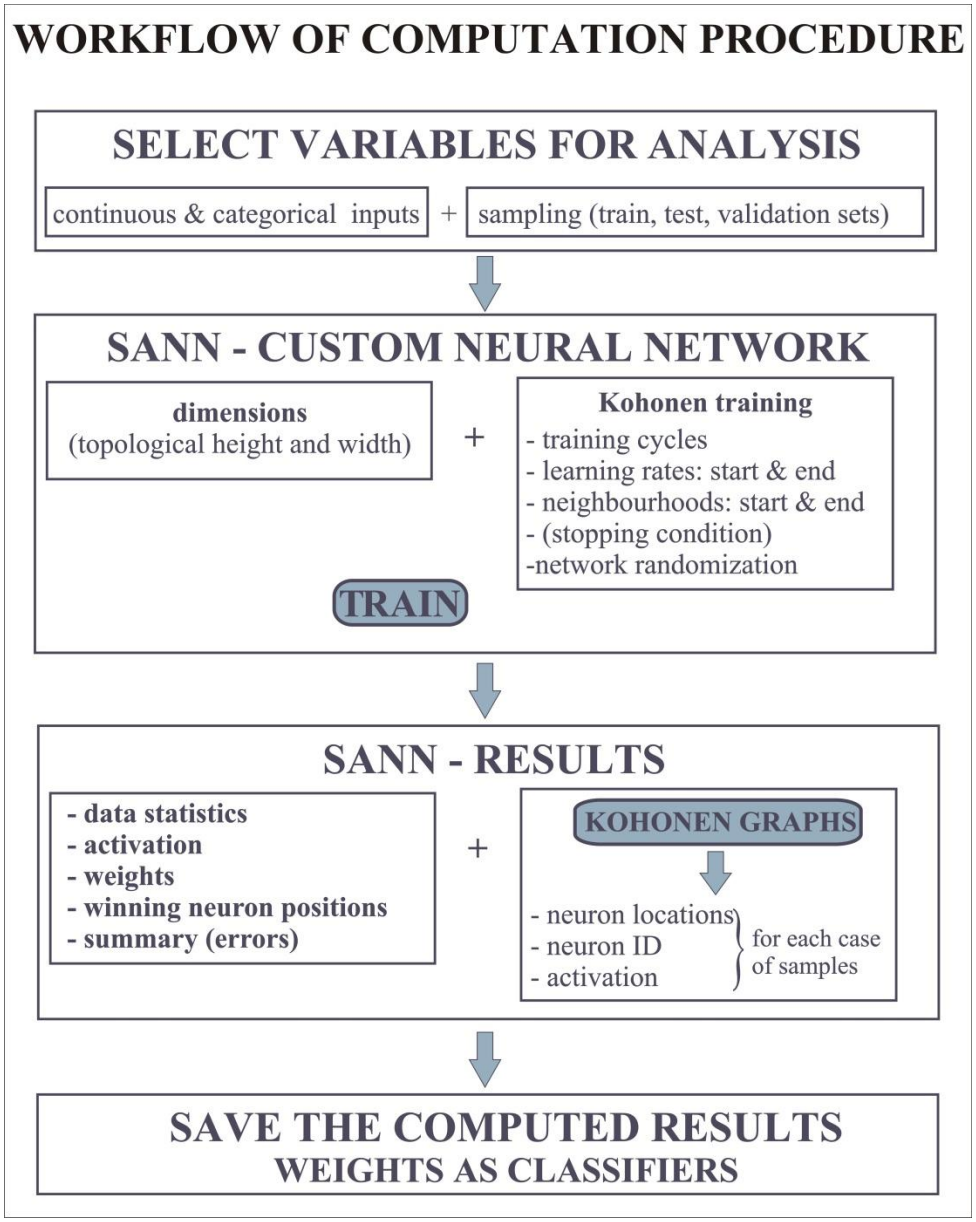


Figure 4-1: The workflow of UNN process

4.2.2 INITIALISINITIALISATION OF NETWOR WEIGHTS, LEARNING RATE AND NEIGHBOURHOOD RADIUS

It is imperative to set initial values for the training rate and the neighbourhood radius. Over the course of the algorithm, the learning rate and the size of the neighbourhood are altered linearly from the first to the last training cycle and gradually reduced. In this way, the Kohonen-learning rate controls the degree of adaptation of the centres to the training cases.

The learning process terminates when there are only small corrections to the weights. This means that the neighbourhood radius reduces to zero and the learning rate reaches a very small value.

The training length is defined by training cycle because the learning is an iterative process. This training stops at the end of the last cycle or it can also be stopped by validation error. Errors in the validation set are monitored during the iterations. An error in the first iteration decreases rapidly. When the reduction in training rate becomes slower a decrease in validation error follows, until the singular extremum. This is called ‘early stopping’ (Fig.4-2), and helps to avoid an overfit of training. When the validation begins to increase, the network has overfit the data.

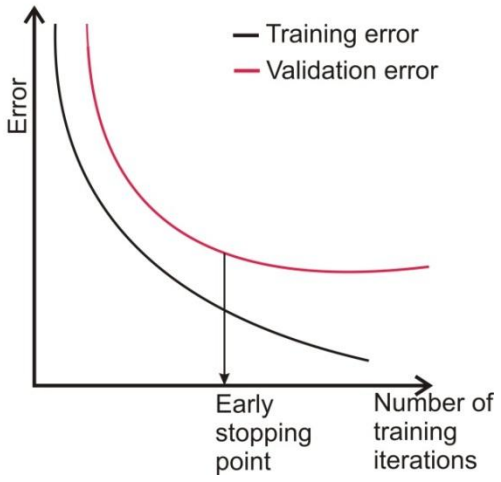


Figure 4-2: Early stopping of UNN learning based on singular extremum of validation error

In addition to problems of overfit, there can be difficulty in predicting new data. Perfect fit results in a zero training error, if indeed possible, but when new data is presented to the network the error is large. The goal of the system is to memorize the structure of the input

data but to avoid generalisation. Generalisation results in errors in new situations, and as a result a more important issue is how to construct UNN which is able to predict new information most effectively. There is a relationship between the overfit of training data and flimsy generalisation. It must therefore be decided which is more important: performance or generalisation.

The SOFM network typically organises the radial layer in two dimensions, but a one dimensional network can also be defined. In this case, the neurons are arranged into only a row. The neighbourhood radius in the radial layer determines the adjacent area centred on the winning unit, and if the Kohonen layer applied is one-dimensional the neighbourhood radius size is one. This way, only the weight of the winner and the weight of the units in the immediate vicinity of the winner change in the next learning cycle. Note that if the neighbourhood radius is initialised to zero, the process will become a simple cluster assignment technique which is can use a Kohonen layer but without neighbourhood definition.

Finally, normal randomisation of weights was used for the training; the mean and variance were specified, and were applied in order to determine the initial weight values.

4.2.3 SAVING COMPUTED WEIGHTS AS CLASSIFIERS OR AS REFERENCE ATTRIBUTES

After the termination of the process, it is possible to save the trained neural network (weights, etc.) for further use. This is advisable because initialisation, finding appropriate parameters, is the most time consuming process of UNN. During initialisation, UNN chooses random values for the initial weight vectors, but the user can also set previous weights according to a saved UNN. This is especially useful if the new dataset is supposed to have similar connections between data points, and in this case it makes it possible to use a supervised network for classification solutions. In this way, the results can be adapted as analogies for similar problems.

Saved computed weights can support the training speed of a neural network and influence the final results through weight initialisation (e.g. Kim and Ra, 1991, in Talaśka and Długosz, 2008).

4.3 STATISTICAL DESCRIPTION AND INTERPRETATION

The clusters formed contain different numbers of elements, and these elements are calculated from the interpreted well log values. In this case the first question is whether the statistics (average, median, etc.) that describe the certain clusters are satisfactory. The different amounts of elements in clusters made it difficult to compare the clusters, and the comparison is not suitable for nonparametric statistics.

In order to resolve the problem of statistical comparison of the clusters, the Monte Carlo simulation was used. Through repetitive sampling the Monte Carlo simulation increased the amount of data and the resolution of the corresponding probability distribution. The simulations retained the characteristic group distributions and properties. In each cluster, the increased dataset reached the number of elements in the measured or interpreted sets of well-log values. These (improved sets and well log values sets) were compared using non-parametric tests.

The Goodman and Kruskal gamma coefficient was used in the comparison to reveal the relationship between two rank-ordered variables. According to the output probability values it is possible to determine the significance level of difference or similarity (Hill and Lewicki, 2005). This is a non-parametric test and the null hypothesis is that the compared samples (samples of created clusters and samples of well log variables) are different. If the test shows similarities the subsequent data analyses and comparison of clusters are based on the extended dataset.

In the characterisation and comparison of clusters several general statistical methods were used. The non-parametric Mann-Whitney test verified that the cluster means were significantly different. These statistical tests were completed by analysis of histograms and box-plots and also through the calculation of within-group and between-group variances.

According to the definition of cluster analysis, the aim is to identify and classify objects based on the similarity of characteristics; another purpose is to create groups which have low variance. Clustering seeks to minimise within-group variance (WGV) and maximise between-group variance (BGV) or at least to create a substantial difference between them. The within cluster variance refers to the spread of objects around the mean and the between cluster variance is a measure of how cluster centroids spread out from one another.

Miller and Khan (1962) showed that, in the case of normal distribution, the total variance was the sum of the within group and between group variances. “If a sample contains *m* groups with the cardinality of n_1, n_2,…, n_m; the members of the j^{th} groups are x^{1j}, x^{2j},…,x^{mj}, then the total variance (TV) of these samples is the sum of the within groups and between groups variances” (Miller and Khan, 1962).

$$\sigma^2 = \frac{1}{SUM}\sum_{i=1}^{m} n_i \cdot \sigma_i^2 + \frac{1}{SUM}\sum_{i=1}^{m} n_i \cdot \left(\bar{x}_i - \bar{X}\right)^2 \qquad \textit{(Eq. 4.3)}$$

The difference in the between group and within group variances expresses the suitability of cluster results. A relatively low WGV and larger BGV means that there are a number of heterogeneous groups with homogeneous contents. A high ratio of BGV and WGV suggests that the clusters are well separated from each other, and there is a high degree of homogeneity within clusters.

Chapter 6 focuses on a comparison of neural network clustering and the widely used K-means technique. In this chapter the comparison of these two methods is based on the results of variance analysis, too. The relatively low WGV demonstrates the efficiency of the two methodologies according to the type of clustering.

4.4 EXTENDING CLUSTERS FOR POINT-LIKE RESULTS USING INDICATOR KRIGING

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 visualisation is also feasible using the Thiessen polygon or Dirichlet tile but an unsampled location is simply allocated to the same category as the nearest observation. Such an inexact interpolation has two weaknesses. It ignores the spatial correlation and probabilities of transition between categories and this does not provide a measure of the reliability of the prediction.

Since membership is a categorical variable its direct mapping (as for a continuous variable) is not possible. The indicator kriging (IK) can offer a reasonable solution for this

task. IK has been designed to estimate uncertainty at unknown points (e.g. at grid points, (Journel 1983, 1986)). To reach this goal it uses a discretized form of the global probability distribution derived from the sample points. The procedure uses several cut-offs for this discretization if the variable is continuous data in the sampled locations. In our case the categories are the discretized form. According to the spatial correlation structure measured between (cut-off) categories, this method estimates the probability distribution function (pdf) at any unsampled locations. These estimates (i.e. probability distributions) change when the number of input samples change, and so it is called conditional (i.e. conditioned to the actual amount of data) probability distribution (e.g. Deutsch and Journel, 1992, 1998, Goovaerts 1997, Olea 1999).

In the framework of GSLIB this technique can be used for both continuous and discrete variables (Deutsch & Journel, 1992, 1998). The latter exactly fits the task of extending the clustering results into geographical space since the clustering of any objects in the geographical space (whether it be UNN or K-means clustering) results in disjunctive sets of spatial points. Their "gridding" process means a lateral estimation of cluster memberships as qualitative objects at each grid node over a domain. This estimation can be than after the definition of as many indicator variables as the number of clusters. The implementation of IK in GSLIB calculates probability for each input category (i.e. for each cluster membership) at every grid node. Note that at any particular grid node, the sum of these probabilities must be one. (The details of the IK process are presented in a flow chart in Appendix 1).

It is clear that a particular grid node should be assigned to those cluster memberships which have the largest estimated probability (e.g. Bierkens and Burrough 1993). Using a simple logical function (Eq. 4.1) we can select every grid node in a cluster that has the greatest probability and these selected clusters can be visualized through, for example, a post map (Fig. 4-3).

$$IF(p_k) = \begin{cases} C_i, & p_k = MAX(p_1, p_2, \ldots p_n) \\ 0, & p_k \neq MAX(p_1, p_2, \ldots p_n) \end{cases} \qquad \textit{(Eq. 4.1)}$$

where

- *IF* is a mathematical logical function (where the function values are 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.

Another choice within the framework of GSLIB is a map showing the change of probability belonging to a particular cluster membership (Fig. 4-3). It is necessary to find a reasonable probability contour to clearly outline the shape of a supposed physiographic unit. The borders of clusters change according to the different probability levels. It is necessary to find a reasonably high (greater than 0.5) probability value which is able to outline the shape of physiographic units.

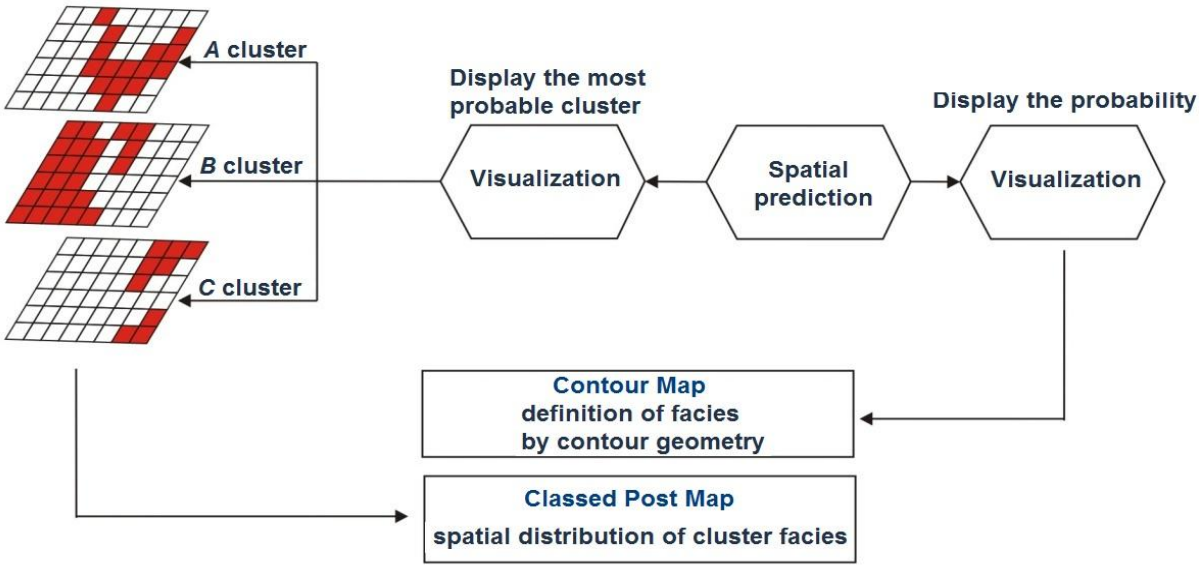


Figure 4-3: Visualization options of cluster members based on spatial prediction (after Bierkens and Burrough, 1993)

It has been shown that different depositional environments can be characterised by special (although, not necessarily different) rock-body morphology (e.g. Moore 1949; Pettijohn et al. 1972). These geometries of depositional facies are the basis of seismic stratigraphy, multiple point simulations, and object-based simulation as a facies modelling technique. Following this line of thought, the rock body geometry expressed by probability contours in this work is interpreted in terms of depositional facies.

The analysis of lateral continuity and spatial variance of cluster facies are also based on these probability contours as border of physiographic units.

The defined and described physiographic units can be used as the training images in object-based simulation. The main steps of a field application include a collection of training images, identification and categorization of lithofacies and depositional facies and analysis of spatial variability. The workflow outlined above can assist in establishing such a set of training images.

5 TWO CASE STUDIES FOR THE PRESENTATION OF APPLIED METHODS

5.1 STUDY-I: ALGYŐ FIELD, SZŐREG-1 RESERVOIR, HUNGARY

5.1.1 SOME GENERAL KNOWLEDGE ABOUT THE FIRST STUDY AREA

The first study area, Algyő Field, is located in the Pannonian basin, south-east Hungary. This field is the largest Hungarian hydrocarbon accumulation, consisting of several sandstone reservoirs (Fig. 5-1). The sediments of these reservoirs accumulated in Lake Pannon in adjacent deltaic and fluvial environments during the late Miocene and earliest Pliocene. This study focuses on one reservoir of this field, Szőreg-1. Stratigraphically it belongs to the upper part of the Újfalui Formation.

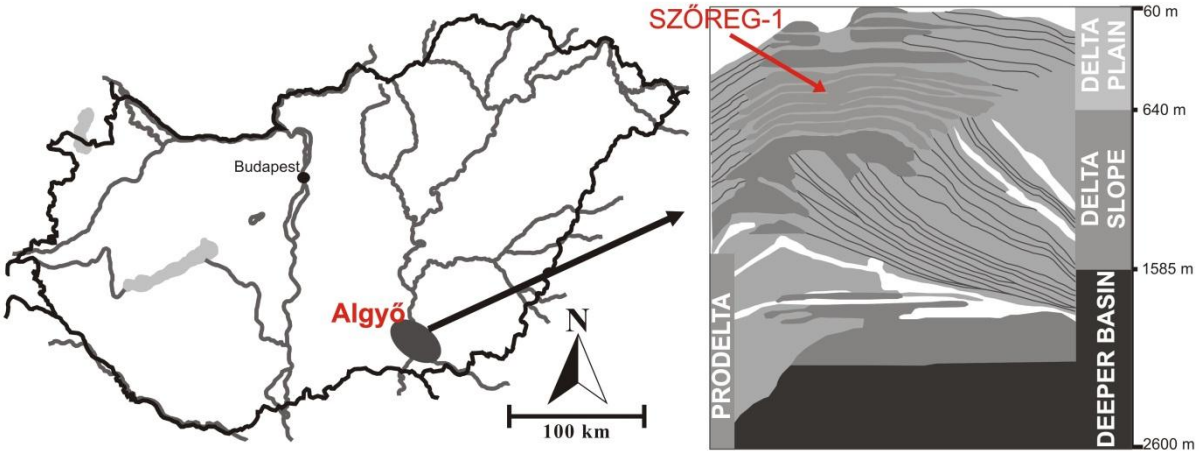


Figure 5-1: Location of study area-1 (Algyő Field, Szőreg-1; after Bérczi, 1988 in Geiger, 2003)

The sedimentation processes of the Algyő field and the Szőreg-1 reservoir are published in several papers in great details. This field has been explored since the 1960s.

Révész (1980) dealt with the deltaic sequences of Algyő Field and showed that it contains the most complete sequence of any general Pannonian (s.l.) basin filling accumulation.

Later, the increased number of production wells made it possible to develop geomathematical and geostatistical models and small-scale approaches to revealing the 3D heterogeneity of this field. Geiger and Komlósi (1996) introduced an application of the 3D

geomathematical system modelling these rock bodies. Using this system, a new geological presentation and a sedimentological model of Algyő Field were created in 1998, and published in 2002 and 2004 by Geiger.

According to these publications, the Algyő field was located in a shallow interdistributary bay during the Late Pannonian period. It can also be assumed that the sources of accumulation were at least two main river/distributary channels. The bay started to fill following the lateral merging of crevasse splays and distributary mouth bar deposits. Because of the periodic abandonments of the river processes, the interdistributary bay sedimentation recurred several times in the area. This processes resulted in depositional cycles. The origin of this delta plane in a significant amount of fluvial channel sedimentation has been demonstrated by Révész (1982), Geiger (2002, 2004), and Sebők-Szilágyi and Geiger (2012).

The present dissertation focuses only on the Szőreg-1 reservoir, in the whole reservoir series (Fig. 5-1). In this work Szőreg-1 is used to demonstrate the workflow outlined in previous chapters and the subjects of the applied UNN are two horizontal surfaces selected from the Szőreg-1 reservoir.

Sebők-Szilágyi (2011), and Sebők-Szilágy and Geiger (2012) presented an improved depositional model of this reservoir with 3D modelling of sub-environments and a detailed description of the sedimentological structures. They analysed this clastic deposit at small scale (cell sizes were 100x100x0.5 m) using a geostatistical method. These works helped to select the two well-known units. These two units belong to the delta plain record, with minor mouth bars developing in front of them, and their lateral accretion infilling the bay area.

The horizontal surface means that the reservoir transformed into a stratigraphic coordinate system where the bottom of the low permeability seal above the reservoir is 0m vertically. In this case the longitude and latitude is the same as in the geographical system but the vertical coordinate was measured from this pre-defined surface. This surface is the almost flat terrain of the deposition and characterised by a massive marl deposit. The average gross thickness of the complete reservoir varies between 35-40m. The rock body in this new coordinate system was cut by lateral surfaces which are parallel with the top (Fig. 5-2). Originally, in the work of Geiger (2004, 2006) and Sebők-Szilágyi and Geiger (2012), the vertical distance between each lateral surface was 0.5m.

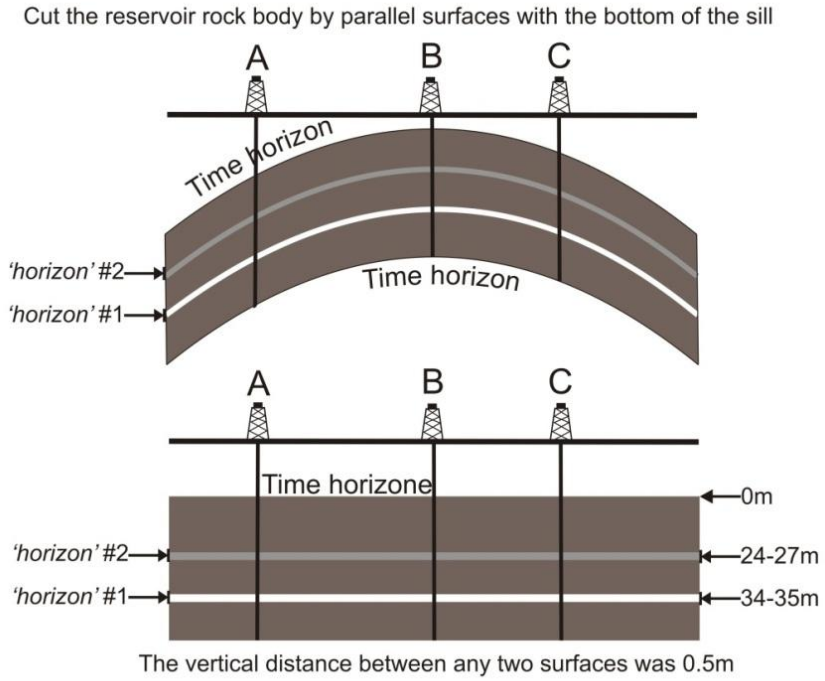


Figure 5-2: The vertical decomposition of the rock body (based on the palinspastic principle; modified after Geiger, 2004)

According to the sedimentation process the selected units indicate distributary mouth bar development (lower surface), and transection of the distributary mouth bar through bifurcation channels (upper surface). The first unit in this work, the distributary mouth bar development, contains a rock body between 34-35m vertically. This is the lower lateral surface of the dissertation (Horizon#1 in Fig. 5-2). The second selected unit locates upper the Horizone#1, between 24-27m vertically (Horizon#2 in Fig. 5-2).

The selection was made based on the fact that these units include different but connected sub-environments in the delta plain area. Their lateral position and character is known from 3D models and detailed descriptions. Thus, it seems to be an appropriate study-area to test the methodology that is the subject of this dissertation.

5.1.2 ROCK TYPES OF SZŐREG-1

The general sequence of Szőreg-1 begins with argillaceous marl units with claystone, marl and coaly argillaceous marl. These rock types can also be seen in higher stratigraphic positions where they are interbedded with lignite units. This succession evidences the

abandonment processes of the previous delta lobes (Révész, 1982; Geiger, 2002, 2004; Sebők-Szilágyi and Geiger, 2012). The grey and dark grey fine siltstones contain a great amount of macrofauna, which in some places even resembles lumashell. This rock type is generally rich in coalified plant fragments. The sediments coarsen gradually from fine to coarse siltstone, but can also be interbedded with thin fine sandstones. Coarse siltstone is usually a grey colour but ochre discolouration is seen frequently. This rock type contains a huge amount of mica and fine sand. Usually it has a laminated structure, with coalified plant fragments. Fine sandstones are characterised by their grey colour. Their laminated structure alternates with coalified plant fragments and mica. Coarse sandstones have a yellowish-grey colour and they can be laminated too, although, the more distinctive features of coarse sandstones are their massive and structureless development, and the high and low angle cross bedding/lamination sedimentary structures. Rock types, with the exception of the underlying argillaceous marl, do not form laterally continuous layers. Laterally, they can be followed in tens or hundreds of metres. Interfingering of the different rock types is a typical characteristic almost everywhere in the studied section.

5.1.3 SOURCE OF DATA WITHIN THE SZŐREG-1 RESERVOIR

Interpreted quantitative petrophysical data was available from 512 wells. The well density is very high in this reservoir (Fig. 5-3). The petrophysical record was measured at every 0.2m intervals. The averages of these records between the selected vertical intervals were used as input data in the UNN clustering. The input variables were porosity, hydraulic conductivity and sand content values. This data derived from quantitative well log interpretation.

The clustering process used hydraulic conductivity as the auxiliary property, which can be in a complex, non-linear relationship with the other two petrophysical features. Despite the relationship between porosity and hydraulic conductivity showing weak correlation in practical reservoir analysis, the relationship between these variables can characterise different depositional facies (Sebők-Szilágyi, 2011). In her dissertation, Sebők-Szilágyi showed sandstones from, for example, the natural levee complex and the channel, described using different correlations of porosity and hydraulic conductivity.

The study assumed that reasonable and real depositional classification can be achieved using the applied parameters, knowing that the porosity is highly dependent on digenesis. It can be presumed that sediments are able to retain more strongly to the depositional characteristics, and diagenetic effects do not modify or only slightly vary the original character.

The lower unit (the emerging distributary mouth bar), penetrated by 209 wells, lies 34-35m below the sill (Fig.5-2). The second (prograded bifurcation channel) is located 24-27m below the top (Fig. 5-2). In this case the dataset came from 344 wells.

The farther up listed variables from both units were used together as input. The cluster process and statistical analysis was not separated for these two horizontal surfaces, so UNN input data was based on all variables from both units. This parallel analysis supports that the progress of cluster facies (development, lateral extension, progradation, etc.) can be followed through two lateral surfaces. The goal is to use the same depositional facies to characterise and define any cluster that appears in both horizontal surfaces. All data was thus imported simultaneously to the UNN, to separate subsets.

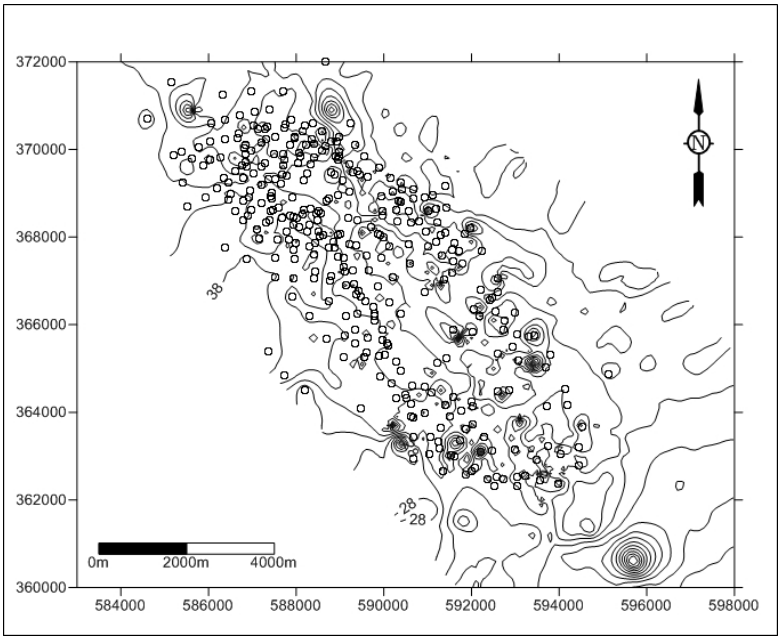


Figure 5-3: Well density of Szőreg-1 reservoir with gross thickness contour

5.1.4 APPLIED METHOD AND INTERPRETATION OF RESULTS FOR THE SZŐREG-1 RESERVOIR

5.1.4.1 Settings of UNN and generated clusters

After the import of input data into the spreadsheet, the size of training set was fixed as 60% of all data points. For the validation and testing, 40% of the whole set was used, evenly divided. These three subsets were collected by the network in a random way to avoid bias. The training set is used by SOM to build a neural network. The validation set is applied to tune the parameters of a classifier and to determine the end of the learning process. The test set lends itself to assessing the performance of the trained clusters.

The dataset was not previously normalised. It was applied within UNN as an additional option.

The number of cluster was pre-fixed for separation. During the final clustering, UNN separated six clusters. Usually it is hard to determine the appropriate number of clusters. This number depends on the user. In this case the number of clusters (six) was equal to the number of depositional facies which had been proved. (In the first few experiments segmentation was attempted using both less and more than 6 clusters, but these gave misclassification, or clusters which were difficult to interpret geologically).

The variance analysis and the comparison of clusters by non-parametric test can help to decide the number of separated clusters, but only subsequently.

The Kohonen-learning rate converges monotonically in the [0,1] interval from the first to the last training cycle. The start value was specified as 0.5 and 0.03 for the end value. The neighbourhood radius designates the adjacent area centred on the winning unit. In the constructed UNN the size of radius was only 1 because of the low cluster number, and it specified a 2X3 square.

Normal randomisation of weights was applied in the training; the mean and variance were specified and are used to draw the initial weight values.

Another important parameter is the learning cycle. UNN learns the relationships of data in each cycle. This process stops at the end of the last cycle or when the test error starts to increase (Fig. 4-2). In this case, the number of cycles was 5000. It is also possible to stop the network process early. This is when a test error breaks the run with increasing convergence. Early stopping is also used if the learning cycle is hard to define. It is certain that after a 4-5000 cycle the results of training did not change because the test error stopped learning. Consequently, since the main objective was, with the constructed UNN, to reach

close validation and low training with the best possible initialisations and settings, the stopping condition was also applied.

According to the constructed UNN (Fig. 5-3) the input is organised into clusters controlled by the neurons. Under the above conditions the input dataset is divided into six clusters. These separated clusters comprise the data points from both horizontal surfaces. In this manner the clusters can be followed in both horizontal surfaces, for example if they changed spatially or if any cluster belonged to only one surface. The sematic construction of UNN, and the cluster averages are shown in Figure 5-4.

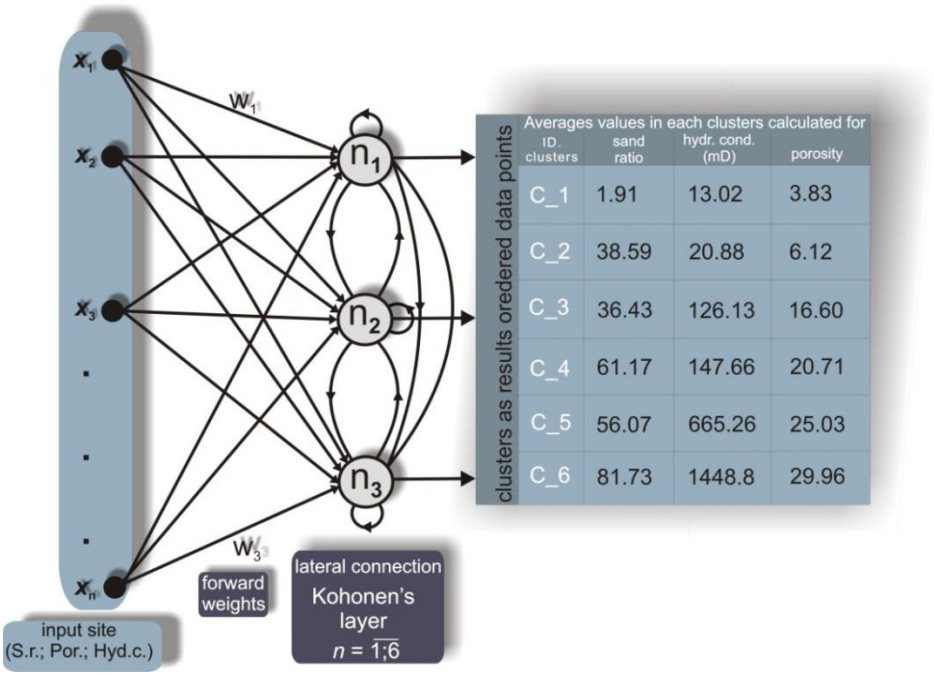


Figure 5-4: Sematic graph of constructed UNN with input and output in Szőreg-1 Reservoir

5.1.4.2 Statistical comparison of clusters

The separated clusters are of different sizes. Figure 5-5 shows the number of points in each cluster in Horizon #1 and Horizon #2. Figure 5-6 shows that Cluster C_1 appears only in the lower horizon. Conversely, the lower horizon contains fairly few data points from C_5 and C_6 clusters. These clusters dominate in the upper horizon (Fig. 5-6).

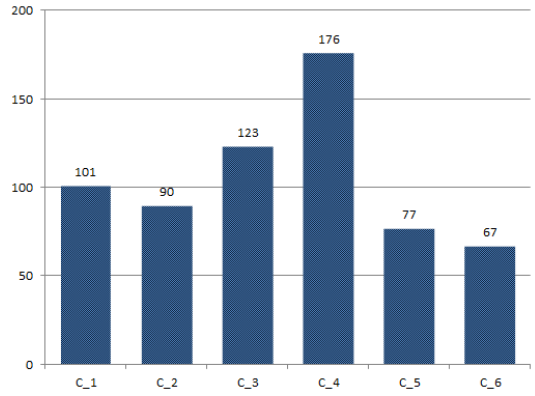


Figure 5-5: Frequency histogram of elements of clusters on both horizontal surfaces in the Szőreg-1 Reservoir

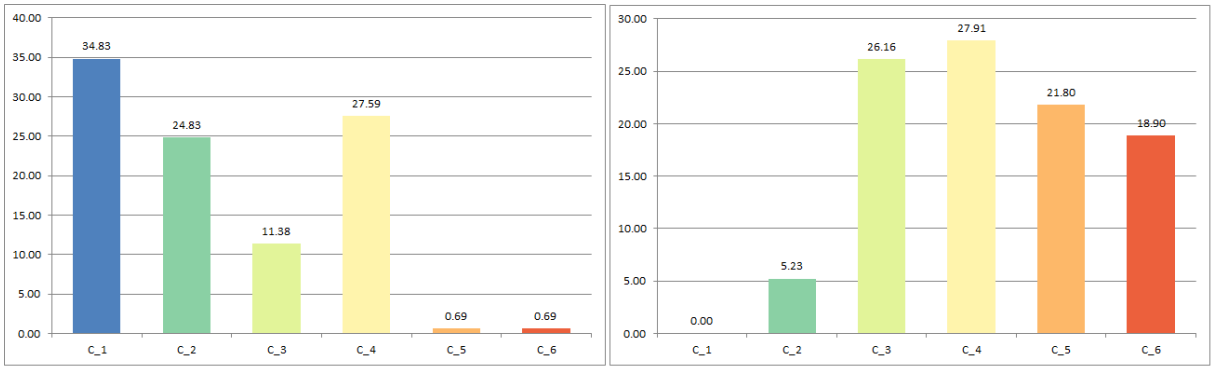


Figure 5-6: Frequency histograms of elements of clusters in certain horizontal surfaces in the Szőreg-1 reservoir (left part represents the frequencies from first surface and the right part from the second surface)

For stable statistical comparisons it is necessary to increase the size of clusters in order to retain the character and distribution of clusters. This was the reason for application of Monte Carlo simulation. In this way the cluster sizes were increased to 650, which is the amount of all input data.

These simulated sets were compared with the original well log-values. Goodman and Kruskal's gamma coefficient was applied to check the significance of similarity. The result was that sets came from the same distribution at the 0.05 significance level. Consequently, the sets of averages were suitable for describing clusters.

The comparison of particular clusters was based on non-parametric statistics (Mann-Whitney test) and additional graphical statistics (frequency histograms and box plots).

Cluster C_1, appearing in one rock horizon, is characterised predominantly by a low porosity and sand ratio (Fig. 5.4). This cluster with low sand content and porosity values can be described as deposits from still water in the interdistributary bay.

According to the histograms for Clusters C_2 and C_3, these subsets have similar pdf in sand content and hydraulic conductivity values. Both have a polymodal histogram in sand content (Fig. 5-7, first column), but the box plot of these properties and the porosity values segregate them significantly. In the case of C_3 this plot shows that the samples are described by moderate negative skewness, which means that the distribution concentrated on the larger sand content (about 40%). For C_2 and C_3, the porosity values mostly confirm the existence of an isolated group and this separation was confirmed by the Mann-Whitney test.

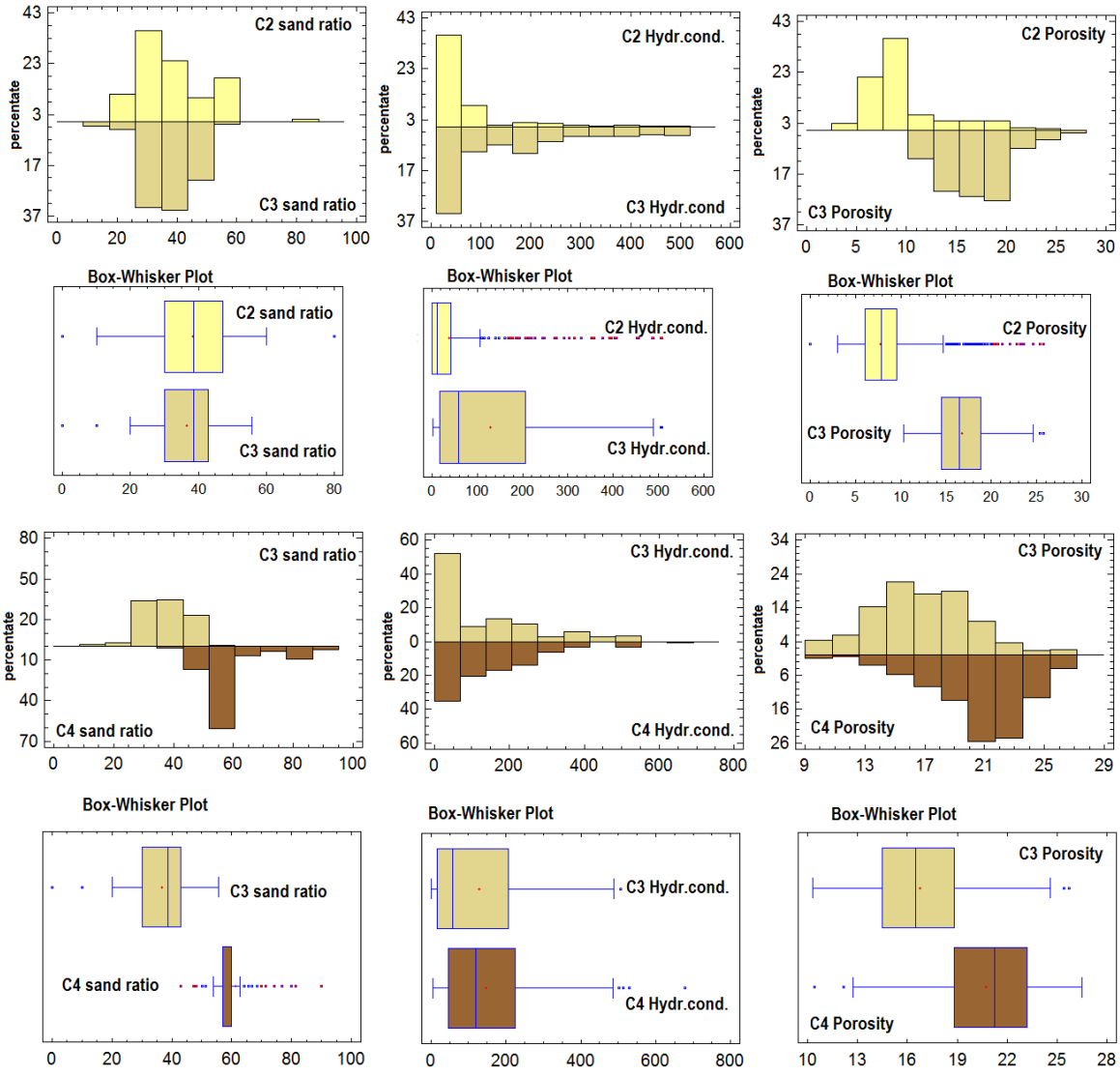


Figure 5-7: Statistical comparison of clusters C_2, C_3 and C_4 (Szőreg-1 Reservoir)

According to the cluster means, Cluster C_4 differs from the others, although hydraulic conductivity has the same pdf in C_3 and C_4 (Fig.5-7, middle column). Sand content can be differentiating between these groups (Fig. 5-7, left column). The Mann-Whiney test also shows this similarity. According to the non-parametric test these clusters are not significantly different based on the median of hydraulic conductivity. As a result of the auxiliary variable, hydraulic conductivity was analysed in relation to porosity.

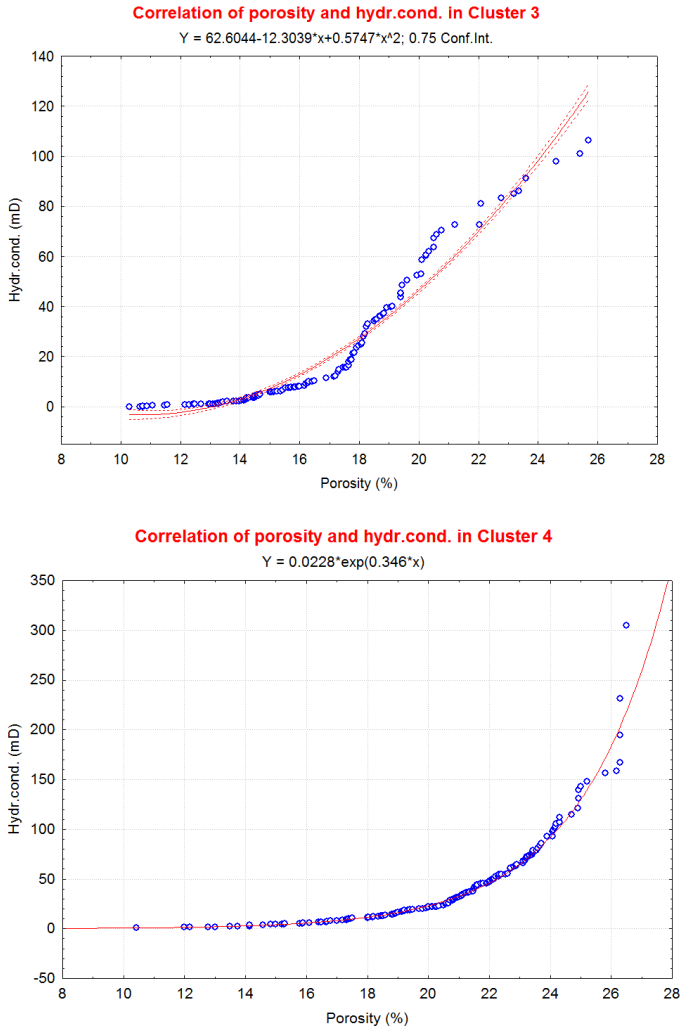


Figure 5-8: Comparison of connection between hydraulic conductivity and porosity in clusters C_3 (upper plot) and C_4 (lower plot)

The relationship between porosity and hydraulic conductivity was compared in these two clusters because sometimes the database contains clusters which cannot be separated by any hyperplane. This usually means that the variables are sometimes similar but that the

connection between the variables is non-linear. In this case the cluster member connects in an interfingering way, or like chains in the property space.

In the comparison of the relationship of porosity and hydraulic conductivity in Clusters C_3 and C_4, we can see that both clusters are described by different correlations (Fig. 5-8). Cluster 4 includes only one population of deposition which is described more or less by exponential Q-Q plot. Conversely, Cluster 3 shows weaker correlation and it can be approximated by polynomial function. According to this Q-Q plot, two populations belong to this depositional cluster facies. Of course this analysis cannot reveal any exact correlation function between these properties, but it can be said that both cluster facies characterised different complex relationships between the porosity and hydraulic conductivity.

Since C_3 and C_4 contain more elements in Horizon#2, it can be suggested that sediments with medium porosity and relatively higher sand content are more widespread in the upper rock unit.

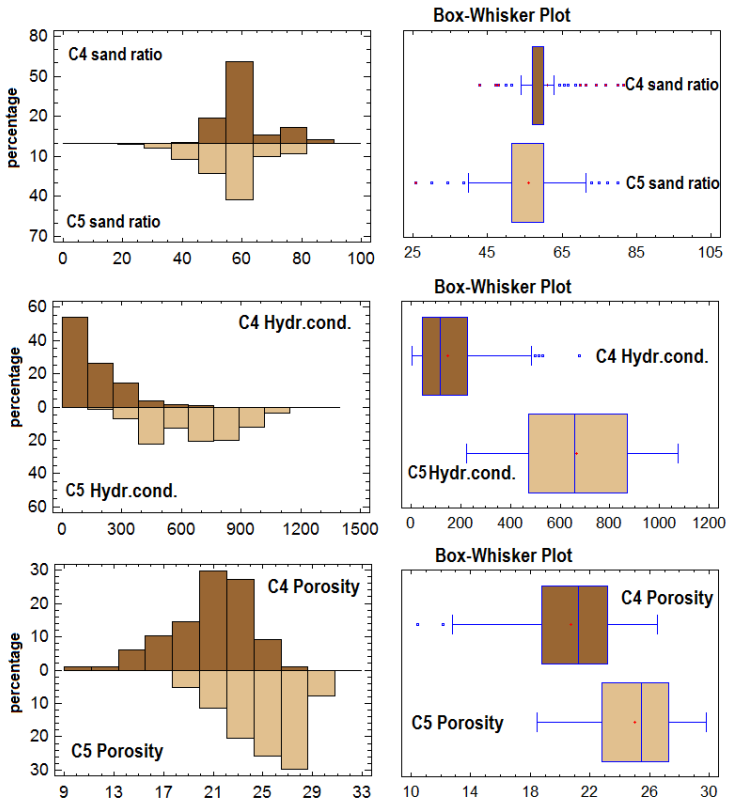


Figure 5-9: Statistical comparison of clusters C_4 and C_5 (Szőreg-1 Reservoir)

C_5 and C_6 clusters dominant only in Horizon#2 (Fig. 5-6). C_5 and C_6 are characterised by a highest porosity value and very large sand ratio. These differ significantly from the other four clusters. The pdf of sand content shows a relationship between C_4 and C_5 (Fig. 5-9), but other statistical character distinguishes C_5.

The question is whether they are two significantly separated cluster facies or not. C_5 seems to be relatively close to C_6 based on the pdf of sand content (Fig. 5-10, first column), but nevertheless, based on the other two variables and the non-parametric test, the separation is confirmed. The Mann-Whiney test supports the difference at a 95% confidence level. The hydraulic conductivity and sand ratio also confirm the existence of an isolated group in the cases of C_5 and C_6. The non-parametric tests also support this differentiation; p-values converge to zero below the 95.0% confidence level.

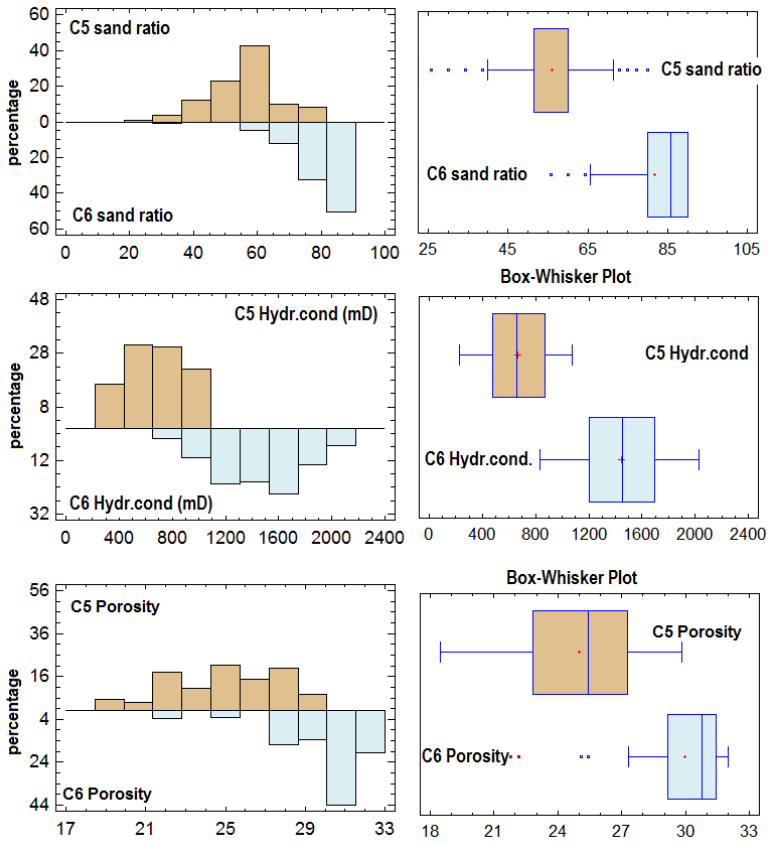


Figure 5-10: Statistical comparison of clusters C_5 and C_6 (Szőreg-1 reservoir)

These two clusters (C_5 and C_6) are dominant only in Horizon#2 which lateral surface was selected as a unit of bifurcation channel system. It can therefore be presumed that

these two cluster facies both belong to this system. Considering the sand content and porosity, C_6 may be defined as the channel fill deposit of a bifurcated channel and C_5 as the levee of this system. But could these be clearly separated from each other? These facies spatially interfinger and their characteristics are also similar.

If this assumption is right, this complex is described by a polymodal pdf in porosity and sand content values. It is supported according to the work of Sebők-Szilágyi and Geiger (2012). The distribution of sand content is characterised by 60, 80% largest frequency, and 22, 30% frequencies describe the distribution of porosity (Fig. 5-11).

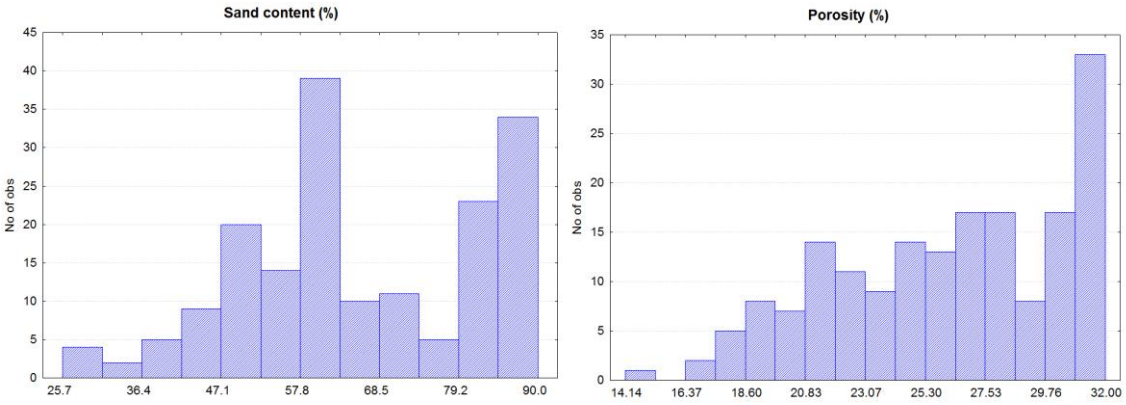


Figure 5-11: Frequency histogram of sand content (left) and porosity variable (right) in C_5 and C_6

Data separation seeks to minimise WGV and maximise BGV to collect mostly homogeneous objects. The previous non-parametric tests analyse the separated groups if they are significantly different, if the subsets are heterogeneous. This means that the measures for spreading the cluster centroids are far enough from each other in the property space. This can be expressed as the value of BGV and WGV shows the homogeneity. This variance analysis is also able to validate the number or clusters. This variance was calculated based on Equation 4.3.

Table 5-1 shows WGV and BGV and their related parameters for the six defined clusters. It suggests that the within group variance is much lower than the between group variance. Consequently, UNN formed clusters which show more homogeneity inside than that can be measured between them.

For all six clusters the parameters in Equation 4.3 are the following:

$m = \overline{1,6}$

n = 634

n_1=101; n_2=90; n_3=123; n_4 = 176, n_5 = 77 and n_6 = 67 (n_i comes from frequency histogram in Fig. 5-5).

Table 5-1: Calculated within group variance and between group variance for all six clusters

	WGV	BGV	TV	WGV / TV rate (%)	BGV / TV rate (%)
POR	9.12	96.61	105.73	8.63	91.37
S.r.	1.67	13.70	15.37	10.84	89.16
HK	3783.04	31277.00	35060.04	10.79	89.21

According to the variance analysis the six separated clusters seem to be very heterogeneous, with homogeneous content, because WGV is only about one tenth of the total variance (Table 5-1). This analysis also shows that clusters are divided with similar variance based on all properties. This means that all variables play a similar role in the clustering.

5.1.4.3 Probability of spatial extension of clusters

The extension and the spatial display of clusters have an important role in the interpretation of the UNN results. The lateral extension and pattern of the statistically described cluster facies can assist to define the facies sedimentologically.

In Szőreg-1 Reservoir the shape and main properties of the facies have been reported by Geiger (2003). The following geometrical characters were described:

(i) The distributary mouth bars have kidney-shape or lunar-shaped geometry and in this part the sand content varies between 40-80%

(ii) The distributary channels have elongated geometry, and the boundaries are defined by approximately 40%, but the inner sand content can reach 70-80%

(iii) The interdistributary bay surrounds the elliptical distributary mouth bars, since the sediment is from quiet water; it contains few sandy deposits, and is characterised by low porosity.

In this dissertation the pattern of facies was also analysed using extended cluster facies. This extension used IK and was applied for both lateral surfaces.

5.1.4.4 Preparation of input data for IK process

In practice, IK leads to some computation problems, such as in Horizon#1 where Clusters C_5 and C_6 occur in isolated locations (Fig. 5-12). In such a case it is difficult to fit a variogram for less frequent classes. These isolated locations usually characterise the nugget effect model that involves random spatial arrangement, since these data points are omitted in the IK process.

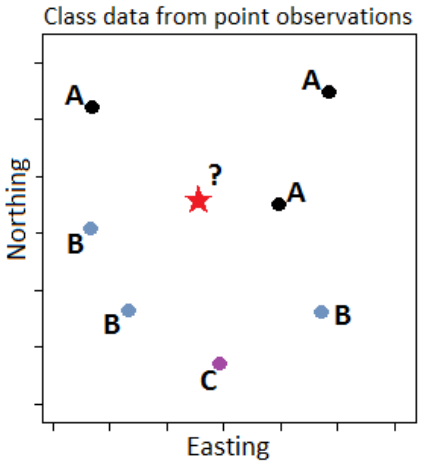


Figure 5-12: Less frequent clusters with isolated location

At first, the variogram surfaces were analysed for all clusters (Fig. 5-13). The last two variogram surfaces in this figure (exemplified by numbers 5 and 6) involved the nugget effect model.

In the second horizontal surface only five clusters were elements. The variogram surfaces of these five clusters were appropriate for use in modelling. Figure 5-14 shows the variogram surfaces of this horizon (labelled with 2-6 values according to the cluster marker).

In the calculation of directional semivariograms, the lag spacing was the half of the average well distance, and the angle tolerance was 22.5 degrees. Of the set of semivariograms, calculated for every tenth degree in a counter clockwise direction, only two were retained. They lay in the direction of the longest and shortest ranges. The

omnidirectional semivariogram was added for modelling purposes. The settings and the final models are shown in Appendices 2 - 10.

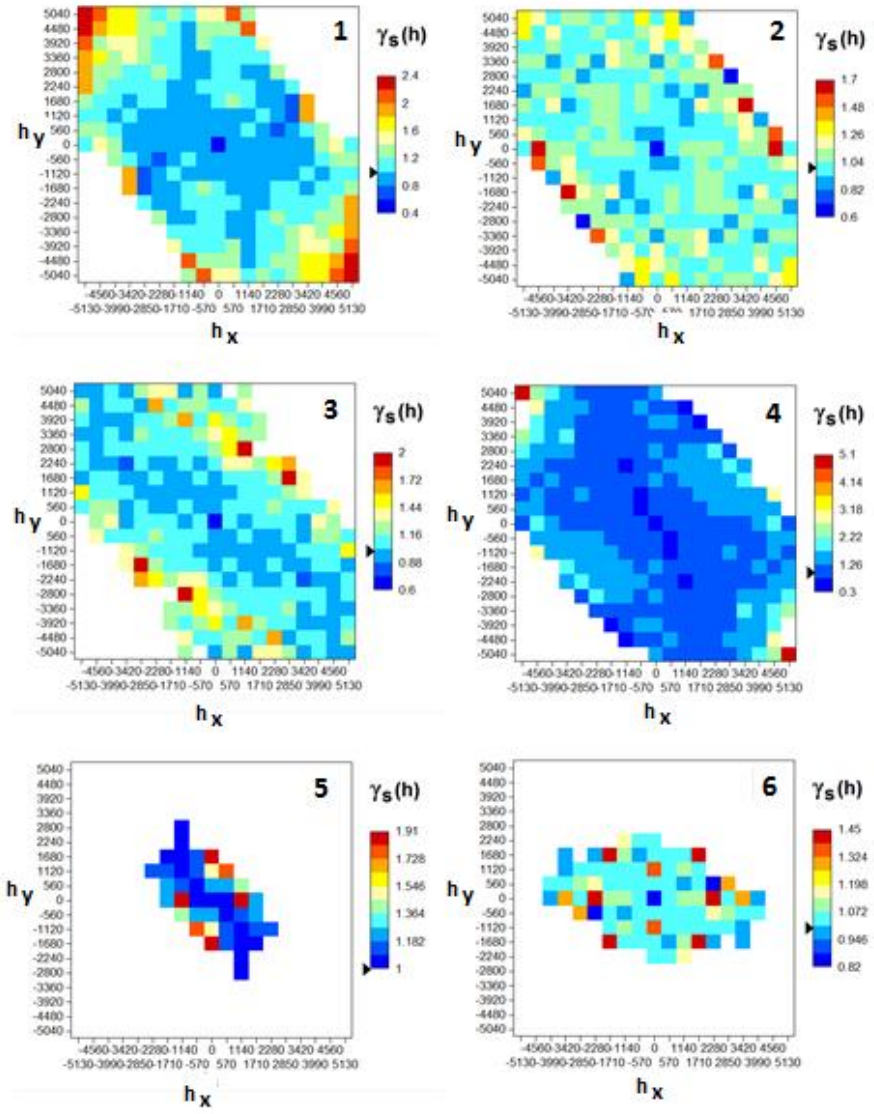


Figure 5-13: Variogram surfaces of all six clusters in Horizon#1, at Szőreg-1 Reservoir

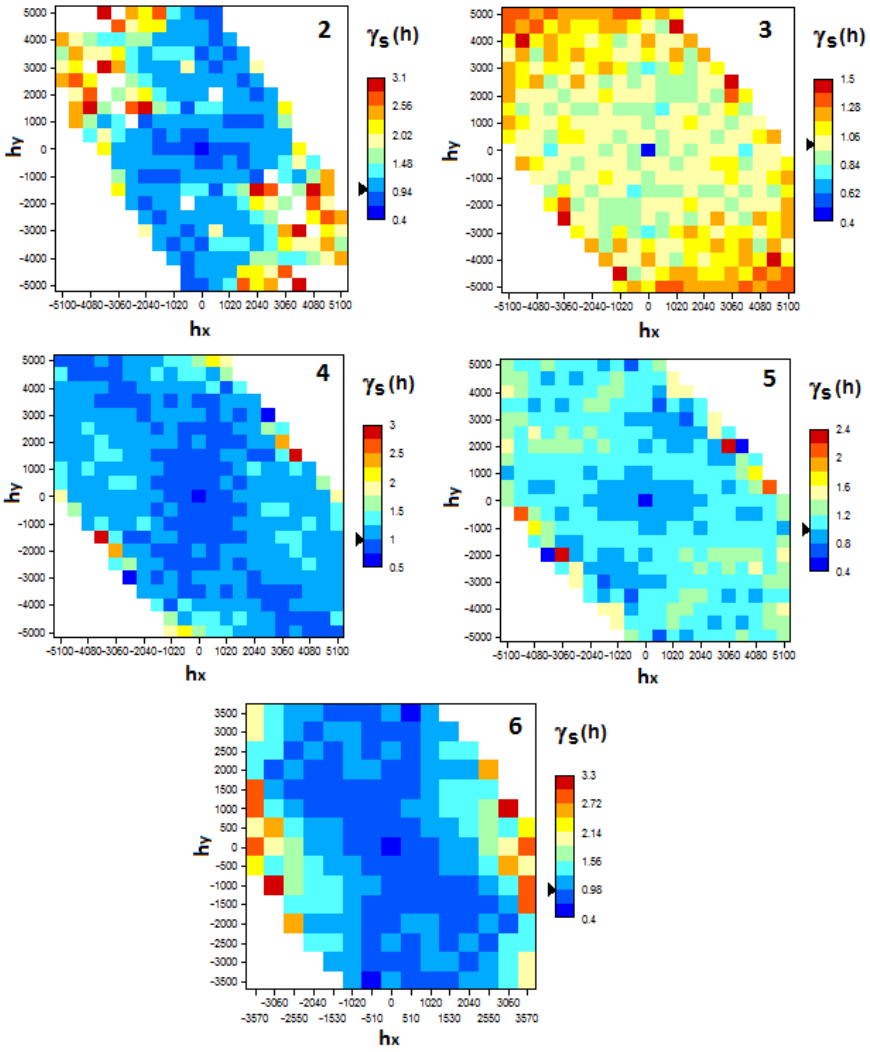


Figure 5-14: Variogram surfaces of five clusters in Horizon#2, at Szőreg-1 Reservoir

During the modelling process some permissible theoretical variograms were fitted to the experimental ones. All models contained two or three structures which can be described using several parameters, such as:

- type of structure in model (type)
- maximum range (hMax)
- minimum range (hMin)
- sill (cc, where the amount of cc value of structures and nugget value equal with 1)
- direction (ang.)
- anisotropy (anis.)
- nugget (ng.)

Table 5-2: The model parameter for probability estimation by IK (horizon#1 in Szőreg-1)

cluster code	structure	type of variogram	hMax	hMin	cc	ang.	anis.	hVert.	ng.
C_1	**first**	spherical	390.693	244.183	0.52	68	1.6	1	0.42
	second	exponential	6900.693	2208.22	0.06	281	0.32	1	
C_2	**first**	spherical	620.539	365.022	0.31	82	1.7	1	0.6
	second	spherical	896.770	320.275	0.09	54	2.8	1	
C_3	**first**	spherical	551.539	172.355	0.24	273	3.2	1	0.71
	second	spherical	344.079	92.990	0.05	299	3.7	1	
C_4	**first**	exponential	781.616	304.830	0.65	315	0.39	1	0.3
	second	exponential	1379.690	317.328	0.05	259	0.23	1	

These parameters are summarised in Tables 5-2 and 5-3. It can be seen that the clusters have a distinct spatial structure with different ranges.

Table 5-3: The model parameter for probability estimation by IK (Horizon#2 at Szőreg-1).

cluster code	structure	type of variogram	hMax	hMin	cc	ang.	anis.	hVert.	ng.
C_2	**first**	exponential	346.59	203.079	0.52	62	1.7	1	0.4
	second	spherical	2080.00	717.446	0.08	9.5	2.9	1	
C_3	**first**	exponential	216.66	127.45	0.55	350	1.7	1	0.4
	second	exponential	520.00	118.18	0.05	10	4.4	1	
C_4	**first**	spherical	910	782.60	0.4	60	0.86	1	0.5
	second	exponential	801.66	364.39	0.1	275	2.2	1	
C_5	**first**	exponential	330.27	181.65	0.5	340	0.55	1	0.45
	second	spherical	2275	947.92	0.05	4.5	2.4	1	
C_6	**first**	exponential	353.67	160.76	0.37	16.89	2.2	1	0.6
	second	spherical	2525.466	7879.21	0.03	210	3.2	1	

The results of IK are shown in Figure 5-15. On this map, the cluster was assigned to a particular grid node which had the largest probability of appearance (Equation 4.1). This figure also shows the facies identification of clusters (Fig. 5-15).

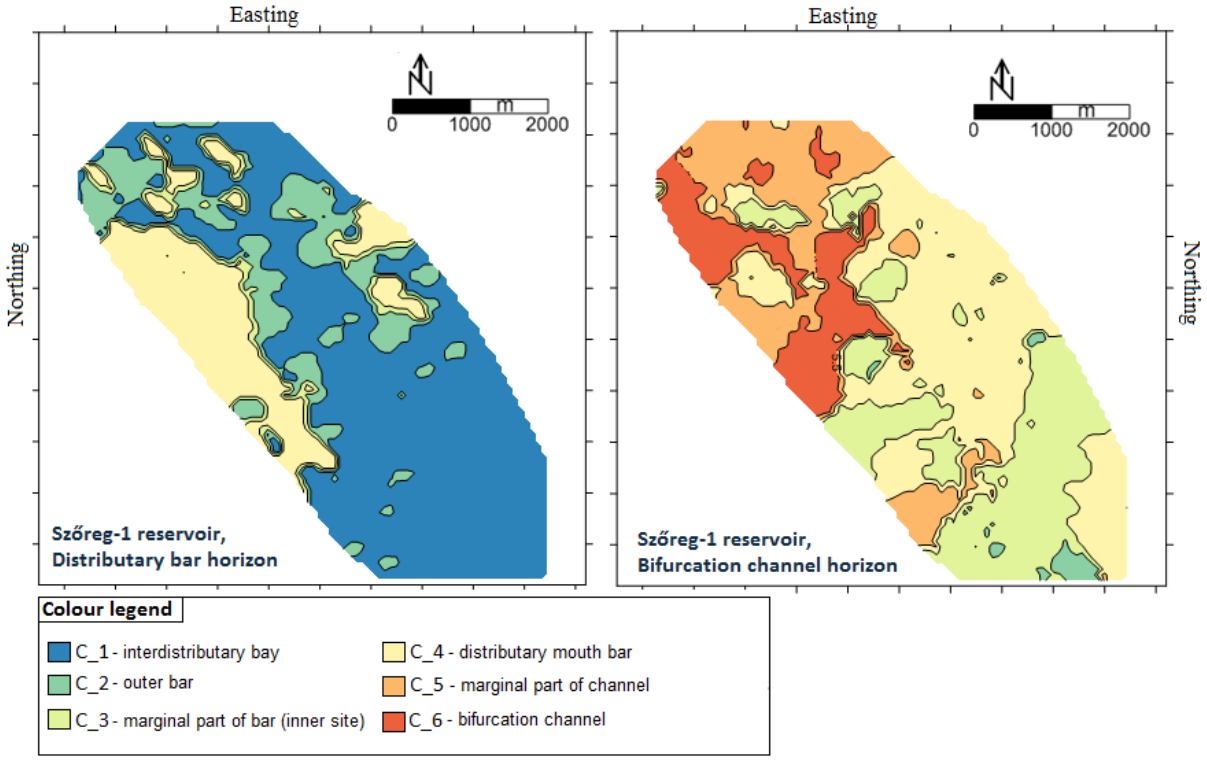


Figure 5-15: The results map with laterally extended and identified clusters at the Szőreg-1 reservoir (The laterally extended clusters displayed by the larges probability values which estimated by IK)

5.1.5 INTERPRETATION OF LATERALLY EXTENDED CLUSTERS AND CONCLUSIONS AT SZŐREG-1

C_1 represents the inter-distributary bay, which is predominantly characterised by a low porosity and sand ratio. This facies surrounds the complex facies of C_2, C_3 and C_4 clusters. C_2, C_3 and C_4 are thus defined as cohesive groups. They are part of a major environment which has a kidney shaped geometry due to lateral extension based on the probabilities. This kidney shaped pattern characterises a distributary mouth bar system. This sedimentary facies has a very complex structure. C_2 interfingers directly with the interdistributary bay where the porosity and the hydraulic conductivity are very low but the sand ratio increases in comparison with bay sediment.

The main part of the mouth bar is characterised by C_4. This facies is characterised by the largest sand content in the lower horizontal surface. It could be defined as the bar crest of the mouth bar deposit. This facies developed from two opposite directions (SW and NE). C_3 also interfingers with C_4. It encloses the body of the distributary mouth bar as a margin. C_3

and C_4 are similar, as shown by statistical analysis, however, they can also be defined as two separated facies. C_2 comprises outer bars, and is a disturbed (due to sliding and slumping) region facing the bay. C_3 represent the marginal parts of distributary mouth bars. This fringe of the mouth bar body is a narrow area in the lower horizontal surface. It is a kind of transition zone with lower sand content and porosity values than the main area of mouth bar.

C_4 is the body of the bars which are involved from two directions. The sedimentation of C_3 and C_4 is controlled by the bifurcated channel system in the hinterland area. The shapes of lateral extensions changed greatly during the development of this area. These changes can be seen in differences between the maps of the two lateral surfaces.

The main difference between C_3 and C_4 is the bimodal porosity distribution of C_3. The genetic background of C_3 is a fact. C_3 was defined as the marginal part of a mouth bar. Here, the bay sedimentation interacts with the accumulation of channel. The resulting silty sand has very bad sorting and quite poor porosity.

The last two clusters appear only in the second, so-called horizontal surface of the bifurcation channel. This is the hinterland area of the developing major mouth bar. C_6 is characterised by the bifurcation channel deposit with a dendroid or finger-like network geometry. This channel cuts to the body of mouth bar area. It developed from the SW direction. C_5 facies followed the facies of this bifurcated channel. C_5 is defined as the marginal part or levee of the distributary channel. These latter two cluster facies together belong to a complex system which is a channel network system with channel fill deposit and natural channel levee.

The sorting of the channel levee deposit is weaker (Fig. 5-10) than in the channel deposit characterised by C_6 due to the porosity values shown in the polymodal histogram and the lower sand ratio.

The size of the two main facies can be determined on the basis of the lateral extension of the clusters. Let’s focus on the two main facies which are represented by Clusters C_4 and C_6. Using the 0.6 probability contour to display their geometry, the shape of C_6 is a channel network with a length of approximately 5km and is around 500m in width and the pattern of C_4 is a 6.5-7km width kidney-shaped bar. The progradation of this major bar might occur from the SW. From the opposite side a minor mouth bar developed with a maximum 3-3.5km width.

These identified facies with their lateral extensions and main directions may reflect two main phases of the depositional history of the Szőreg-1 reservoir, as introduced by the studies of Geiger (2003) or Sebők-Szilágyi and Geiger (2012). Figure 5-16 summarises the phases of temporal development for Szőreg-1 Reservoir. The lateral surfaces selected and represented in the present dissertation correspond to Phases C and D-E in Figure 5-16. The applied approach of facies analysis could reveal the same sub-environments. The map of results from IK (Fig. 5-15) for Horizon#1 shows Phase C when discrete major and minor mouth bars formed from the SW and NE directions (Fig. 5-16, C). The map of results from IK (Fig. 5-15) in Horizon#2 presents Phases D-E when bar bodies were involved due to the progradation and bifurcation channel that broke it up.

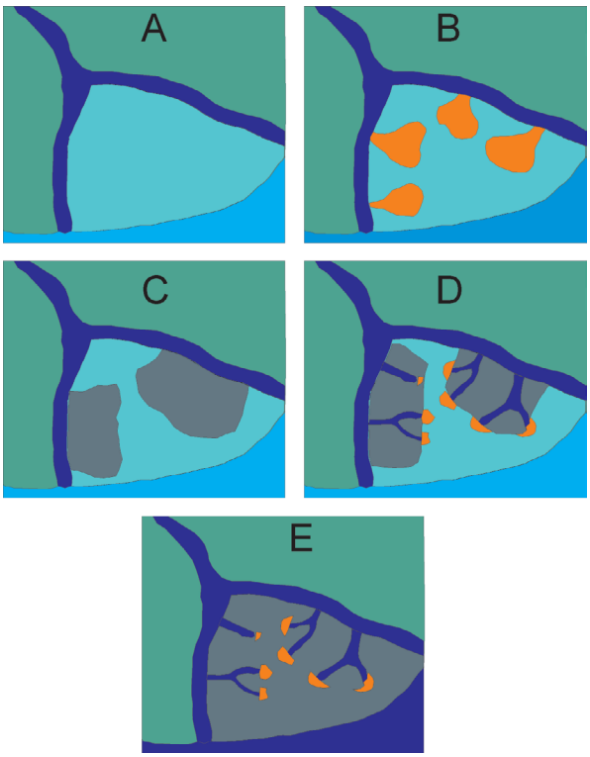


Figure 5-16: Depositional history of Szőreg-1 Reservoir (Sebők-Szilágyi and Geiger, 2012)

5.2 STUDY-II: SAVA FIELD, SAVA BASIN, CROATIA

5.2.1 SOME GENERAL KNOWLEDGE ABOUT THE SECOND STUDY AREA

The study area is a hydrocarbon field situated in the north-eastern part of the Sava Depression, Northern Croatia, about 35km east of Zagreb (Fig. 5-17.). This field is called the Sava Field through the rest of this study.

The first exploration activities began in the early 1940s and 'Sava Field' was discovered in 1963. This field contains 11 hydrocarbon reservoir-units with oil, dissolved gas and gas, in a gas cap (Hernitz et al., 1996).

A total of 87 wells have been drilled in the field so far. There are 42 in production, 13 monitoring, 14 are water injection wells, and 18 have been abandoned (Report[2]). In this case study the input data was derived only from 78 wells (their data was suitable for data pre-processing). 3D seismic survey of the field was carried out at the end of 1998 within the scope of the 3D Sava-1 extension project (Report[2]).

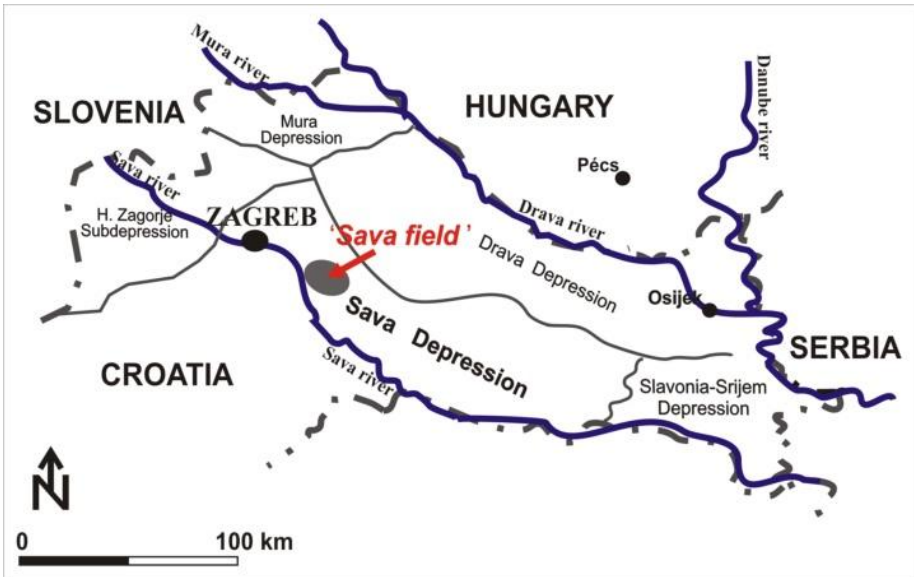


Figure 5-17: Location of Study Area 2 (Sava Field)

According to the genetic stratigraphic sequence concept (Galloway, 1989 in Hernitz et al., 1996) 8 depositional events were originally identified. After reambulation, the number of

[2] Unpublished report about the study field in Sava Depression - Tertiary CO2 Injection (2003): INA, Zagreb

these events increased to 11. The main reservoir rocks are fine grained sandstones with a significant amount of quartz content (lithoarenites).

5.2.2 ROCK TYPES OF RESERVOIRS IN SAVA FIELD

In the entire sedimentation sequence, 11 hydrocarbon pools have been defined in Late Miocene sediments that belong to the Neogene and Quaternary periods. Generally, the Middle and partially Upper Miocene clastic sedimentation was greatly influenced by a pre-Neogene basement palaeo-relief.

As Figure 5-18 shows, Sava Field is an asymmetrical brachianticline. Its axis is in a northwest-southeast direction and a slightly pronounced peak in the southern part of the structure can also be seen. According to the seismic interpretation (Report[3]) the presence of normal and reverse faults were established. These faults (mostly normal) originate from the continuous tectonic activity during Pliocene, and were induced by reverse faults in the south-western part of the field.

In this case study the method was only applied to two reservoirs of Upper Miocene age. The total thickness varies between 120-150m in the whole rock body which contains all 11 reservoir units. The reservoirs would all be quite large scale for the model. The two selected reservoirs are the two largest and these are positioned directly one above the other. These two reservoirs have been divided into two different depositional units in the reservoir rocks. Since the marl between the two units seems to be very thin in some wells, both units were the subject of the depositional environment analysis.

The analysed sequence is made of Upper Miocene marls, siltstones and sandstones. The latter two clastic (psammitic) lithofacies were deposited by periodical turbidity currents in the entire depression (e.g., Šimon, 1980; Novak-Zelenika et al., 2012). This sedimentation was continuous during the Pannonian and Pontian ages, when the entire lacustrine area was constantly reducing in size, depth and salinity (e.g., Vrbanac, 1996; Malvić, Velić, 2011). Detritus was redeposited several times before it finally accumulated (Malvić et al., 2005; Malvić, Velić, 2011). The morphology of sandstone bodies follows the direction of turbidite currents. At the axes of these flows generally thick bedded, fine-grained sandstones were

[3] Unpublished report about study field in Sava Depression - Tertiary CO2 Injection (2003): INA, Zagreb

deposited. Towards the rims of these depocentres, sandstones are gradually replaced by siltstones and marls (e.g. Saftić et al., 2003).

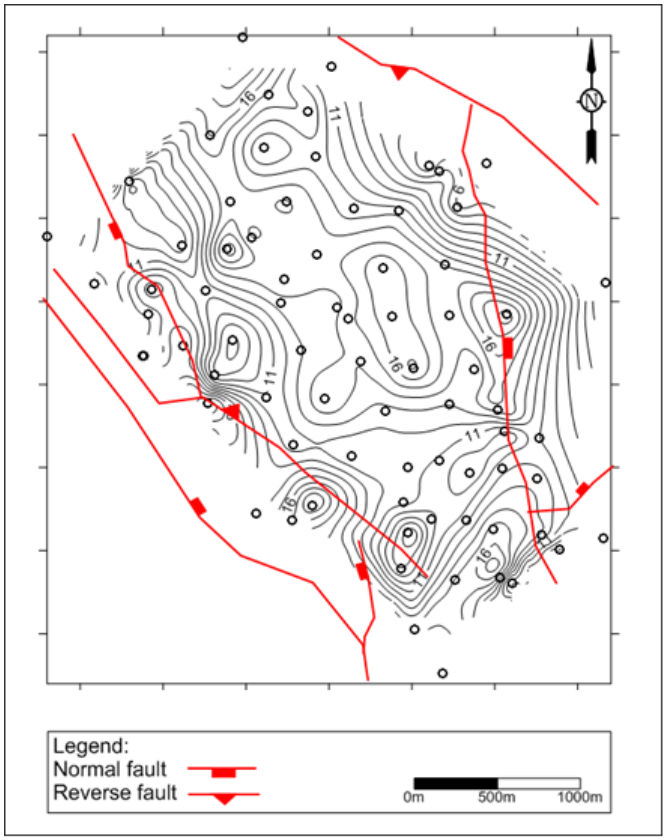


Figure 5-18: Well density of the study area (Sava Field) with gross thickness contour of largest HC pool and the location of fault

5.2.3 SOURCE OF DATA FROM THE RESERVOIR OF SAVA FIELD

The selected reservoir rocks were transformed into a stratigraphic coordinate system (similarly to the first study case). The vertical coordinate was measured from the top of reservoirs.

The average gross thickness of the complete reservoirs varies from several meters to 21m. All reservoir rocks in this new coordinate system were cut by lateral surfaces which are parallel whit the top. The vertical distance between each lateral surface was 1m.

The analysed data comes from 78 wells (Figs. 5-18). Geophysical logs with their quantitative petrophysical interpretations of porosity, water saturation and shale volume were available at 0.2m intervals. Each horizontal surface contained the calculated averages of the

petrophysical variables from any 1m thick interval, so, the original readings were averaged at each 1m thick interval starting at the top. In addition to the interpreted petrophysical data, one categorical data 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’.

5.2.4 METHODS APPLIED FOR SAVA FIELD

5.2.4.1 Settings of UNN and generated clusters

The UNN method was applied for a set containing all samples belonging to the two reservoirs. The input variables of UNN are the petrophysical parameters (porosity, water saturation, shale volume) and a categorical variable, lithotype. The dataset was again randomly subdivided into training, test and validation sets. The sizes of these subsets were the same as for the first study area (60-20-20%). The required number of clusters was four according to the preliminary information about the facies.

The training rate altered from 0.5 to 0.03 during the training cycles which were maximised to 5000 cycles. The training showed a larger test error after 5000 cycles. Because it was hard to define the best number of cycles, the length of learning could be affected by the value of the test error, so, the training was halted independently of the cycle when the test error started to increase (Fig. 4-2).

The neighbourhood radius was only one and the adjacent area specified a 1x4 square which gave a linear Kohonen layer. Normal randomisation of weights was used for the training.

The four separated clusters in the training cycles were controlled by the neurons and their lateral connections. The general framework and the tabulated results of the major averages are represented in Figure 5-19. The averages displayed in Figure 5-19 were calculated in each cluster. The calculation was based on data points which belong to single groups. The most general lithology code was highlighted from each cluster.

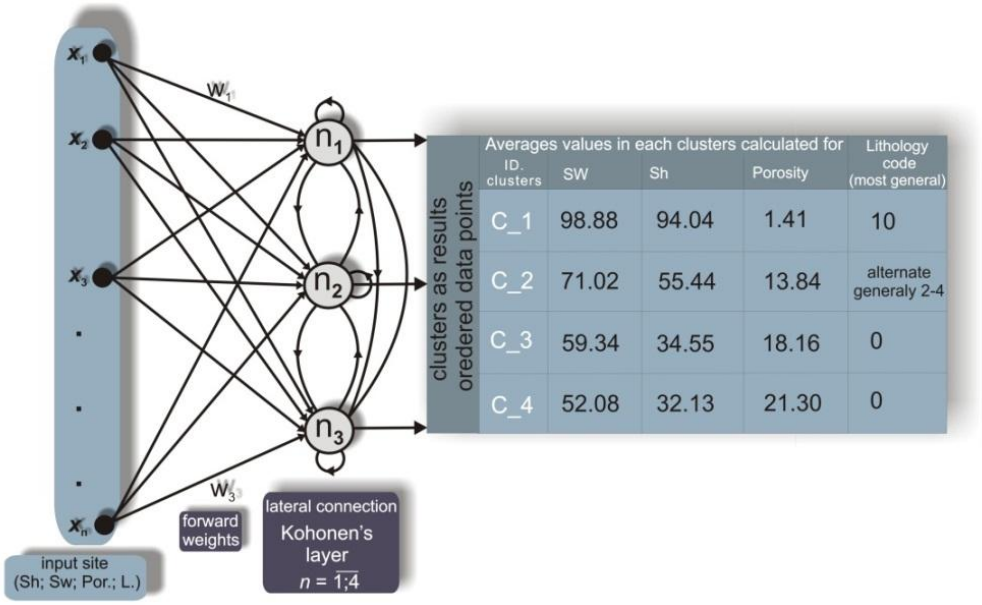


Figure 5-19: Schematic graph of constructed UNN with input and output in the Sava reservoirs

5.2.4.2 Statistical comparison of clusters

The sizes of the created four clusters were different (Fig. 5-20). Altogether, there are 1578 data points in the input sets of 76 wells. 200 data points belong to Clusters C_1, and Clusters C_2, C_3 and C_4, contained 210, 363 and 800 data points respectively.

The average group porosity, and the most frequent lithotype suggested that these reservoirs are primarily dominated by sandy sediments with more than 70% sand content (Fig. 5-20). Consequently, it is a sand-rich system. This will be an important statement in the identification of the environment because the general models of deep-water clastic turbidite systems are characterised by grain size according to, for example, Reading and Richards, (1994), and Richards et al. (1998).

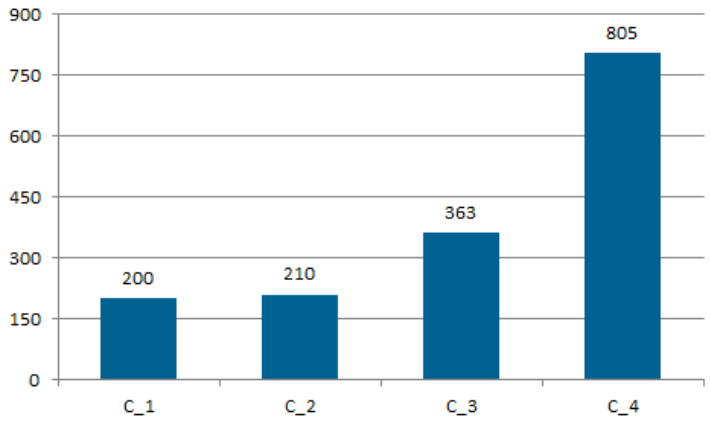


Figure 5-20: Frequency histogram of elements of clusters in both reservoirs in Sava Field

As in the first case study, a Monte Carlo simulation was applied to increase the cluster sizes. The simulated set was checked against the null hypothesis that both the original and the simulated datasets come from different populations. According to Goodman and Kruskal's gamma coefficient those samples were significantly similar using a 0.05 signification level. Consequently, the sets of averages were acceptable in the statistical description and in the comparison between clusters. The following statistical analyses thus derive from this extended dataset (by Monte Carlo simulation) where the increased number of data points was 1500.

Since the analysis confirmed the calculated averages of partitioned clusters, cluster C_1 was regarded as independent from the others. It contains mainly shale lithologies with very poor effective porosity. Its average porosity was not greater than 1.41% (Fig. 5-19). C_1 probably represents massive marl sediment on the basin plain.

The next question was whether the other three clusters were significantly separated from each other. A Mann-Whitney non-parametric test and the variance analysis (comparing WGV and BGV) helped to answer this question. In the non-parametric statistical tests the null-hypothesis was that the ranks of medians were equals.

The results showed significant difference at p=0.05, however a comparison of histograms and box-plots (Fig.5-21) suggests that although the difference is significant, the characters of C_3 and C_4 still were closer to one other than to any other clusters.

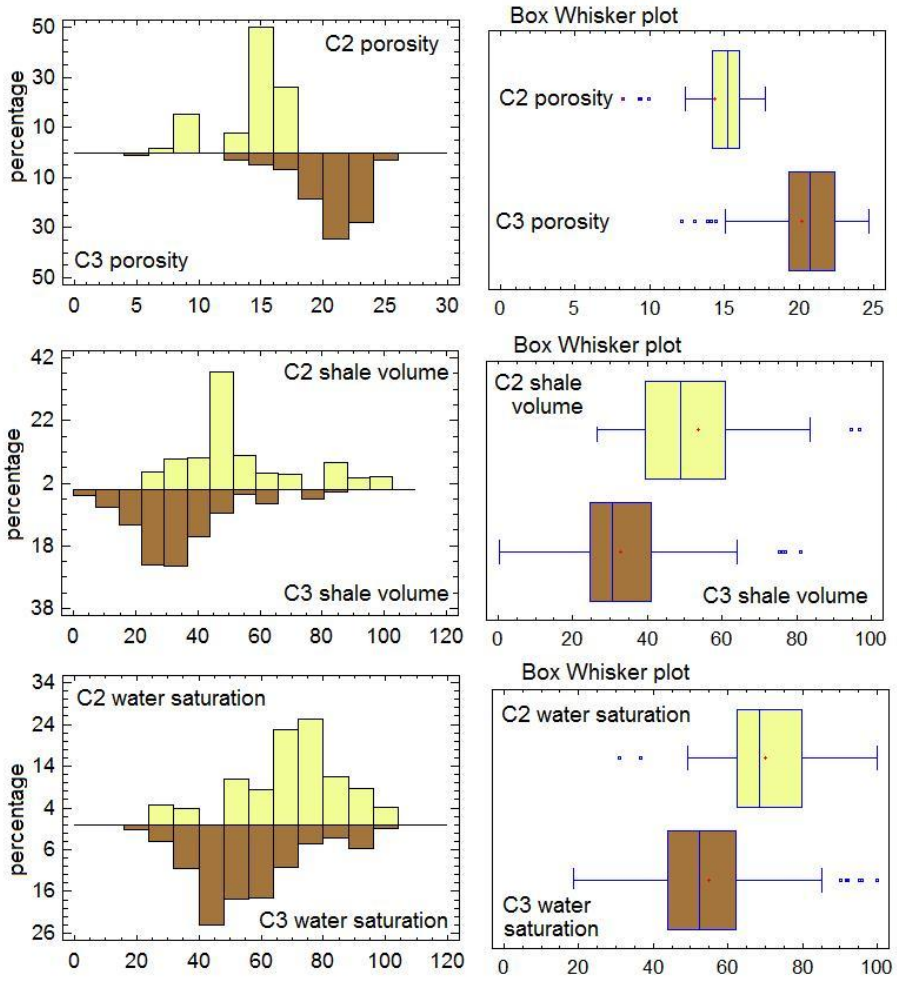


Figure 5-21: Statistical comparison of clusters C_2 and C_3 (Sava Field)

Cluster C_2 was characterised by 13.8% average effective porosity and relatively high shale content (about 55%) (Fig. 5-19). Cluster C_2 had a bimodal porosity distribution where the major mode was at 15% and the minor was only 9%. The shale values also had a bimodal distribution. The major mode corresponded to the average (55%) and the minor mode was 85%. This bimodality may be derived from the lithological character. The sandy deposit often alternated with shale. In this cluster the lithology code changed at relatively large intervals, but the most frequent lithology code was 2-4 suggesting that only a few shale beds interrupted the sandy deposition (Fig. 5-19). This cluster was regarded as laminated sandstones with siltstones and marls.

The table of averages in Figure 5-19 showed that the mean effective porosity was 18% in the third cluster. In the fourth cluster the porosity showed close values, 21%. The shale volumes were also close to each other (34% and 32%). Consequently the descriptive statistics

(Fig. 5-22) suggested that these cluster faces represented very similar sediments with slightly shifted parameters. Both clusters contained sandy deposits, but in C_4 the porosity and the shale volume showed better quality sandstones than in C_3. Cluster C_3 contained samples with low effective porosity but quite high sand content. This cluster can be characterised by thin sandstone with interrupted siltstones and massive sandstones.

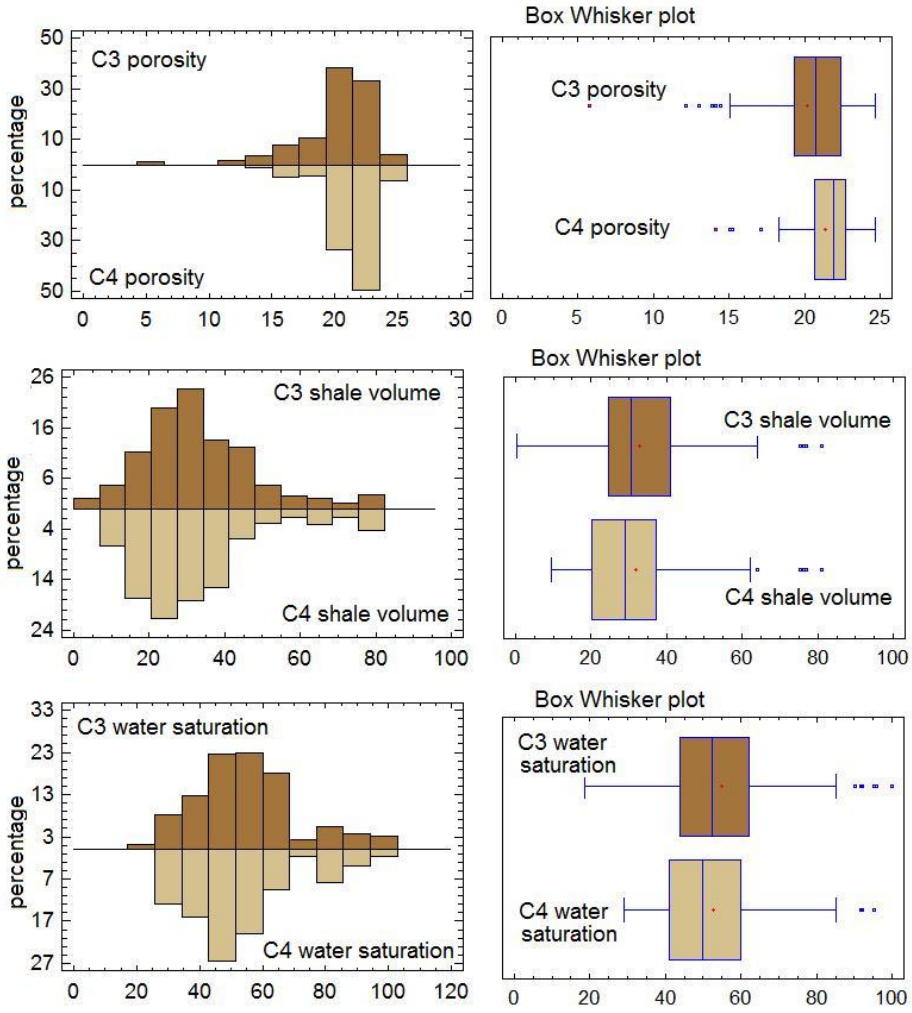


Figure5-22: Statistical comparison of clusters C_3 and C_4 (Sava Field)

The comparison of within group and between group variances (Table 5-4) showed a close relationship between these clusters. The between group variance is lower than the within group variance. Presumably, this analysis shows that (i) clustering of separated clusters which are not heterogeneous enough, or (ii) that these clusters belong to a complex system such as a chainlink dataset (see Chapter 2). In case of assumption (i), it is necessary to pool Clusters

C_3and C_4. In case of assumption (ii), the analysis revealed miss-clustering. Unfortunately, a clear decision could not be made. There are several validation techniques to reveal misclassification but each depends on a different approach to cluster separation definition. Thus, these clusters will be pooled in the lateral extension. This pooling modifies the distributions of cluster elements, as shown in Figure 5-23, where C_4_m is the merged C_3 and C_4 clusters.

In the variance analysis the applied parameters were as follows:

$m = \overline{1,2}$

n = 1168

n_1=363; n_2=805; (from Fig.5-4)

Table 5-4: Comparison of WGV and BGV for Clusters C_3 and C_4

	WGV	BGV	TV	WGV / TV rate (%)	BGV / TV rate (%)
Sw	174.5	240.17	414.68	42.08	57.92
SH	261.21	195.55	456.75	57.19	42.81
POR	0.0004	0.0002	0.0006	64.17	35.83

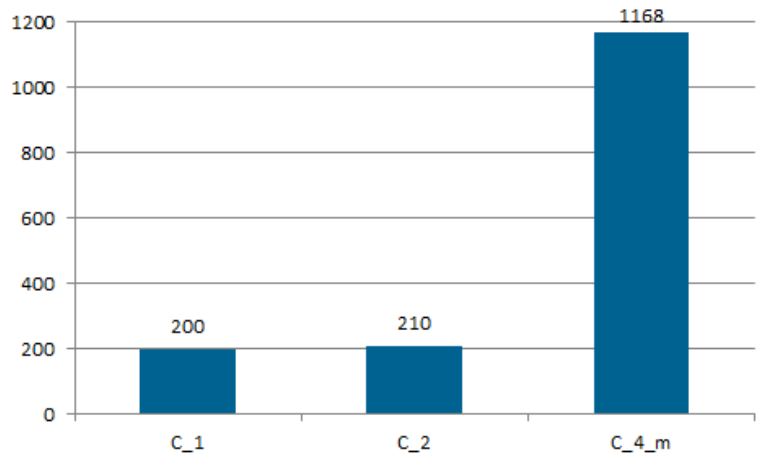


Figure 5-23: Frequency histogram of elements of clusters after pooling C_3 and C_4 in both reservoirs in Sava Field

5.2.4.3 Lateral extension of clusters using IK in Sava Field

According to the statistical information, the clusters are defined lithologically, but during the identification of facies it is necessary to visualise the clusters on a map. The extended cluster geometry can define the facies sedimentologically. The geometry and other

properties of submarine-fan systems were collected in several papers, including Reading and Richards (1994) and Richards et al. (1998). These papers restricted the terms of submarine fans to single point source turbidite systems. This paper follows the scheme of Reading and Richards (1994) which is the most accepted and most often cited work on this issue. The identification of facies follows the visualisation of clusters (based on these papers).

As this is a sand-rich system, the potential facies are as follows:

(i) On the proximal area of the deep-water siliciclastic system, there are channels which are characterised by elongated finger-like geometry, where sand content is greater than 60-70% and the porosity values are the largest.

(ii) On the distal area of submarine fan, the channelized lobes are in elongated patterns which are often broken up with bifurcated channel geometry. They also usually have greater than 60-70% sand content.

(iii) Where the sand content is less than 60-70%, the channel levees follow the channel beds.

(iv) The basin plain surrounds the submarine fan system, with low porosity and large shale volumes.

These listed points can be used to recognise and describe the extended cluster facies which were visualised by IK.

5.2.4.4 Preparation of input data for the IK mapping process

The well density was first analysed in the pre-processing of the lateral extension. The lateral distribution of wells is almost uniform on the field, so the wells were all applied in the variography analyses.

At the beginning of the study both reservoir rock bodies were cut by surfaces parallel with each other and by the top surface of the reservoirs. Vertically they were 1m apart. As a result, both reservoirs decomposed approximately 20 horizontal surfaces. The UNN process separated the clusters using data from each well in all horizontal surfaces, but the lateral extension was not applied to all horizontal surfaces; instead, some lateral surfaces were selected in both reservoirs.

This selection depended on two things: (i) the alternations of clusters distribution; and (ii) the alternations of porosity and sand content properties in the single surfaces. Eight

horizontal surfaces were selected in the reservoirs altogether. In the case of the lower reservoir, they were 13, 11, 10, 7, and 4m from the top. In the case of the upper reservoir they were 9, 6, 3m from the reservoir top. The lateral distributions of porosity and sand content on these surfaces are shown in Figures 5-24, and 5-25. These figures present the differences in spatial distribution of clastic sediments during the different deposition phases.

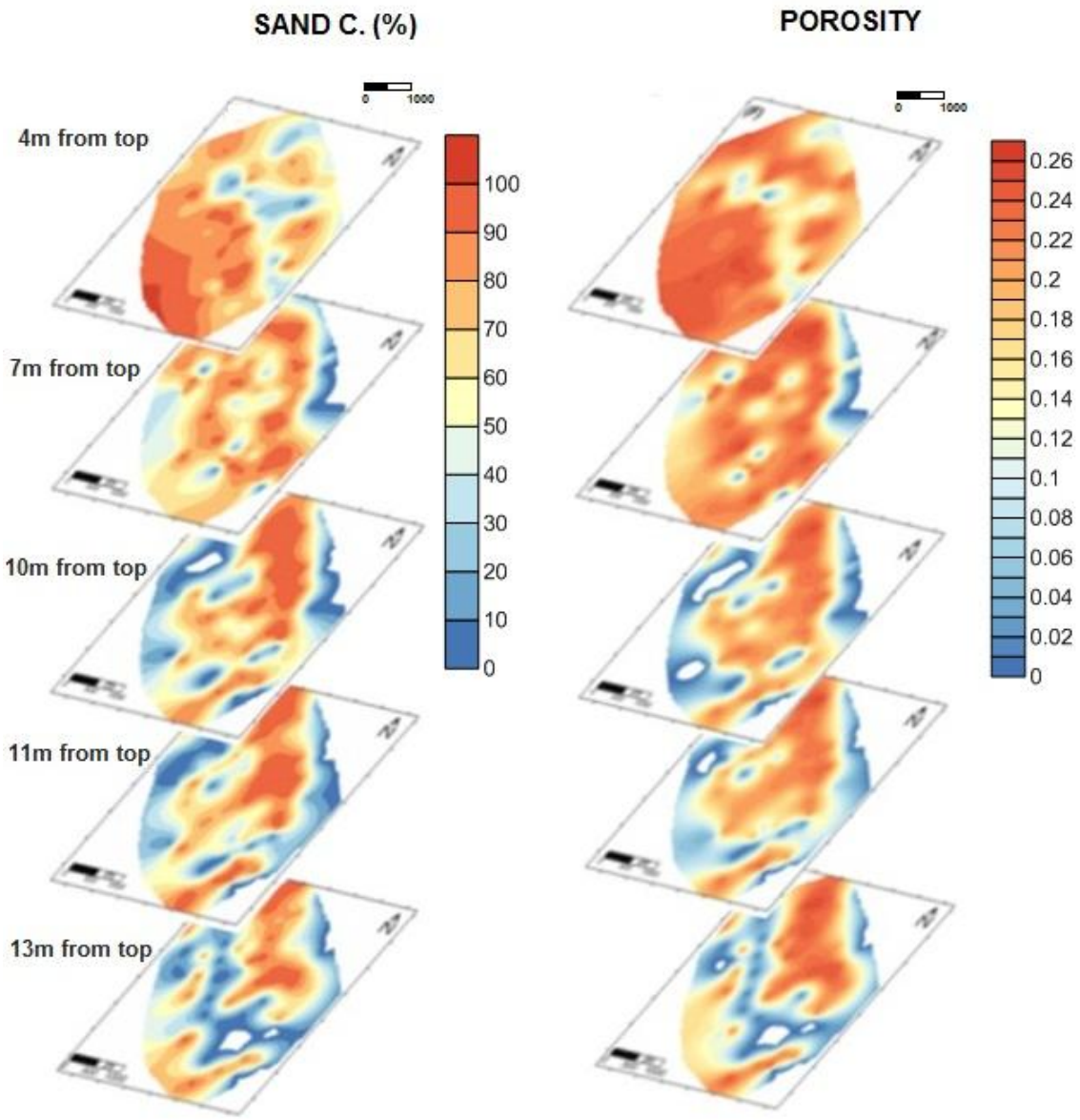


Figure 5-24: The selected horizontal surfaces from the lower reservoir where clusters were laterally extended (Surfaces 13, 11, 10, 7 and 4m below the low permeability seal)

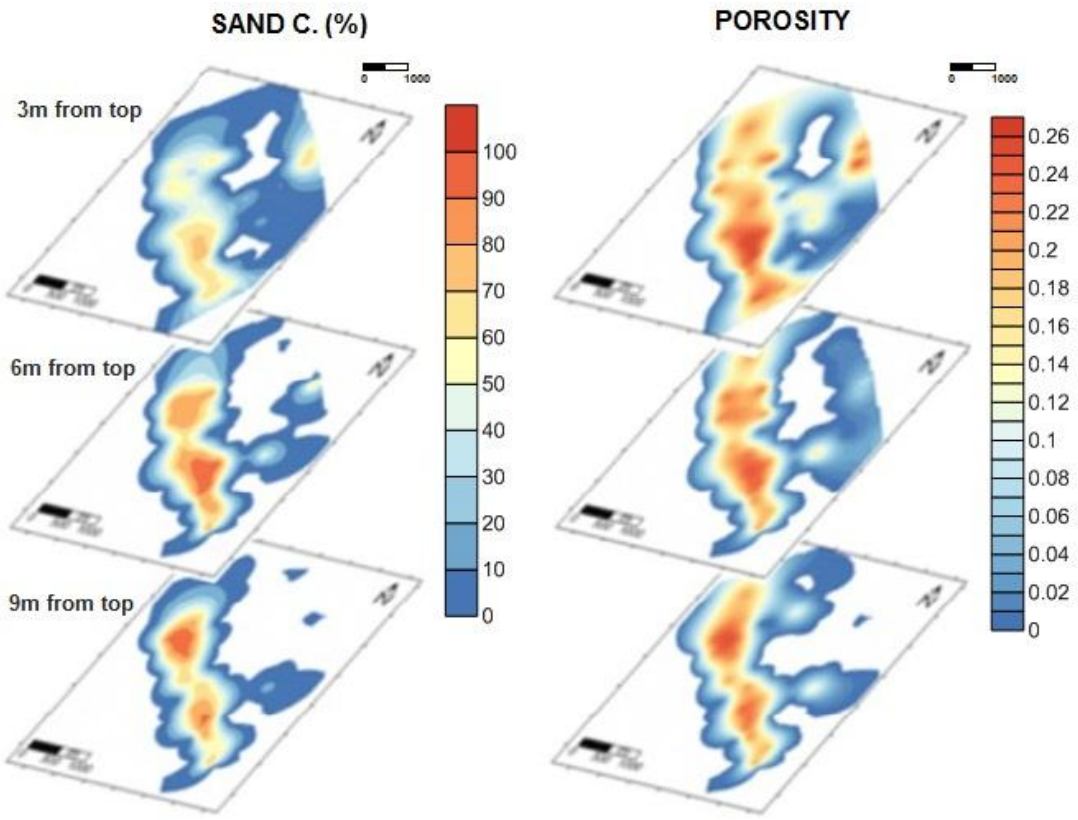


Figure 5-25: The selected horizontal surfaces from the upper reservoir where clusters were laterally extended (Surfaces 9, 6, 3m below the reservoir top)

Figure 5-25 shows that the differences were insignificant in the upper reservoir. The depositional phases did not change greatly, the lateral and vertical differences in the spatial distribution of sediments are not significant. This is why the lateral extensions of clusters using IK were prepared only in these three selected surfaces. On the other hand, the number of elements of the principal cluster (pooled C_4_m) also had to be considered in the selection of surfaces. The unit with high sand content defines the pattern of deposition well. The deepest horizon surface contains only a few wells which represent the C_4_m cluster. In this way it is difficult to fit a variogram model for the less frequent cluster in an isolated location. As a result the selected lowest surface lies 9m from the top.

It was also found that boreholes which belong to Cluster C_1 and C_2 were reduced. Conversely the number of elements in Cluster C_4_m increased in the upper surfaces. This was seen in both reservoirs. Because of this, for example, the upper reservoir top cluster C_2 appeared in the surface 10m below the lower reservoir top or 9m below only in few isolated

locations. In this case it is not possible to fit the variogram model to this cluster location. The variogram surfaces of this cluster could refer to random spatial arrangements.

In the pre-processing of IK the variogram surfaces was analysed for all clusters separately in each of the selected surfaces. Appendices 11 - 18 show the variography surfaces of the clusters in all 8 selected horizontal surfaces. The variogram surfaces of the clusters were appropriate to use in modelling.

The grid line geometry also derived from the variogram surface, and was imported to the IK estimation. Usually the modelling of continuity is based on two directional variograms. The first was chosen from [0°;90°] and the next was selected from [90°;180°] with 22.5° of angular tolerance. Every cluster was modelled individually. The settings and model profiles are shown in Appendices 12 - 38. Every model contained two or three structures which were characterised by seven parameters. These were the same coefficients that were listed in the Szőreg-1 case study (App. 39 - 46).

The results of IK are shown in Figures 5-26 - 5-33. These figures show the lateral distribution of clusters and the contour geometry of the principal cluster (C_4_m) according to the probability contour map.

5.2.5 INTERPRETATION OF LATERALLY EXTENDED CLUSTERS IN SAVA RESERVOIRS

Figures 5-26 to 5-33 represent the lateral extensions of point-like results on the contours of porosity and sand content. Each figure comprises the lateral distribution of clusters (Part A in Fig. 5-26 - 5-33). Part B represents the spatial probability of the principal clusters (C_4_m with large sand content and the highest porosity value). Parts C and D represent the contour of porosity and sand content maps combined with probability map of the principal cluster. The last two parts (E and F) show the spatial distribution of sand content and porosity in the surface generally.

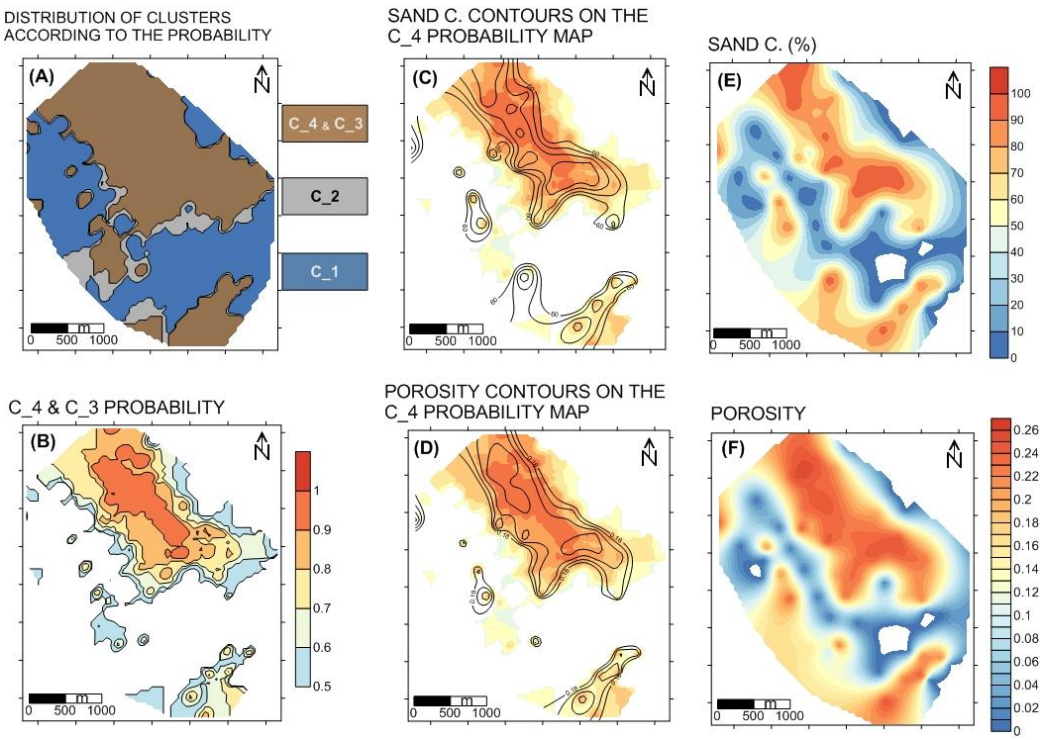


Figure 5-26: Results from the surface 13m below the reservoir top (in the lower reservoir)

In the three selected surfaces (13, 11, 10m below the reservoir top) the main clusters appeared in elongated and bifurcated geometry with large estimated probability. This channel shape extends in a NW-SE direction. It suggests a progradation process in a NW to SE direction. This progradation can be followed surface by surface in Figures 5-26 - 5-28. These maps show an increasing number of distributary channels. In these maps the geometry of C_4_m defines the channel or channelized lob facies by dendroid geometry in the turbidite system. In this system cluster C_2 appears at the forefront of lobs, between the bifurcation channels, or follows the channel system at the marginal parts. These clusters are surrounded by Cluster C_1 which represents basin pelitic deposits.

In all three surfaces (13, 11, 10m below the reservoir top) the extended principal cluster with large probability corresponds to the contours of sand content (60%) and effective porosity (18%). The result maps show that in addition to the main direction there is a secondary direction where the sand body spreads. The depositional strike of this elongated finger geometry suggests a south-north transport direction. The development of this sandy area was also seen on the lowest selected surface. The progradation continues until the sandy deposits from two directions interfinger (in the surface 10m below the low permeability seal).

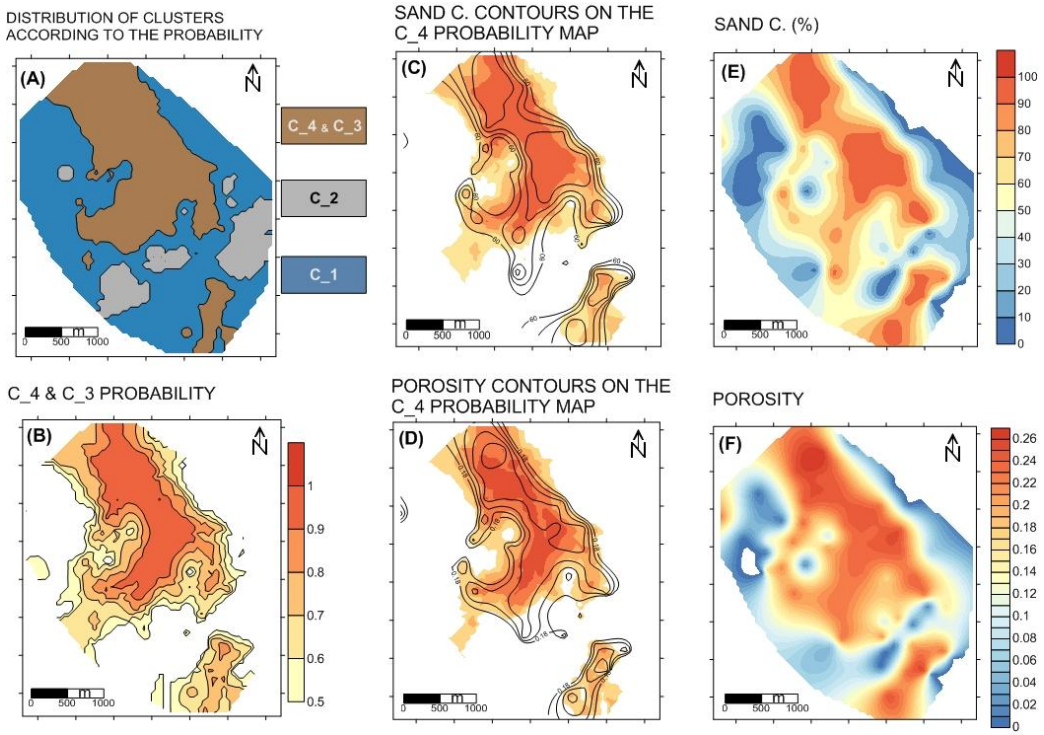


Figure 5-27: Result for the surface 11m below the reservoir top (in the lower reservoir)

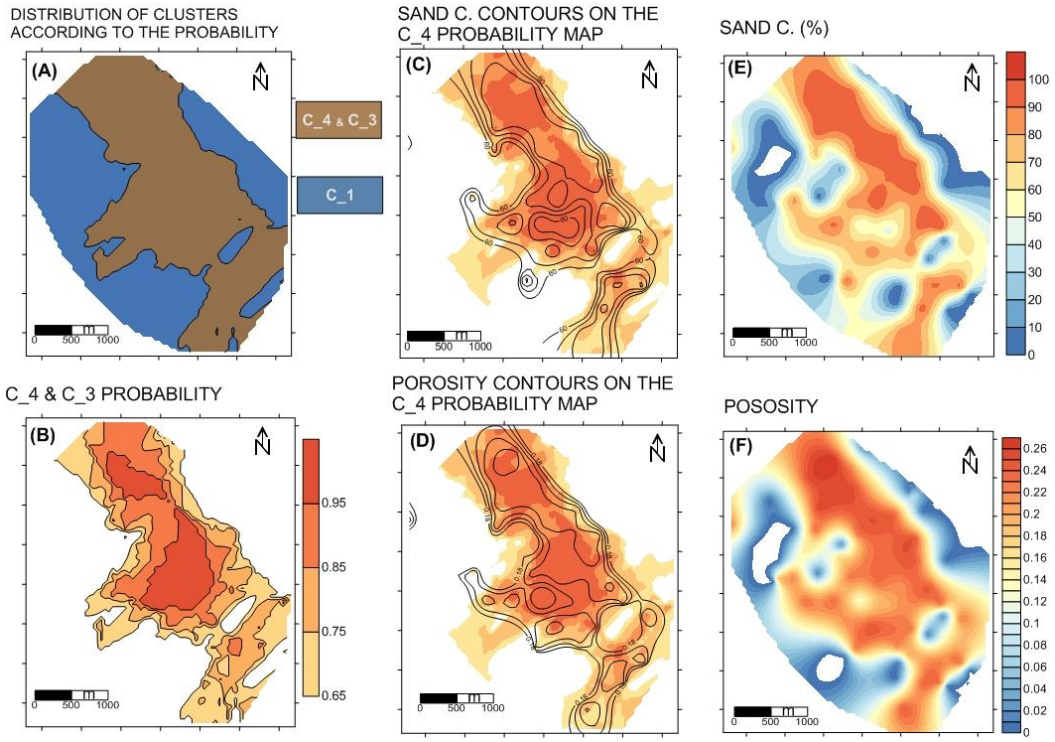


Figure 5-28: Result for the surface 10m below the top (lower reservoir)

Due to the progradation of the submarine fan, three large distributary channels developed in the next selected surface (7m below the top).

C_2 is detected sporadically, only in a few boreholes, 7m below the top. As a result, C_2 does not appear in this surface. In contrast with the previous surface, the results map contains the lateral extension of Clusters C_3 and C_4 separately. Both clusters were separated by UNN in a relatively large number of boreholes. This made it possible to fit variograms to both cluster locations and analyse the lateral distribution of these clusters despite the variance analysis. The IK process supported the decision that these clusters were also spatially separated, and they can be defined as two different facies within the C_4_m cluster facies.

The channel system contained an increasing number of advanced bifurcate channels (Fig. 5-29, E and F). It was interesting to determine which part of the main sandy deposit was characterised by cluster C_3. This sandy deposit was of lower quality than C_4 according to the porosity values and shale content. Group C_3 appeared among the channelized lobs.

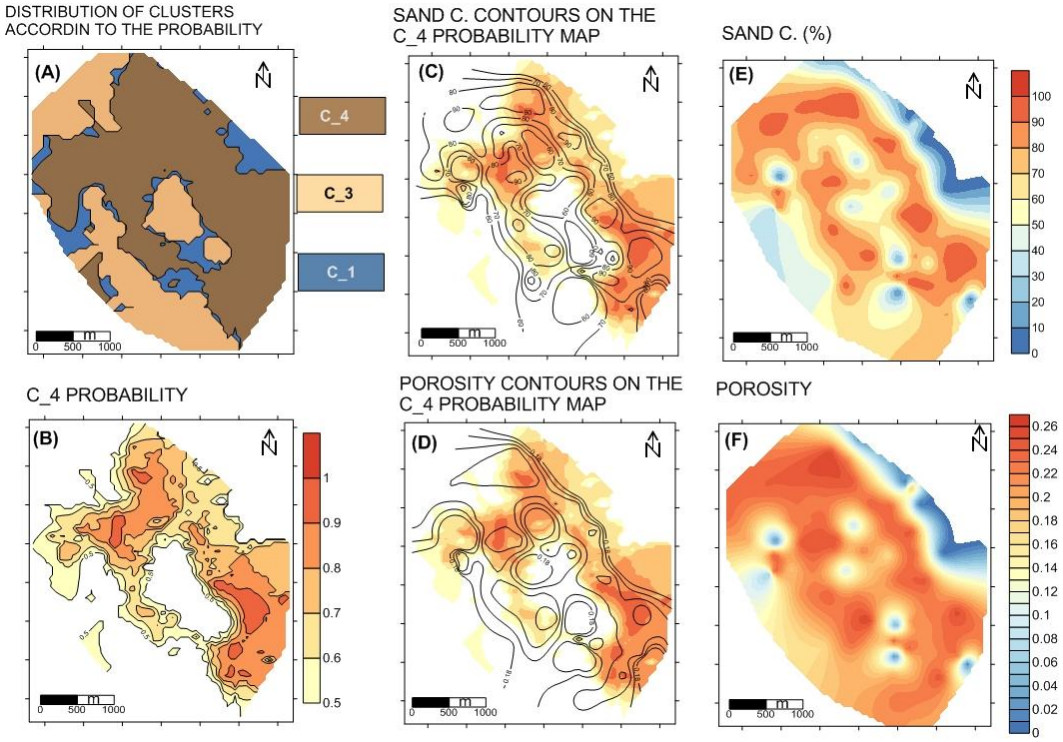


Figure 5-29: Result for the surface 7m below the top (lower reservoir)

C and D maps of Figure 5-29 represent a slightly larger variability in porosity and sand content within the contour of Cluster C_4. Sand content varied between 50-80% and the porosity changes between 16-18%, generally. In the lower surface channels, channelized lobes are characterised by around 20-22% of mean porosity and 70-80% of mean sand content. Within the contour of Cluster C_3 the average porosity is around 14-16 and sand content is 40-50%.

The definition of C_3 as the inactive lob areas of a deep-water fan system was based on three reasons: (1) it is located among the channelized lobes; (2) its shape seems to be elongated lobs; (3) its porosity values and sand content are smaller by one order than that of C_4.

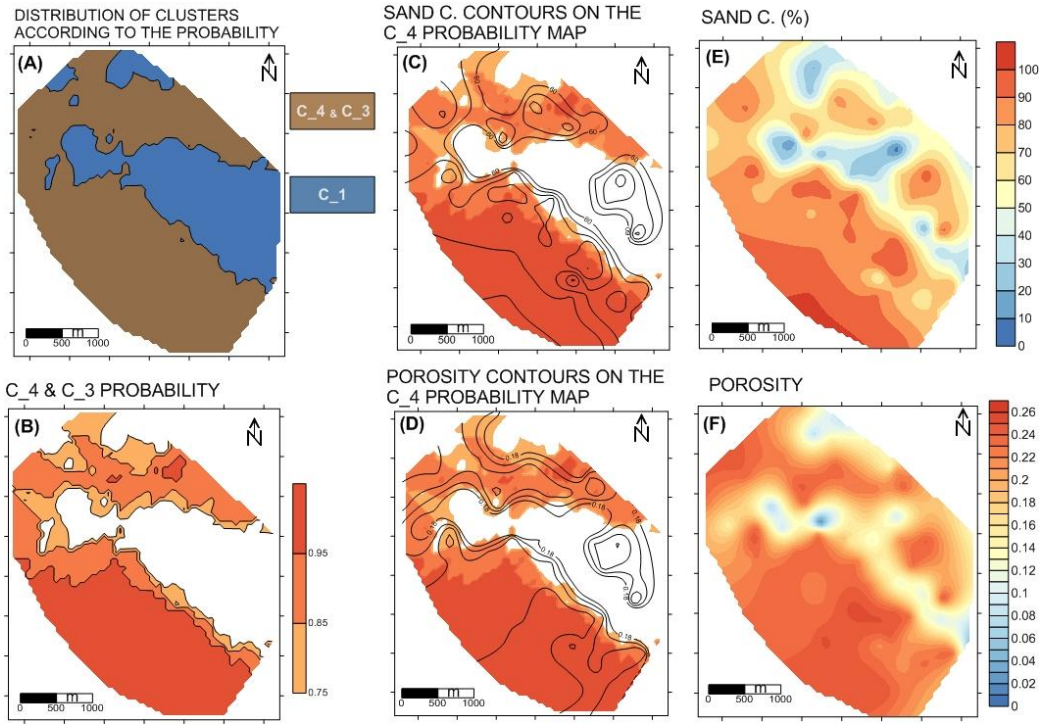


Figure 5-30: Result for the surface 4m below the top (lower reservoir)

In the final selected surface, 4m below the reservoir top, the geometry of the sand deposit completely changed and a generally sandy deposit dominated in the whole area (Fig. 5-30). In the northern part of the field an elongated channel complex extended along the NW-SE line. The bifurcation channel system characterised by the sand-rich principal group (C_4_m) prograded to a SE direction and also showed lateral movement. This sand-rich deposit is more extensive towards the eastern and western edges.

Farther up, the next reservoir clusters are appeared only in three selected lateral surfaces. According to the porosity and sand content contour map it was apparent that the spatial distribution of clastic rock in this reservoir differs from the previous. In addition to the geometrical aspect, there is another difference between the lower and upper reservoirs. In all surfaces the sand body geometry identifies as an elongated braided channel without bifurcation or lob geometry. This shape is a so-called shoestring geometry. The point-like extension of clusters also reveals this spatial distribution in the reservoir. The probability of the main clusters corresponds to the porosity or sand content values (Fig. 5-31 - 5-33). The cluster extension follows 50% sand content and 15-16% porosity contour lines.

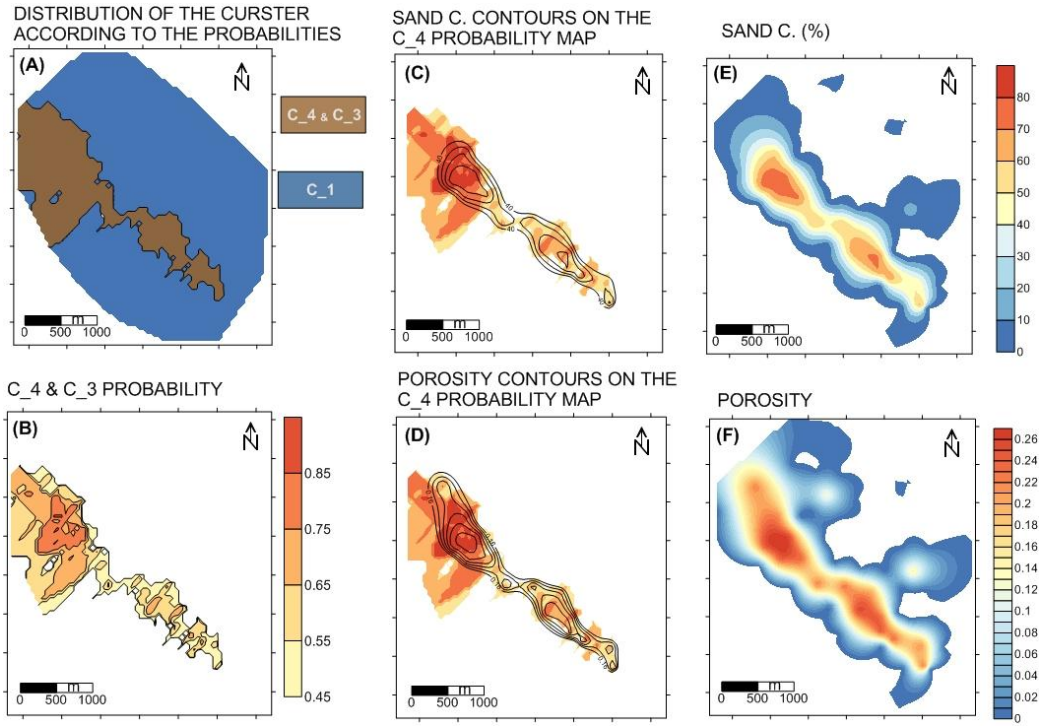


Figure 5-31: Result for the surface 9m below the top (upper reservoir)

The area of the identifiable channel system varies in all surfaces in the vertical series. Its lateral dimension becomes larger, and its elongated shape is more widespread in the SE direction. It is a clear progradation mechanism. The geometry and the relatively large porosity and sand content values indicate that in this reservoir the main clusters can be identified in the same way as for the previous reservoir. However, it is also noticeable that these large sand content and porosity values are slightly lower in the Sava reservoir (this difference is generally only 3-4% of porosity and 10-20% of sand content). This means that the same facies

are characterised by somewhat worse petrophysical properties. This is the effect of the cluster algorithm which made the outliers in some clusters possible.

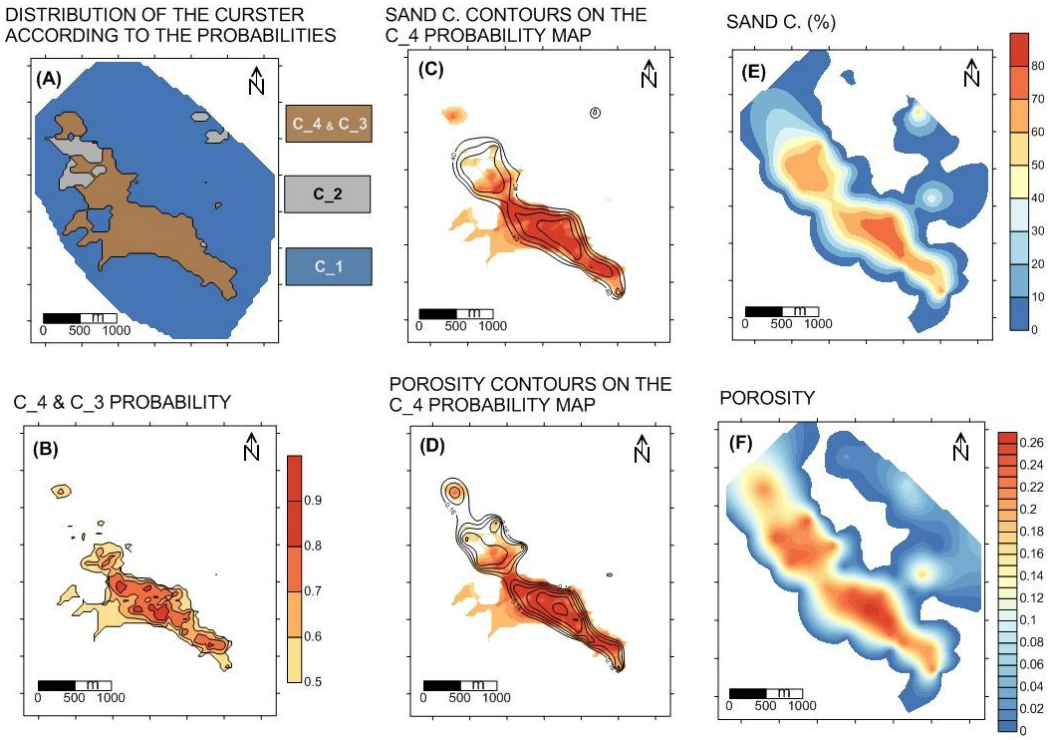


Figure 5-32: Result for the surface 6m below the top (upper reservoir)

Another difference between the two reservoirs is that the relative frequency of Cluster C_2 decreases from the lower towards the top surfaces. This cluster appears first at 6m below the top and has larger area above this surface. This spatial pattern is seen in the north-west part of the field.

In this reservoir Cluster C_1 surrounds other facies as well. It can be characterised as having lower porosity and sand content and it is defined as basin plain sediment. This cluster does not have a well-defined spatial structure and geometry.

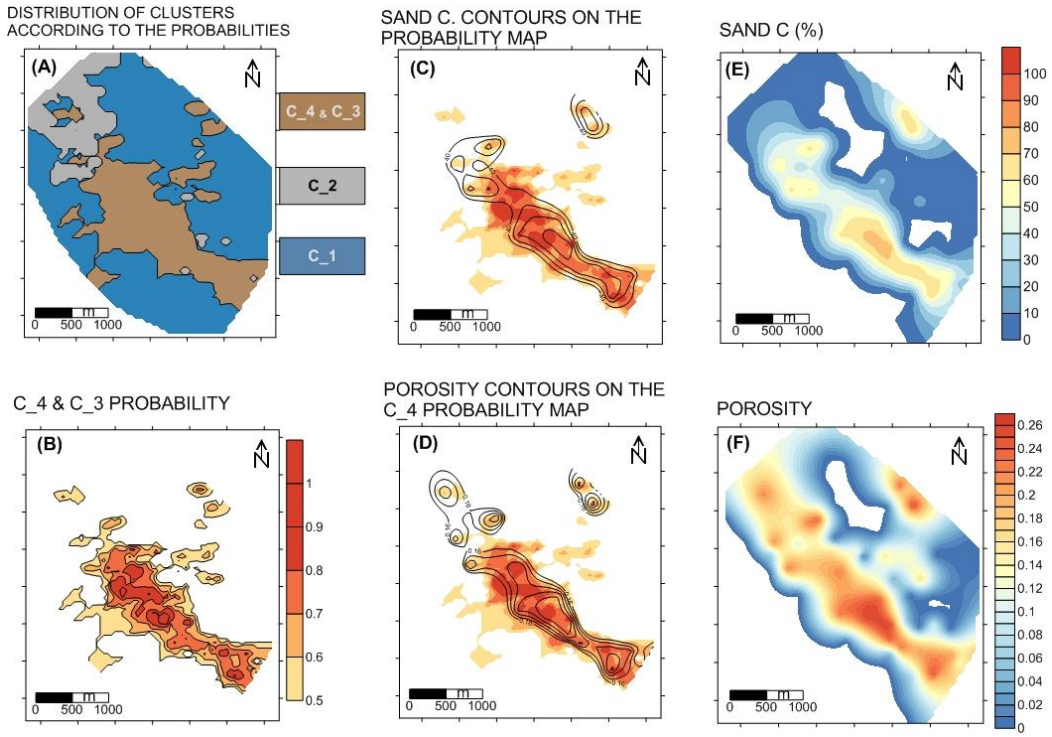


Figure 5-33: Result for the surface 3m below the top (upper reservoir)

5.2.6 CONCLUSIONS OF THE INTERPRETATION OF DEFINED CLUSTERS

The analysed clusters were defined lithologically on the basis of their statistical characters. In addition to this, their probability based point-like extension also suggests certain sub-environments in the deep-water fan system. Because Clusters C_3 and C_4 (sandy deposits) were predominant in the clastic rock body, this clastic deposition is identified as a sand-rich submarine fan system. This is important because the efficiency or range of turbidity currents can be defined. Figure 5-34 demonstrates the dissimilarity between the different systems based on the dominant grain size.

According to Reading and Richards (1994) the transport efficiency and the scale of the depositional system are inversely proportional to the grain size. All types have a distinctive characteristic according to the geometrical aspect, the range and the dominant grain size (Fig. 5-34). The sand-rich system has the smallest range (generally 10km, but it is never more than 50km) and the most compact lobate geometry (the length and radius volume is similar, but it also depends on the basin geomorphology). The efficiency describes how far the turbidite

flows can transport the sediments. Small grain sizes indicate that the transport mechanism soon stopped. In a sand-rich system this efficiency thus means the smallest distance.

On the other hand, this model may also be useful in the characterisation of the main reservoir property. The main depocentre of the sand deposition is the mid-fan or the close proximity of the inner fan. In Normark's (1978) terminology, the mid-fan area is the suprafan lobe. The volume of mud within turbidites is low. This means that the reservoir homogeneity is fairly good so the connectivity of sand bodies may also be good both horizontally and vertically.

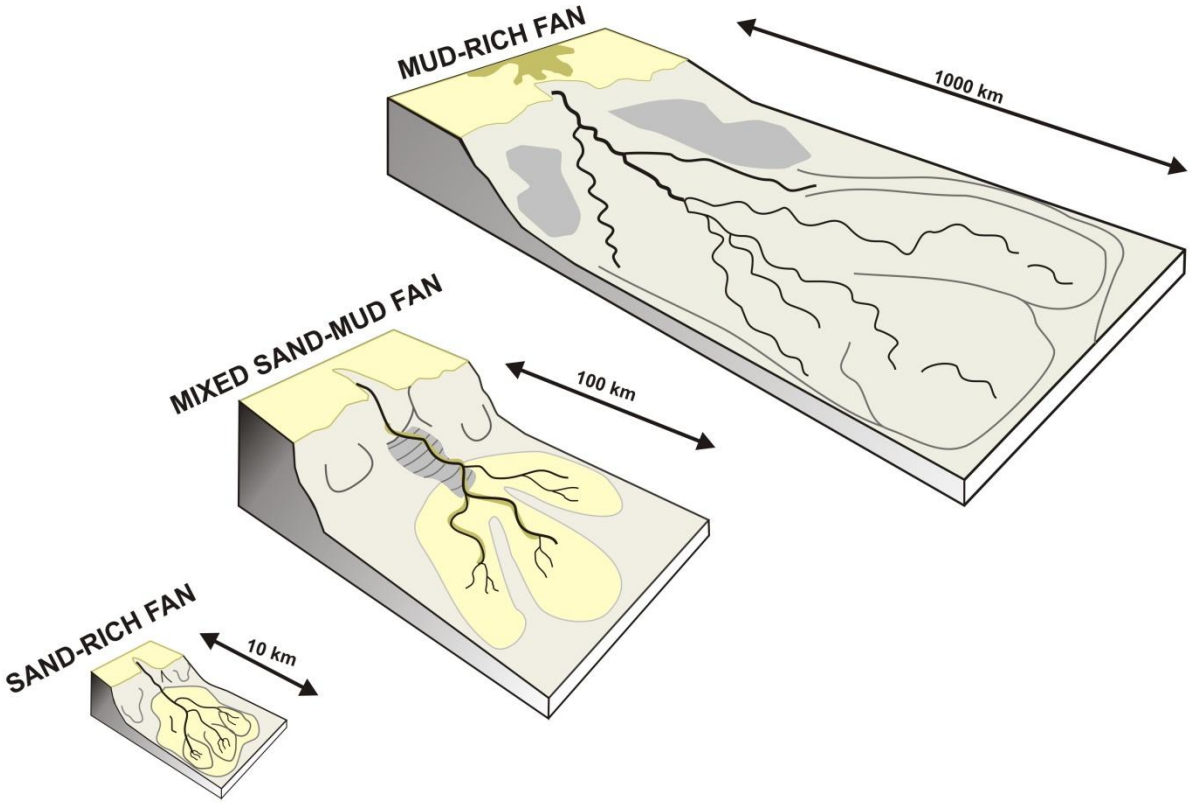


Figure 5-34: [4]Schematic block diagrams of deep-water clastic submarine fans according to dominant grain size and the range of turbidity currents

It is also important to define the types of sand-rich clastic system. Based on Reading and Richards (1994), there are three types. These are the point source submarine fans, multiple source submarine ramps and the linear source slope aprons. Considering the

[4] Figure 5-35 is prepared after http://www.sepmstrata.org/page.aspx?pageid=40, 15.08.2013

character of the depositional basin area, the study area may correspond to the first type. According to the geometrical aspects and considering the statistical characteristics, the defined facies describe a sand-rich point source submarine fan system. The result maps (Fig. 5-26 -5-33) show elongated shapes, and their corresponding petrophysical properties define channel systems: braided channels with shoestring geometry, bifurcation channels with dendroids, and lobate sand deposit with a radial fan shape.

In the vertical series of lateral surfaces the system suggests some phases of progradation (P1 and P2 in Fig. 5-35). These phases are represented by the results maps for the surfaces 13, 11, 10m below the top of the lower reservoir (Fig. 5-26 – 5-28). In the next phase (P3 in Fig. 5-35) the proximal part of the mid-fan area is shown by the upper horizontal surfaces of the lower reservoir (Fig. 5-29 – 5-30).

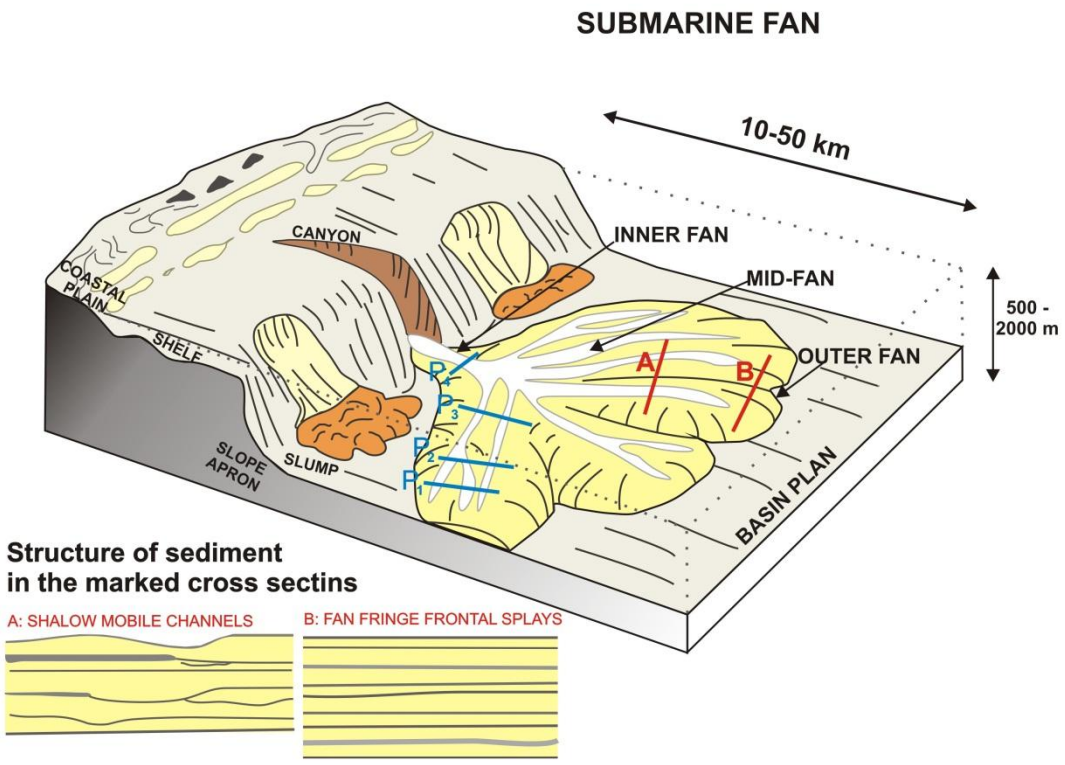


Figure 5-35: Summary block diagram illustrating a schematic sand-rich fan system including phases of progradation (P1 - P4) and structures of deposition in two cross-sections in mid-fan (A, B) (after Reading and Richards, 1994)

In the upper reservoirs the deposition processes are characterised by a braided channel without bifurcation. It may be the main channel in the inner fan area. In this case this facies

outlines the main transit zone, the so-called feeder channel. In the progradation mechanism this is demonstrated by phase P4 in the summary block (Fig. 5-35). Usually this facies is described as a sandy deposit of lower quality (in porosity values and sand content). In this upper reservoir this area is characterised by a 10-20% lower sand content and 4-5% lower porosity values than in the lower reservoir.

The probability contours of the principal cluster facies changed continuously according to certain horizontal surfaces. This change shows the sedimentation history of the Sava reservoirs, which corresponds to phases P1-P4 in the summary (Fig. 5-35). This history can be seen in lithology columns in Figure 5-36.

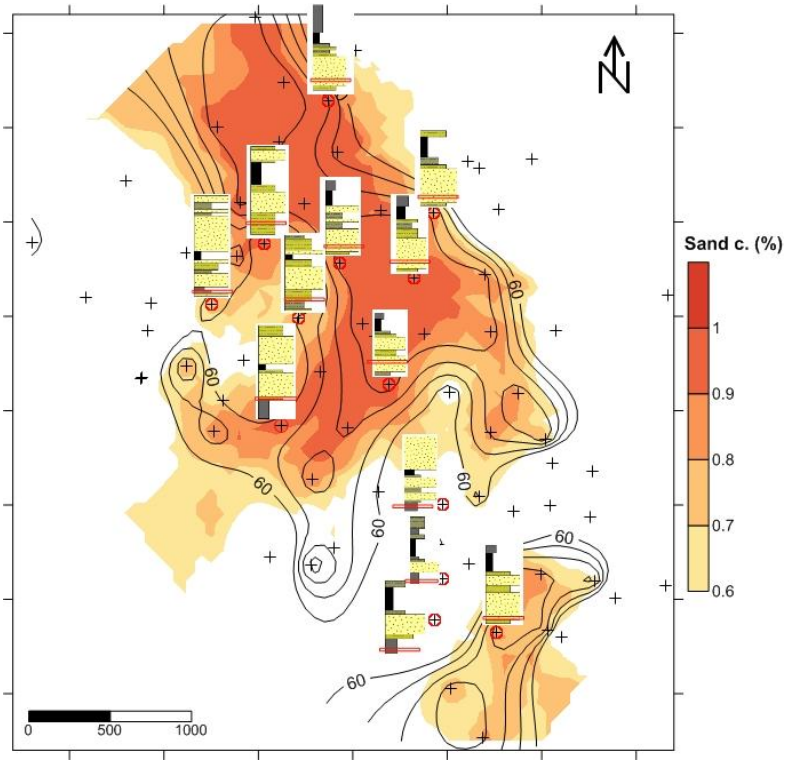


Figure 5-36: Lithology columns are displayed in the sand content contour map and the probability map of cluster C_4_m. (The horizontal surface 11m below the lower reservoir is emphasised in a red rectangle in the lithology columns).

The turbidity currents were active in the central area. In this area the lithology columns show sand bodies several meters thick (Fig. 5-36). When the turbidity current shifted to another area, an abandoned lob facies developed within the formerly active region. Here the deposits formed thinning- and fining-upward sequences. These sequences can be identified in the lithology columns. These abandoned lobs appeared between the active ones.

Figure 5-37 shows this phase where the inactive lob area is outlined with a red line. Figure 5-38 shows the location of the inactive lob according to Normark (1978). The lithology columns in Figure 5-37 show the alternations of active and passive areas. Here the massive sandstones are replaced by thin sandstones and interrupted siltstones. The difference between Figures 5-36 and 5-37 demonstrates that the active accumulation moved from the central area towards the edges and the central lob died. This is also suggested by the lower, sand content, less than 60% .

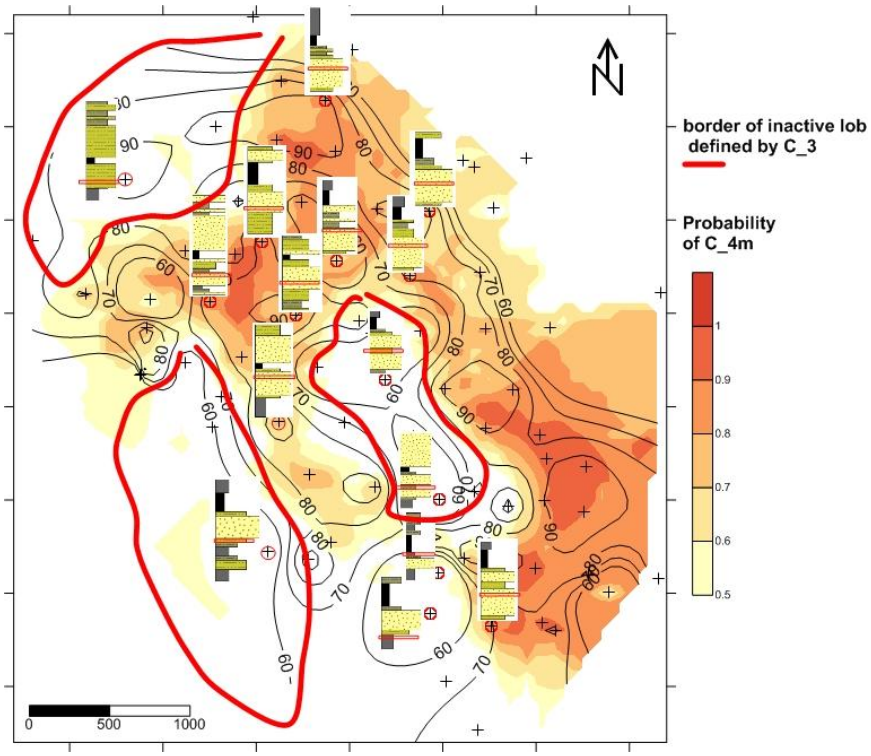


Figure 5-37: Lithology columns are displayed on the sand content contour map and by the probability map for Cluster C_4. (The horizontal surface 7m below the lower reservoir is emphasised in a red rectangle in the lithology columns and the inactive lob area is outlined with a red line).

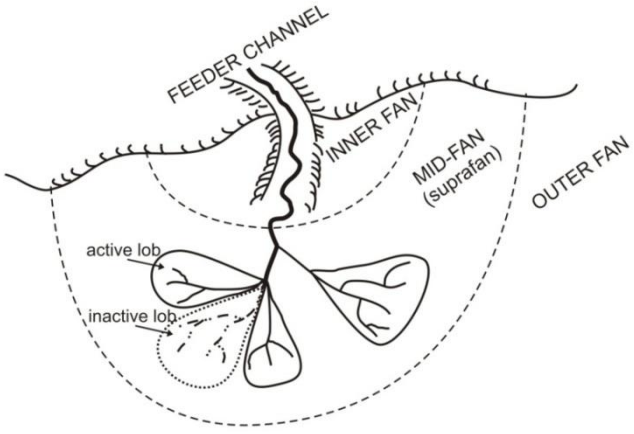


Figure 5-38: Active and inactive lobs in the suprafan area (after Normark, 1978)

The change in the type of deep-water system is demonstrated by a turbiditic flow zone in Figure 5-39. This transport area shifted to the south-west wing of the field. The lithology columns demonstrate several meter thick massive sandstones without any alternation. In the other part of the field the deposits formed thinning- and fining-upward sequences and even reached the persistent massive marl sediments.

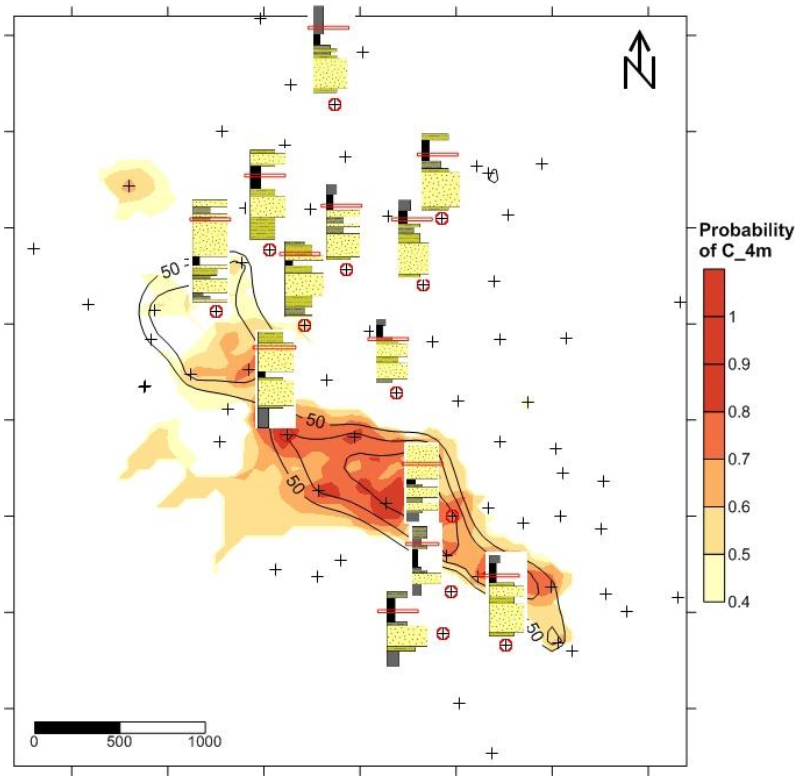


Figure 5-39: Lithology columns are displayed on sand content contour map and by the probability map for Cluster C_4_m (The horizontal surface 6m below the upper reservoir is emphasised in red)

The statistical analysis and description of clusters, the point-like extension of clusters by IK and the general shapes of clusters indicate C_1 to be a massive marl sediment with very low effective porosity and dominantly pelitic grain size. It was deposited continuously by normal lacustrine basin pelitic sedimentation during the Upper Miocene in the Sava Depression. The other cluster includes C_2, C_3 and C_4 groups, which were deposited directly from the turbiditic current. Within these clusters C_3 and C_4 represent the main sediment transport directions of the densest part of the turbiditic current. These clusters generally correspond to the Tb-Tc turbidite facies (Reading and Richards 1994).

Eventually the C_2 group represents laminated sandstones with siltstones and marl deposits accumulated between the bifurcations channels which generally correspond to the Td-Te Bouma facies.

The described clusters perfectly match the lithofacies defined for the Upper Miocene sedimentation in the Sava Depression (Vrbanac et al., 2010). It could be established that the thick-layered massive sandstones (F1) relate to Cluster C4; the facies of thin sandstones and interrupted siltstones (F2) correspond to C_3, and the facies of laminated sandstones, siltstones and marls (F3) match Cluster C_2. Finally, the facies of the massive marls (F4) were found to be the equivalent of Cluster C_1. These results are in accordance with previously published models of the depositional history of the Sava Depression (e.g., Šimon, 1980; Vrbanac, 1996; Malvić, Velić, 2011).

5.2.7 Spatial variance and lateral continuity analysis of defined depositional facies

The point-like extension of clusters and the statistical analysis made the identification of depositional facies possible, but the analysis of the lateral continuity of porosity and sand content of bifurcated channels, lobate deposits and main channels may also have valuable consequences for improving the oil production.

In this step, the contours of the above mentioned facies were first blanked in porosity and sand content grids. The contour of the geometries was based on the 0.7 probability contours of IK results. The boundary of the bifurcated channel comes from the horizontal surfaces 13-11m below the lower reservoir top. On these surfaces the clusters appear as a well-developed distributary channel system with elongated geometry (Fig. 5-40).

The channel orientation reflects a NW-SE progradation. It is 1200-1300m long and a maximum of 750m wide. Within the blanked channel geometry this facies is characterised by more than 22% porosity and more than 70% sand content properties.

The channelized system in the mid-fan area is the path for turbidity currents. It developed where the turbulence loses energy and transforms to sustained flows. Because of this, the sorting of grain sizes is better and the porosity is larger than in the proximal part or the inner-fan area. This process is also expressed in the histograms of Figure 5-41. In this figure the sand content forms a polymodal histogram and the mean is 81%. The porosity histogram shows better distribution, its mode is 23.6% and the mean is 22.48% (Fig. 5-41).

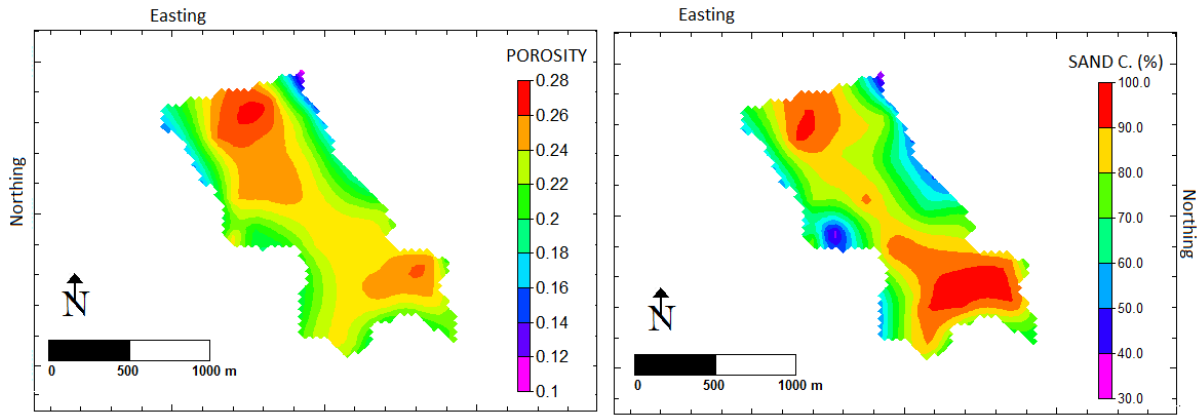


Figure 5-40: Average porosity values and sand content grid which is blanked by a contour of 0.7 probability value of principal cluster (derived from surfaces 13m and 11m from the top of the lower reservoir.)

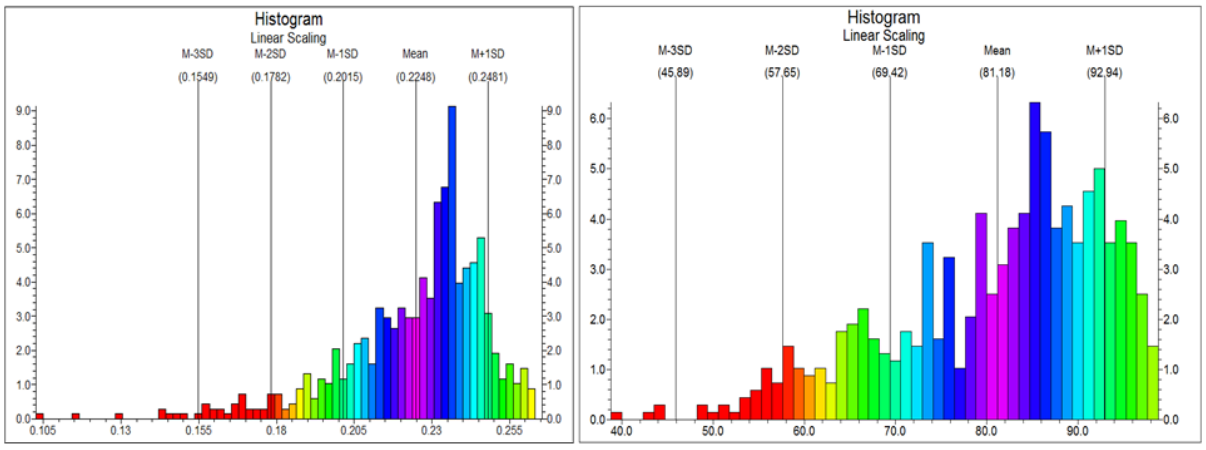


Figure 5-41: Frequency histogram of porosity values (left part) and sand content (right part) within the defined channel geometry in the lower reservoir of Sava Field

Variography analysis has been performed for both average sand content and porosity grids. The results are shown in Figure 5-42. For both variables the principal continuity direction is around 130 degrees, with 650m lateral ranges. In the perpendicular direction the range is 390m for both porosity and sand content. Since the nugget effect is low, less than one-tenth, this model can characterise almost 90% of the total variance in linear geostatistics.

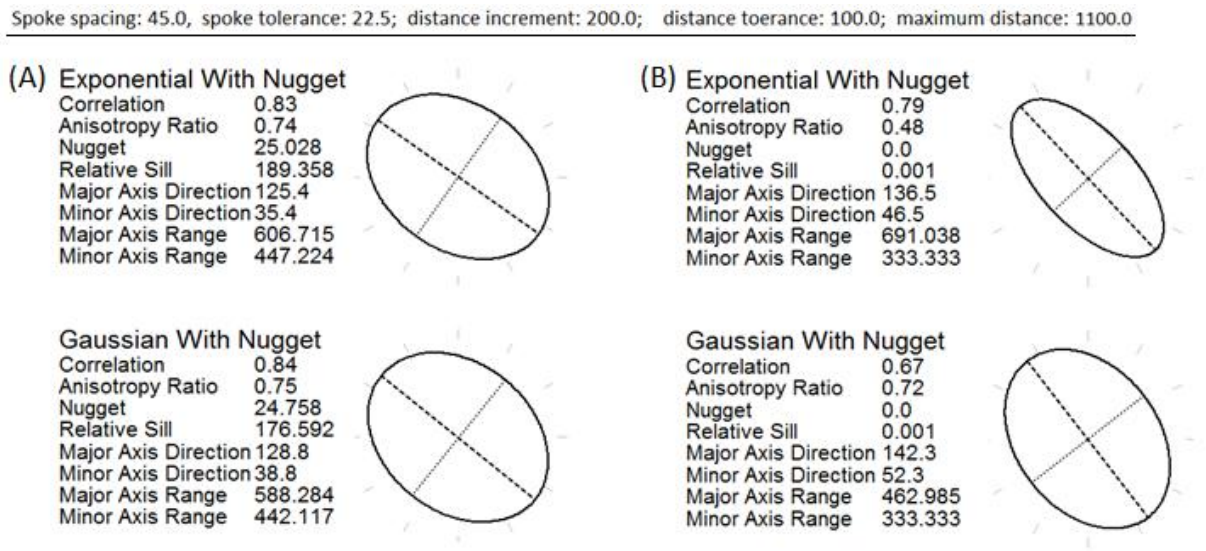


Figure 5-42: Model specifications and anisotropy ellipses (A- sand content, B- porosity) of channel system in the lower reservoir of Sava Field

The next unit analysed is that of the lobate depositional facies. In this case the surface 10m below the top of the lower reservoir was selected for blanking. Figure 5-43 shows the blanked porosity and sand content grids. This facies can be found at the distal end of the channels. They correspond to the farthest depositional area in fans where most turbidity currents die. The selected lobate unit has a geometry around 900m long and 700m wide in a NW-SE direction. In the central part, where porosity is greater than 20% and sand content is greater than 70%, the range along the minor axis decreases to 450-500m. The NW-SE direction of progradation thus coincides with the bifurcating channel outlined above.

The frequency histogram for sand content has three modes (Fig. 5-44). The major mode is very large, 91%. The first minor mode lies around the mean, 73%. The last minor mode is between them, at around 80%. This polymodal histogram describes a complex facies. It implies that the greatest sand content (major mode and minor mode) outlines channel facies within the channelized lob. The second mode characterises the sandy lobes deposit. The porosity histogram is also polymodal but less characteristic. The largest porosity values for the major mode (23.5%) correspond to the bifurcation channels, and the lobate deposit is described by 16-20% porosity. This corresponds to the minor mode.

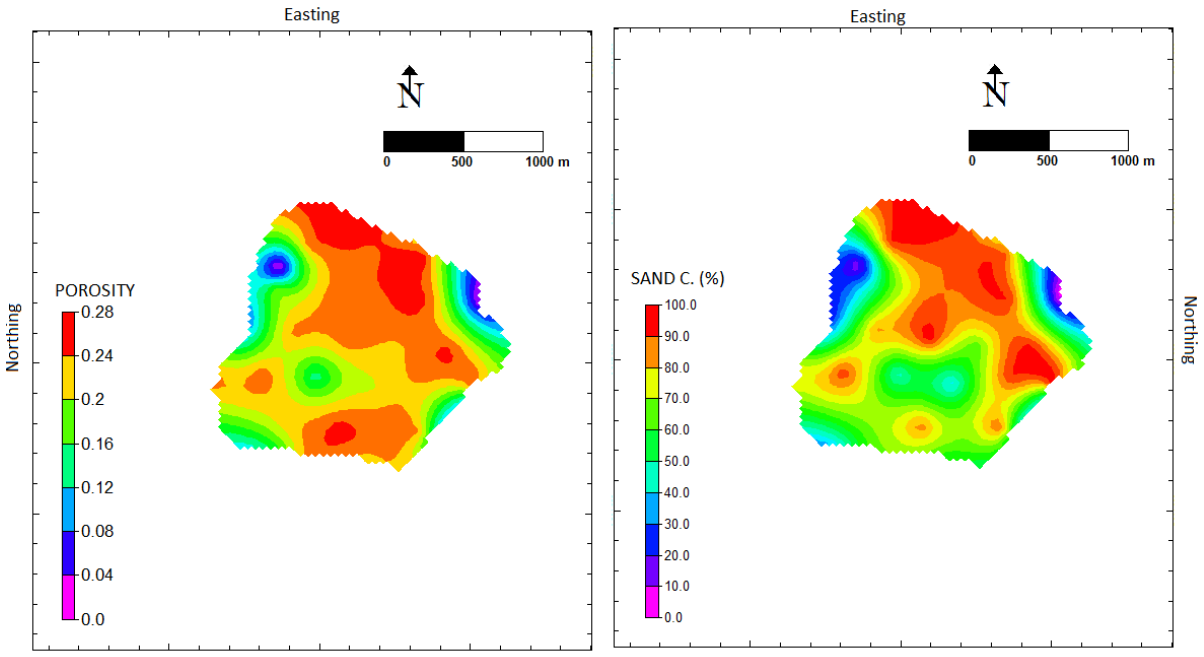


Figure 5-43: Porosity values and sand content grid blanked by the contour of 0.7 probability value of principal cluster (derived from surface 10m at the top of the lower reservoir.)

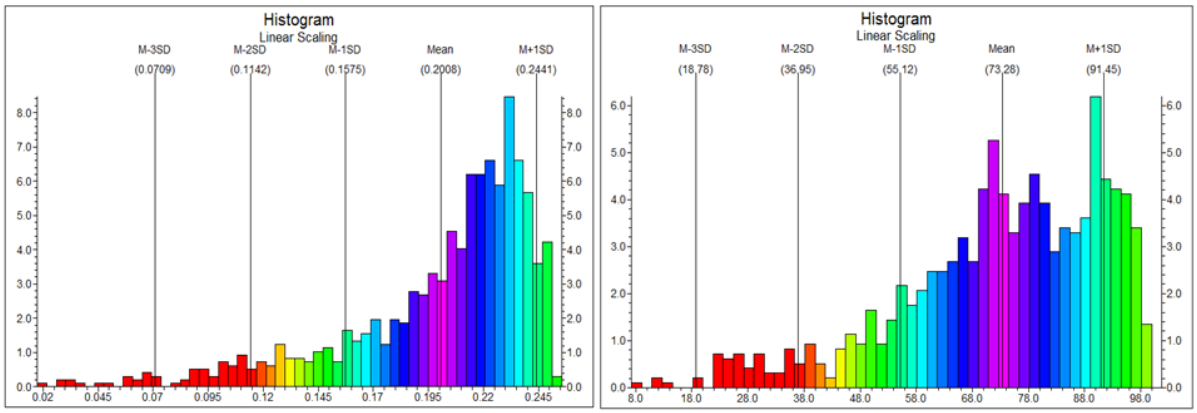


Figure 5-44: Frequency histogram of porosity values (left part) and sand content (right part) within the defined channelized lob geometry in the lower reservoir of Sava Field

The lateral continuity of sand content is stronger than that of porosity, according to the correlation of Gaussian and exponential semi-variogram models (Fig. 5-45). For both properties the main continuity direction is 160 degrees. The principal axis coincides with the channel axis in this lateral surface. The anisotropy ratios are 0.73 for sand content and 0.66 for porosity values. Usually, in the case of sand-rich systems the lobate deposits have a radial rather than rounded structure and the anisotropy ratios also support that geometry. These

elongated lobes can merge, and in this way they can form a broad sand sheet at a larger scale (e.g. Link and Welton, 1982; Heller and Dickinson, 1985; Chann and Dott, 1983; Kleerlaan, 1989; Busby-Spera, 1985; Smith, 1995; in Richards and Bowman, 1998).

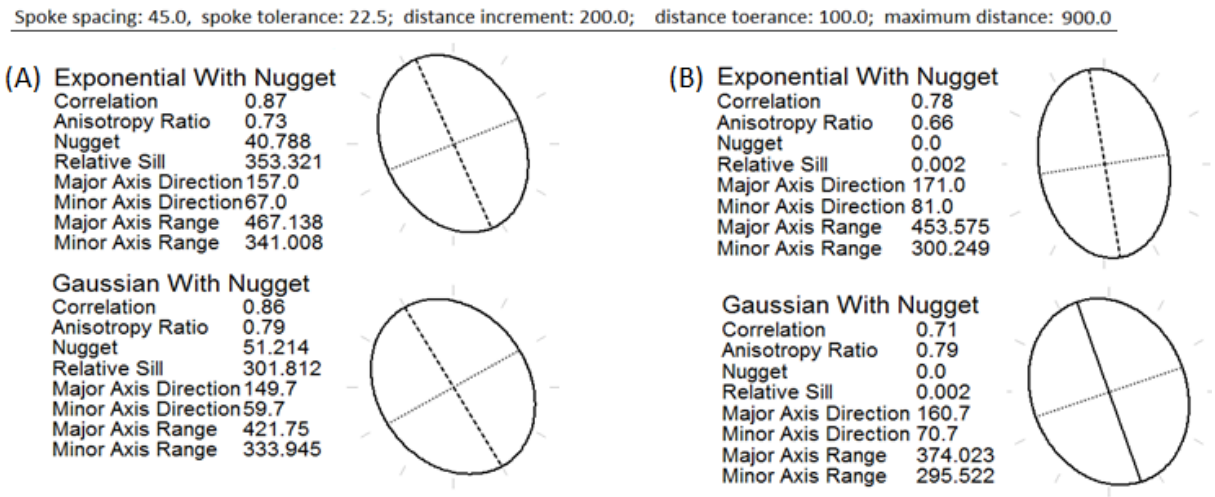


Figure 5-45: Model specifications and anisotropy ellipses (A- sand content, B- porosity) of channelized lob in the lower reservoir of Sava Field

The last physiographic unit is a channel developed near the feeder channel area in the he upper reservoir. Figure 5-46 shows the blanked porosity and sand content grids. The elongated geometry (2000m long) suggests that the direction of progradation was from NW to SE. The width of this channel is 800-900m.

Within the blanked facies geometry the average porosity values and sand content are one magnitude lower than in the channel facies in the lower reservoir. The average of the porosity values is about 18% and the average sand content is only 50%. This is also supported by the frequency distributions (Fig. 5-47). Both histograms are polymodal and reveal poorly sorted sediments.

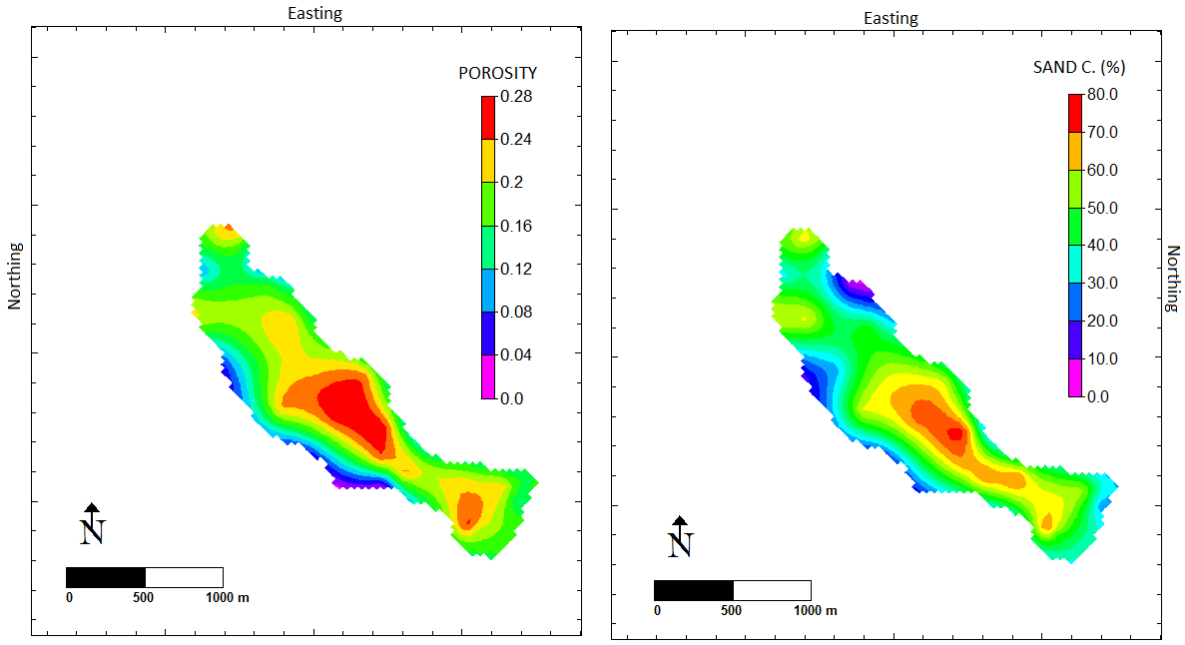


Figure 5-46: Porosity values and sand content grid blanked by the contour of 0.7 probability value of principal cluster (derived from surface 3m from the top of the upper reservoir.)

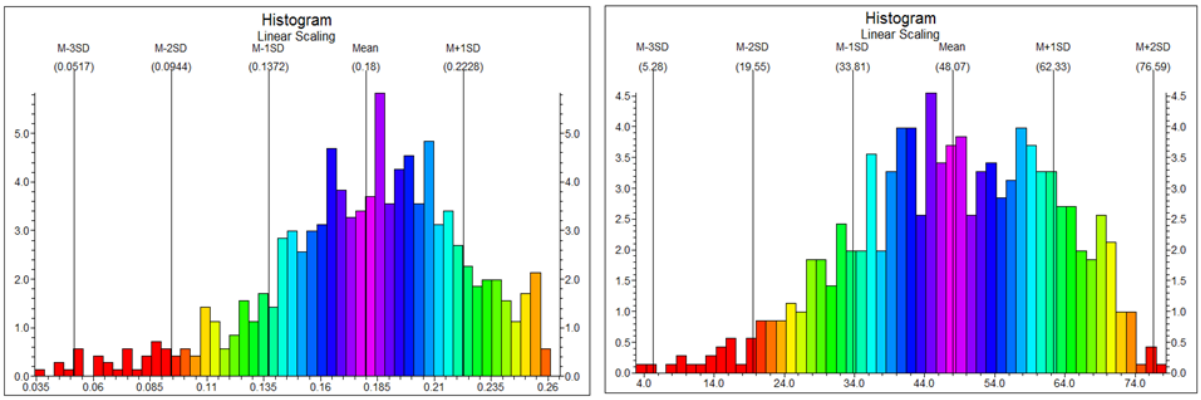


Figure 5-47: Frequency histogram of porosity values (left part) and sand content (right part) within the defined channel geometry in the upper reservoir of Sava Field

The exponential and Gaussian type models gave almost the same results. Figure 5-48 shows that both models characterise a large anisotropy ratio. The major axis direction is 142 degrees (depending on both variables), and in this direction the range is more than 610m. In the perpendicular direction it is 430m. The nugget affect suggests that this model may characterise almost 90% of the total variance.

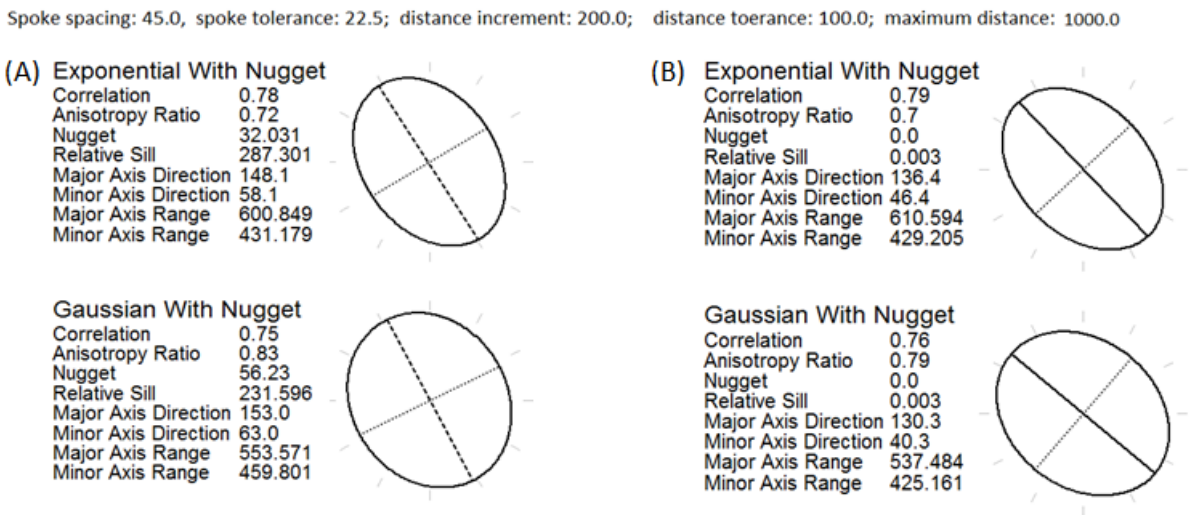


Figure 5-48: Model specifications and anisotropy ellipses (A- sand content, B- porosity) of channelized lob in the lower reservoir of Sava Field

6 COMPARISON OF THE APPLIED UNN AND K-MEANS CLUSTERING IN THE LIGHT OF THE RESULTS OF STUDY AREAS

This dissertation focuses not only on the workflow of the identification of depositional facies, but this chapter compares the applied UNN and the widely used K-means clustering procedure. It has been proved that both methods have advantages and disadvantages. K-means clustering has been a widely used process from the 1960s because it gives robust, fast and efficient processing for large and high-dimensional datasets, however several authors found that K-means sometimes failed to find any reasonable clusters. In this study the comparison of these two methods relies on the results of variance analysis, and it is particularly based on the classification results of both case studies.

In the comparison the same input dataset was used to separate cluster facies. In addition, the number of clusters was pre-defined using a K-means algorithm in the same way as for Kohonen clustering. K-means separated six clusters in the case of Szőreg-1, and four in Sava Field. The cluster centres were determined by maximisation of initial between-cluster distances. This process seeks to maximise the between-group variance (BGW) and minimise within-group variance (WGV).

Both UNN and K-means clustering break the observations into groups that are as internally homogeneous as possible, and are as different from each other as possible. As a result, the efficiency of these processes can be measured using the ratio of WGV and BGW. So, the variances were calculated separately using Equation 4.3 in both cases.

The K-means algorithm has given significantly different cluster-results especially for Szőreg-1. In the Sava reservoirs the identified clusters are similar, with few differences between the two applied methods, but based on the WGV and BGV rate of total variance, the UNN clustering gave really better solution for separation.

First, let’s focus on Sava Field. On the basis of variance calculated for all four clusters it would be difficult to say which solution is the better. Tables 6-1 and 6-2 display WGV/BGV ratios. These tables reveal significant differences between UNN and K-means. In the case of UNN clustering, porosity has the greatest weight in the separation. The objects in particular clusters are highly homogeneous due to the porosity. Based on other properties the within-group variances are only one fourth of the total variance. In the case of K-means

algorithm the porosity values separate groups poorly. According to the porosity values the homogeneity is characterised by 24% within-group variance (Table 6.2).

In the case of UNN, Clusters C_3 and C_4 were separately analysed. Both are defined as sandy deposits and as merged clusters extended by the IK process, these clusters seemed to be similar using nonparametric statistics. In addition, IK suggested that these cluster facies can be individually defined sedimentologically as active (C_4) and inactive (C_3) lobate deposit. The extended Cluster C_3, with a lower porosity value and sand content, outlined the inactive lob area of the midfan in the lateral surface 7m below the lower reservoir top (Fig. 5-29).

The analysis of WGV and BGV for these two clusters revealed considerable differences. Table 5-4 collected the results of variance analysis for UNN and Table 6-3 shows the variances of clusters by K-means. In the case of K-means, the variance analysis reinforced the idea that these clusters were not separated in an acceptable way, since WGV could characterise almost 96% of the total variance for SH content and 88% for porosity (Table 6-3).

Table 6-1: Comparison of WGV and BGV for all clusters generated by UNN in Sava Field

NN all cluster	WGV	BGV	TV	WGV / TV rate (%)	BGV / TV rate (%)
Sw	177.45	506.72	684.18	25.94	74.06
SH	282.64	768.08	1050.72	26.90	73.10
POR	0.0006	0.0059	0.0065	9.49	90.51

Table 6-2: Comparison of WGV and BGV for all clusters generated by K-means in Sava Field

K-m all cluster	WGV	BGV	TV	WGV / TV rate (%)	BGV / TV rate (%)
Sw	111.64	240.17	683.58	16.33	83.86
SH	136.64	913.83	1050.47	13.01	86.99
POR	0.0004	0.0049	0.0065	24.18	75.82

Table 6-3: Comparison of WGV and BGV for Clusters C_3 and C_4 generated by K-means in Sava Field

K-m C_3 and C_4	WGV	BGV	TV	WGV / TV rate (%)	BGV / TV rate (%)
Sw	97.61	313.58	411.18	23.74	76.26
SH	114.43	5.13	119.56	95.71	4.29
POR	0.0007	0.0001	0.0008	88.26	11.74

In the comparison of the two clustering processes the cluster members from K-means separation were shown on classed post map in Figure 6-1 containing four lateral surfaces as examples. These maps showed the sand content and the overlay results of K-means algorithm. Sand content was used in this analysis, since its contours followed the defined UNN cluster facies very well. Using K-means Cluster C_4 was clearly characterised as a massive sand deposit (marked by a black symbol in the maps of Figure 6-1). The other two clusters contained data points that UNN detected as C_4, or interchanged C_2 and C_3 (marked with a red symbol in maps of Figure 6-1). This may be the result of misclassification. In this case the within-group variance needs to be relatively low. This calculation also verifies that these clusters are not homogenous enough (Table 6-4). K-means algorithm usually identified data points with large porosity and sand content as Cluster C_2 and C_3 if the sample originated from the upper reservoir. One possible reason for statement is that there are two characteristically different depositional facies in the study area: the lob systems channelized by turbidite current and the quiet water deposits. Using UNN clustering, the first depositional facies were divided into more sub-facies, which were not seen in either reservoir. In the upper reservoir the cluster facies outlined only channel fill deposit and basin floor sediment. This channel facies was characterised by a one magnitude lower value in porosity and sand content than the lower reservoirs. This may be why K-means misclassified several samples (Fig. 6-1).

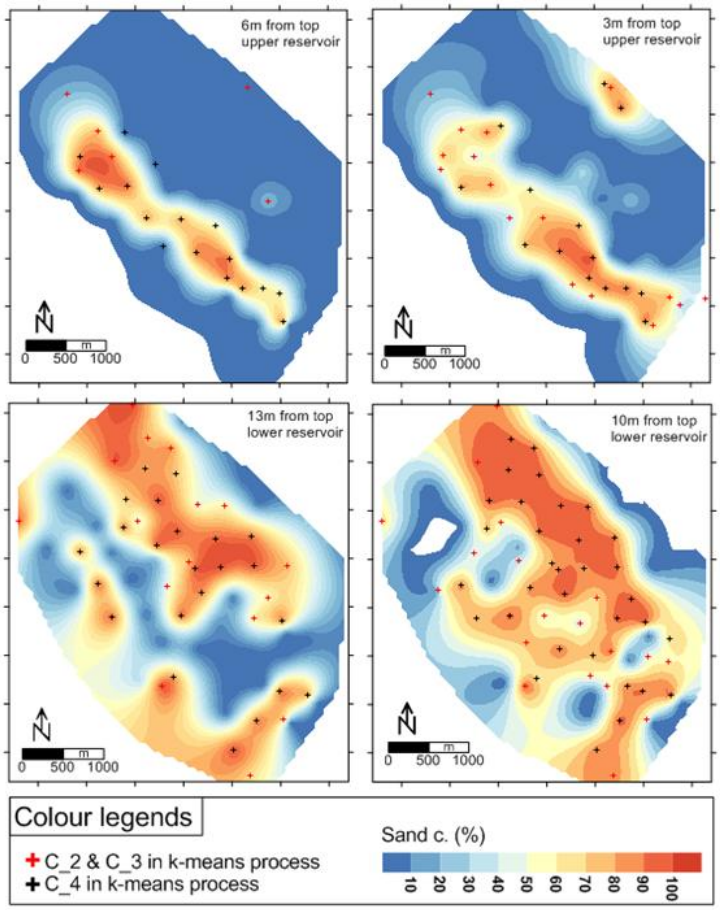


Figure 6-1: The classed post maps of clusters separated by K-means algorithm overlapped with the sand content map of several surfaces of reservoirs in Sava Field

Table 6-4: Comparison of WGV and BGV for Clusters C_2 and C_3 generated by K-means in Sava Field

K-m C_2 and C_3	WGV	BGV	TV	WGV / TV rate (%)	BGV / TV rate (%)
Sw	183.48	155.82	339.30	54.75	45.92
SH	181.76	233.26	415.03	43.79	56.20
POR	0.0019	0.0001	0.0020	93.76	6.23

In the Szőreg-1 study, there are also significant differences between the two clustering techniques due to the variance analysis. Six clusters were generated by K-means, as in the case of UNN, but these clusters did not correspond directly to clusters by UNN. Only the pairs of pooled clusters corresponded more or less to UNN-clusters. These pooled clusters were as follows: C_1 - C_2 as sedimentation from quiet water defined by UNN; C_3 - C_4 as distributary mouth bar by UNN clustering; C_5 - C_6 as bifurcation channel by UNN-clusters. The relationship between UNN and K-means clusters was thus recognisable. In a comparison of pooled clusters C_1-C_2 with the clusters of quiet water deposit (defined

clusters by UNN and IK) or the merged C_5-C_6 with the bifurcation channel (identified as UNN clusters), the relationship was stronger than for the C_3-C_4 pair. The variance analysis supported this observation (Table 6-5 - 6-7), however this merging could not yield enough homogeneous clusters compared to the UNN clusters (Table 5-1). In the case of the C_3-C_4 pair, the variance analysis showed the greatest differences (Table 6-6).

Clusters C_3 and C_4 by K-means were visualised overlying with the UNN result maps (Fig. 6-2). The main difference could particularly be detected in on Horizon#2 (bifurcation channel surface in the right part of Figure 6-2) where the number of facies increased. Conversely, in Horizon#1 (distributary bar surface in the left part of Figure 6-2) the correspondence is better.

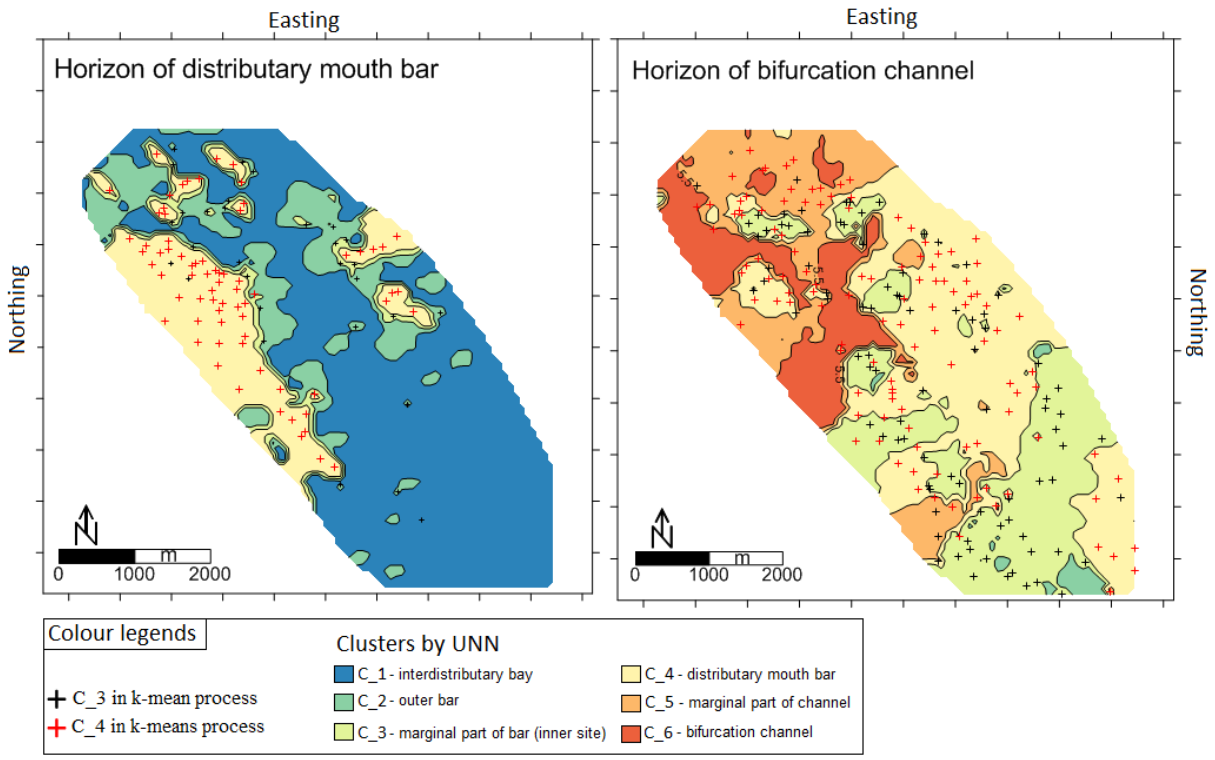


Figure 6-2: Classed-post maps of Clusters C_3 and C_ 4, overlapped with results map of extended clusters by UNN (The classed-post maps display clusters which are separated by K-means)

Table 6-5: Comparison of WGV and BGV for Clusters C_1 and C_2 generated by K-means algorithm in Szőreg-1 reservoir

K-m	WGV	BGV	TV	WGV / TV rate (%)	BGV / TV rate (%)
S.r.	151.29	280.56	431.85	35.03	64.97
POR	7.77	6.26	14.03	55.38	44.62
HK	209.26	16.88	226.14	92.54	7.46

Table 6-6: Comparison of WGV and BGV for Clusters C_3 and C_4 generated by K-means algorithm in Szőreg-1 reservoir

K-m	WGV	BGV	TV	WGV / TV rate (%)	BGV / TV rate (%)
S.r.	81.36	131.67	213.03	38.19	61.81
POR	11.27	2.748	14.02	80.40	19.60
HK	1783.89	105.42	1889.32	94.42	5.55

Table 6-7: Comparison of WGV and BGV for Clusters C_5 and C_6 generated by K-means algorithm in Szőreg-1 reservoir

K-m	WGV	BGV	TV	WGV / TV rate (%)	BGV / TV rate (%)
S.r.	102.17	70.61	172.79	59.14	40.86
POR	5.60	4.03	9.63	58.12	41.88
HK	11937.38	33948.92	45886.29	26.02	73.98

The comparison of data separation using the Kohonen neural network and K-means algorithm demonstrated that UNN is able to recognise clusters as facies even in such situations where K-means clustering techniques fail to find any reasonable depositional units. This can be explained by the difference in the separation algorithm.

In the iteration of K-means algorithm there are main two steps, (i) assign each data point to the nearest mean, and (ii) move the "means" to centres of every single cluster. The number of iterations is previously fixed. The problem is that the separation is almost independent on the length of iteration since clusters can be characterised by the features of the first data vectors and in this way they have greater weights in the definition of centres.

In contrast, the Kohonen network can modify the centres of clusters through the modification of weight according to the training rate and using the test set. Thus, an advantage

of using the Kohonen network is that its cluster-forming ‘capacity’ is self-regulated, which is why it is more efficient than K-means clustering.

Of course there are some advantages to using the K-means technique. With a large number of variables this method may be computationally faster than other clustering techniques, but only in case of fewer clusters.

Other weak points of K-means are that the defined clusters may differ in size, density, and non-globular shapes of dataset. If the clusters are not chained, K-means algorithm may produce tighter clusters, as did UNN. If there is previous knowledge about cluster features or centroids from an analogue area, the partition process may be efficient, robust and very fast.

Both techniques outlined above also have disadvantage. Since it is difficult to predict the number of clusters, an auxiliary process (e.g. variance analysis, statistical test) or precognition is generally necessary.

7 DISCUSSION AND INFERENCES

This dissertation has demonstrated, through two case studies, a workflow for facies identification in clastic depositional environments. The workflow was based on unsupervised neural network clustering and probabilistic extension of cluster members.

UNN separated the input datasets which contained differently measured and interpreted well-log parameters and categorical variables for lithology description. The aim of clustering was to separate subsets that were potential depositional facies. The cluster facies recognition was supplemented by statistical interpretation and lateral extension.

For the Szőreg-1 reservoir two well-known horizontal surfaces were selected for the demonstration of the workflow. Horizon#1 reflects the phase of the development of distributary mouth bars in a bay. Horizon#2 represents the phase when bifurcation channels break through distributary bars. In these lateral surfaces six clusters were defined as depositional facies:

(i) Facies C_1 represents the sedimentation from the quiet water in the interdistributary bay area.

(ii) Facies C_2 is the outer bar area of a distributary mouth. It also represents still water sedimentation.

(iii) Facies C_3 is identified as the marginal parts of distributary mouth bars.

(iv) Facies C_4 is the main body of the distributary mouth bars. Its accumulation is characterised by a channel system. The lower horizon reveals major and a minor bar development from SW and NE. These mouth bars are involved due to progradation which is revealed in the upper lateral surface.

(v) Facies C_5 is identified as the marginal part of the bifurcation channel. It appears around the dendriform channel geometry.

(vi) Facies C_6 defined the bifurcation channel with dendroid shape. This channel developed from a SW direction.

C_4 and C_6 clusters outlined the two most characteristic depositional sub-environments as the distributary mouth bar and bifurcation channel. The probability contours of Cluster C_4 outlined a 6.5-7km wide kidney-shaped bar. The progradation of this major

bar is from SW and from the opposite side a minor mouth bar developed which reached a maximum 3-3.5km width. The shape of C_6 was approximately 5km in length with an approximately 500m wide channel network. The identified facies showed two main phases of the depositional history of Szőreg-1 reservoir (Geiger, 2003).

The second case study came from the Sava basin. Two reservoirs from this area were represented using the detailed workflow. The two sequences analysed were built up by Upper Miocene marls, siltstones and sandstones. The latter two lithotypes had been deposited by periodic turbidity currents over the entire depression.

Both reservoirs were cut by 1m thick surfaces, parallel with the top of the reservoir in a stratigraphic coordinate system. Between any two cutting surfaces, the petrophysical data (from well-log interpretation) of the 1.0 m thick vertical intervals was given by the average of the data falling to these intervals. This dataset was the input for UNN.

Four clusters were separated using the UNN clustering technique and these were identified as lithological facies. The depositional facies were also defined according to their spatial pattern. The analysis demonstrated that UNN can segregate the different clastic lithofacies which are deposited in different sub-environments of lacustrine turbiditic flows. These facies correspond perfectly to the lithofacies which were defined as typical Upper Miocene sedimentation in the Sava Depression (e.g., Šimon, 1980; Vrbanac, 1996; Vrbanac et al., 2010; Malvić and Velić, 2011). The facies are as follows:

(i) C_1 is identified as massive marly sediment deposited by still water sedimentation. According to Vrbanac et al, (2010) it corresponds to (F4) lithofacies.

(ii) Facies C_2 was defined as laminated sandstone, siltstone and marl deposits which correspond to (F3) by Vrbanac et al., (2010). This facies was deposited directly from low-density turbidity current. According to IK results, this cluster appears between the bifurcation channels and lobate sediments. It is defined as the inactive part of a fan system and generally corresponds to the Td-Te Bouma sequence.

(iii) Facies C_3 is the lithofacies of thin sandstone and interrupted siltstone which corresponds to (F2) by Vrbanac et al., (2010). As part of low-density turbidites this sedimentation was regarded as lob-type deposits.

(iv) Facies C_4 is identified as thick-layered massive sandstone which corresponds to (F1) by Vrabanac et al., (2010).

The sedimentary environments were identified according to the characterised cluster facies. In general, the studied rock body of Sava Depression appears to be a mid-fan (suprafan) area of a sand-rich submarine fan system. The process of facies migration suggests that this submarine fan system prograded from NW to SE. During the progradation the lobate sediments shifted laterally. The lower reservoir represents the distal and upper reservoir of the proximal part of mid-fan in the submarine fan system.

The lateral distribution of Clusters C_3 and C_4 represents the main sediment transport directions of the low-density turbidite within the sand-rich submarine fan system. These clusters generally correspond to the Bouma Tb-Tc sequences (Reading and Richards, 1994). The fourth facies outlined the axes of turbiditic flow and lobate deposit according to bifurcation channel shape and the fan-like pattern in the distal part of the submarine fan system. Custer C_3 is also defined as lobate deposit. In the surface 7m below the top of the lower reservoir the probability map of clusters showed that C_3 may be identified as inactive lobate sediment which appeared between the active lobs sedimentation by C_4.

Three physiographic units were characterised from the selected 1-2m thick intervals of the reservoirs:

(i) a channel with an elongated and bifurcated shape in the direction of the main progradation; it is maximum 1200-1300m long and 750m wide. This physiographic unit is characterised by high quality thickly-layered massive sandstones with greater than 22% porosity and more than 70% sand content.

(ii) a lobate deposit with radial pattern which is about 700m in the major axis (NW-SE) and at maximum 500m wide in the perpendicular direction. According to the anisotropy ratios the lob-system has a radial pattern which is a common characteristic in sand-rich systems. This physiographic unit is characterised by greater than 20% porosity values decreasing toward the wings. The sand content is greater than 70% and decreases towards the wings, where it is about 50-60%.

(iii) an elongated channel without bifurcation representing the proximal part of the suprafan area. It has an approximately 2000m long axis from NW to SE, at maximum 800m wide. According to porosity and sand content, this unit has a clearly different character compared to the channel facies in the lower reservoir. The elongated pattern has shoestring shape, and the one magnitude lower value for porosity and sand content indicates a position

closer to the feeder channel. Here the porosity changes between 16-20% and the average sand content is 50-60%.

On the basis of the characterised physiographic units, the heterogeneity of this reservoir is very low, so the reservoir continuity and connectivity is very good laterally. In the upper reservoir, the connectivity is very good, probably also vertically since the elongated channel is not characterised by lateral migration.

The demonstrated workflow was based on data clustering but it was not classical, such as K-means. This dissertation tried to show that the applied UNN is able to recognise clusters in those situations where K-means clustering techniques fail to find any reasonable cluster facies. Another advantage of using UNN in facies analysis is that its cluster-forming ‘capacity’ is self-regulated, so it is more efficient than ‘classic’ clustering.

SUMMARY IN ENGLISH

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 the objects in the sample, 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.

There are many, widely used clustering techniques but in this dissertation the separation of subsets, based on a neural network approach, the so-called Kohonen network was demonstrated. This method was applied because neural network cluster separation requires associative ability, learning ability and non-linear separation techniques. Often, a database cannot be divided in a linear way. This may be the reason that in some cases the separation procedures misclassify at relatively large rates. Using a suitable non-linear transformation these linked clusters can be separated.

The Kohonen neural network is a non-linear separation technique. The K-means approach is regarded the most similar to the Kohonen clustering. But the K-means is one of from classical clustering techniques. Some papers have dealt with their comparison using statistical tests and found that K-means sometimes failed to find any reasonable clusters. In this study a comparison of these two methods relies on the results of variance analysis. Clustering seeks to minimise within-group variance (WGV) and maximise between-group variance (BGV) and it can rarely reach a substantial difference between them. The difference between WGV and BGV can demonstrate the suitability of cluster results. The relatively low WGV and larger BGV mean that cluster analysis has a number of heterogeneous groups with homogeneous contents.

A comparison of data separation by Kohonen neural network and K-means algorithm pointed out that: (1) UNN is able to recognise clusters as facies even in such a situation where K-means clustering techniques fail to find any reasonable depositional units; (2) one of the

advantages of using UNN in facies analysis is that its cluster-forming ‘capacity’ is self-regulated, which is why it is more efficient than ‘classic’ clustering.

The applied Kohonen neural network is an unsupervised neural network. This is also an analogy of the manner by which the human brain can logically arrange data, and new information. This is a kind of associative memory, which supports the systematic organisation process without any external help. To this end it was used as a clustering process to separate the cluster facies in the data space.

The results of clustering mean only categorical information in well points. The question is how can we spatially interpret and extend these point-like results? The workflow described in the dissertation demonstrates the probabilistic approach of the extension. The process uses indicator kriging. IK uses a discretized form of global probability distribution derived from sample points, and the procedure uses discretization. The cluster categories are the discretized form. According to the spatial correlation structure measured between categories, this method estimates the probability distribution functions at any unsampled locations. The goal of IK is thus the estimation of probability for each input cluster category at every grid node. A particular grid node should be assigned to those cluster memberships which have the greatest estimated probability. Another choice is a map showing the change of probability belonging to a particular cluster membership. It is necessary to find a reasonable probability-contour which can outline the shape of a supposed physiographic unit well.

Clustering any objects in the geographical space (whether UNN or K-means clustering) results in disjunctive sets of spatial points. Using lateral extension by IK these clusters can be defined spatially as interfinger facies. This also implies that the applied methodology does not contradict the definition of either clustering nor depositional facies.

The identification of depositional facies was based on the geometry of laterally extended clusters and their statistical characters. The different depositional environments can be characterised by special (although, not necessarily different) rock-body morphology. Thus, in this work the rock body geometry expressed by the probability contours was interpreted in terms of the depositional facies. These contours can be used as a blank-polygon. In this case, the differently defined physiographic units are blanked, and within the contours it is possible to analyse the lateral continuity and spatial variance of the porosity and sand content.

Two case studies from different clastic sedimentary environments demonstrate the workflow. The first represents a delta plain environment. The second one shows a deep water

turbidity system. In addition to these, the type of available variables and the amount of data were also different in these two cases.

The first study area was located in the Szőreg-1 reservoir, Algyő Field, south east Hungary. In this reservoir, UNN was applied in two depositional sub-environments: (1) emerging distributary mouth bar; and (2) prograded bifurcation channels. Their corresponding stratigraphic positions were 34-35 m and 24-27m below the top argillaceous marl of the rock body.

The second case study involved the Sava Field, north Croatia. The focus of this study was only two selected reservoirs located one above the other in the reservoir series. The two sequences analysed were built up by Upper Miocene marls, siltstones and sandstones. The latter two lithotypes had been deposited by periodic turbidity currents over the entire depression.

In both cases the depositional rock bodies were transformed into a stratigraphic coordinate system. In this coordinate system only the vertical coordinates were changed. It is measured from the top of the rock body, from the bottom of the massive clay-marl deposit which separates the reservoir units. This system was sectioned by surfaces being parallel to the top, and thus, it is possible to follow the change of depositional facies in a small-scale analysis.

In the case of Algyő Field, six NN-facies were defined. According to BGV and WGV these were clearly well separated. These clusters seemed to be very heterogeneous with homogeneous content and WGV was only about one tenth of the total variance. The lateral extension of these clusters was based on IK. Due to the statistical characteristics and the spatial pattern the clusters were defined as follows: (i) interdistributary bays; (ii) outer bars facing to still water sedimentation; (iii) marginal parts of distributary mouth bars developed as the result of channel and bay interactions, (iv) main body of the distributary mouth bars with 6.5-7 km width and kidney-shape geometry; (v) marginal part of the channels involved and (vi) bifurcation channels with dendroid geometry. This channel pattern was characterised by a length of approximately 5km and a channel network approximately 500m wide

The facies identified showed two main phases of the depositional history of Szőreg-1 Reservoir. In the first phase discrete major and minor mouth bars formed. During the second phase bars prograded and a bifurcating channel cut into their rock body.

For Sava Field, four NN-facies were defined. Cluster facies were extended laterally using a probabilistic approach to several selected horizontal surfaces of the whole reservoirs.

In some horizontal surfaces Clusters C_3 and C_4 were extended as pooled clusters. These clusters together represent a main depositional facies due to the porosity values and sand content. The C_3 cluster in some lateral surfaces has isolated locations. In addition, these clusters were pooled because the variance analysis showed that these groups are not clearly heterogeneous in the geographical space.

According to the statistical character and the spatial geometry of clusters the following four facies were defined: (i) Massive marls with low porosity representing still water sedimentation. (ii) Low-density turbidity currents resulting in sandstones interbedded by siltstones and marls. They were accumulated between the bifurcating channels of a fan system. (iii) Thin sandstone and interbedded siltstone of low-density turbidites were regarded as belonging to lob-type deposits. This facies describe the passive-lob area. (iv) Massive sandstone was probably deposited at the axes of turbidity channels of a sand-rich turbidity fan. This facies was outlined by elongated shoestring and dendroid geometry. The sedimentary environment was identified according to the depositional facies characterised. In general, it is proposed as the mid-fan area of a sand-rich submarine fan system. The processes of facies migration suggest that this submarine fan system prograded from NW to SE. During the progradation the lobate sediments also shifted laterally.

The workflow demonstrated in the dissertation may aid depositional facies analysis and the identification of depositional environments. This methodology may also support object-based modelling and define training images. The weights belonging to clusters derived from UNN process are applicable to the analysis of similar depositional environments if the identification of facies is based on neural network classification.

SUMMARY IN HUNGARIAN

A fácies elemzés területén már számos szakirodalom és cikk foglalkozott az elmúlt néhány évtizedben a különböző többváltozós statisztikai módszerek alkalmazásával. A munkák jelentős része szintén valamely csoportosító vagy klaszterező eljárásra támaszkodva tette lehetővé a mintatér felosztását, hogy abban valamely szempont szerint homogén litológiai egységeket definiáljanak.

A dolgozat témája szintén az üledékes fácies elemzéséhez kapcsolódik. Jelent tanulmány a klaszterező módszer alkalmazásának és a szeparált csoportok valószínűségi alapú laterális kiterjesztésének kombinálásával valósítja meg az üledékes egységek azonosítását. A dolgozatban bemutatott módszer továbbá kiegészítő statisztikai összehasonlításra, illetve variancia analízisre egyaránt támaszkodik.

Számos, széles körben elterjedt klaszterező eljárás ismert, ennek ellenére a disszertációban leírt eljárás egy neurális háló alapú osztályozáson alapul, jelen esetben ez az ú.n. Kohonen háló. A mintatér felbontására azért egy neurális háló alapú eljárás lett alkalmazva, mivel a csoportok kialakításában az asszociatív képességnek, a tanítási elvnek és a nem-lineáris szeparációs technikának volt szerepe. Gyakran előfordul, hogy az adathalmaz nem szeparálható lineárisan klaszterekre. Ez lehet az oka, hogy egyes esetekben a csoportosító vagy egyéb szeparációs eljárások nagy tévesztési rátával hajtják végre a feladatot. Egy megfelelő nem-lineáris transzformációt alkalmazva, a lineárisan nem szeparálható, összefűzött klaszterek a tulajdonságtérben lineárisan szeparálhatóvá alakíthatók.

Kohonen neurális háló alapú klaszterezést leginkább a K-means típusú csoportosítással hozzák összefüggésben. A két eljárás összehasonlítását elsősorban tesztek alapján végezték és írták le. A disszertáció lehetőséget adott a két módszer gyakorlati szempontú összehasonlítására. Ez az összehasonlítás a csoportokhoz tartozó külső-belső szórások arányára támaszkodott. Ez a fajta összehasonlítás lehetővé tette a klaszter definíciójából következő csoporton belüli homogenitás és a csoportok közötti heterogenitás kiértékelését. A két szórás közötti arány fejezi ki a klaszterezés eredményének hatékonyságát. Azaz, minél alacsonyabb a belső szórás és nagyobb a külső szórás aránya, annál heterogénebb klasztereket sikerült elválasztani relatíve homogén mintaelemekkel. A tanulmányban bemutatott példák alapján belátható, hogy a K-means, mint gyors eljárás alkalmas ugyan kis klaszterszámmal előzetes információt adni a fáciesek típusairól, viszont azok teljesen

elfogadható szeparálására nem képes. Ez azt is jelenti, hogy a Kohonen neurális háló képes felismerni egy fáciest még azon helyzetben is mikor a K-means algoritmus már nem képes az adott üledékes egység ésszerű besorolására.

Az alkalmazott Kohonen háló egy felügyelet nélküli eljárás a neurális hálók széles körében. A csoportosító eljárások közül maga a klaszterezés is tanító halmazpárt nélkülözve alakítja ki a szeparált részhalmazokat a mintatér struktúrája, vagy mintázata alapján. Ez az emberi agy asszociációs képességének az analógiája, hiszen az emberi agy is a felismerésen és a hasonlóságok keresésén alapulva képes objektumokat besorolni és így klasztereket kialakítani.

A kialakított klaszterek, mint pontszerű információk jelennek meg a térben. A bemutatott módszer második pontja azzal a kérdéssel foglalkozik, milyen módon lehet ezeket a kialakított klasztereket kiterjeszteni, és a klaszter fáciesek közötti határfelületet megadni. Ehhez az indikátor krigelés nyújtott megoldást.

Az IK lehetővé teszi a diszkrét kvalitatív információk laterális kiterjeszthetőségét is. Az eljárás minden egyes gridpont körül a valószínűségi eloszlás becslését végzi a térbeli korrelációs struktúra alapján. Miután a klaszterekhez tartozó indikátor változókat definiáltak, az IK becslést ad a klaszterek laterális kiterjedését illetően azáltal, hogy minden egyes rácsponthoz hozzárendel csoportonként egy-egy valószínűséget. A valószínűségi értékek alapján már könnyen definiálható a legjellemzőbb klaszter. A másik alternatíva, ha a valószínűségeket kontúr térkép jeleníti meg minden egyes csoport esetében. Ebben az esetben fontos megtalálni azt a valószínűségi értéket, amely képes körvonalazni az adott üledékes környezet jellemző geometriáját.

A klaszterek matematikailag diszjunkt halmazokként értelmezhetők, a valószínűségi kiterjesztés alapján azonban már térbelileg összefogazódott fáciesenként jeleníthetők meg. Ez egyben azt is jelenti, hogy az alkalmazott eljárás nem sérti sem a klaszterezés matematikai, sem a fácies üledékföldtani definícióját.

A kiterjesztett klaszterek geometriája és a csoportok statisztikai jellemzői alapján lehetőség van a csoportok azonosítására. Az üledékes környezetek jellemző térbeli megjelenéssel, alakzattal írhatók le, még ha azok nem is szükségképpen különbözőek. Emellett, a megfelelően kiválasztott valószínűségi értékek kontúrjai vágási poligonként is alkalmazhatók. A poligonnal körülhatárolt üledékes al-környezet porozitás és homoktartalom tulajdonságainak laterális folytonosságát lehetett így vizsgálni.

A disszertáció a részletezett munkafolyamatot és kiértékelési szempontokat két, eltérő tanulmányterület bevonásával szemléltette. Az első tanulmányban egy deltasík felhalmozódás, míg a második esettanulmányban egy mélyvízi hordalékkúp elemzése történt meg. Emellett a rendelkezésre álló adatok mennyisége és az adatok típusai is különböztek.

Az első mintaterület a DK-Magyarországon található Algyő mezőből származik. A módszertani eljárás kiértékelésében nem a teljes üledékes rendszer, hanem csak a Szőreg-1 rezervoárból kiválasztva két felhalmozódási egység lett kielemezve. Az üledékes felhalmozódás vertikumából kiválasztott egyik egység egy torkolati zátonytest üledékes sorozatát, míg a másik a zátonytesten megjelenő elágazó medrek rendszerét foglalja magába.

A második mintaterület mélyvízi üledékes környezetet képvisel. Ez a mintaterület a Száva medencében, É-Horvátországban található. Maga a tanulmány fókuszában lévő terület csak a medence egyik rezervoár sorozatának két, egymás felett elhelyezkedő tagját öleli fel.

Az elemzésbe bevont két egységet Felső Miocén korú agyagok, aleuritok és homokkövek építik fel, amelyek szakaszosan ismétlődő turbidit áramlatok hatására halmozódtak fel.

Mindkét tanulmány esetén az üledékes kőzet egy sztratigráfiai koordinátarendszerbe lett elhelyezve. Minden esetben a rezervoárokat elválasztó kitartó agyagmárga talpa, azaz a rezervoár tetőszintje volt az a kitüntetett időhorizont vagy felület, amellyel párhuzamos felszínek mentén a kőzettestek fel lettek szeletelve. Így lehetőség volt az üledékes kőzetek kisléptékű vizsgálatára és üledékes fáciesek vertikális változásosainak nyomon követésére is.

Az algyői esettanulmányban a neurális háló alapú csoportosítás alapján hat klaszter fácies lett definiálva. Az elkülönített klaszterek a hozzájuk tartozó külső és belső szórás alapján egyértelműen jól definiáltak, minden bemenő változó alapján relatíve homogének, azaz a belső szórás kis arányt képvisel csak a teljes minta varianciájában. Az klaszter fáciesek kiterjesztése indikátor krigeléssel történt. A statisztikai jellemzők és geometriai megjelenés alapján a következő fáciesek lettek azonosítva: (i) Az öböl területe, ahol a delta síkjához tartozó torkolati zátonytestek fejlődtek ki. (ii) A külső zátony területe, amely elsősorban a nyugodt üledékképződés térszínéhez kapcsolódik szorosabban. (iii) A torkolati zátonytest szegélyterülete, amely a petrofizikai adottságai alapján gyengébb tulajdonságú homokkövekből épül fel. (iv) A torkolati zátonytest jellegzetes vese alakú megjelenésével, amely 6.5-7km szélességben követhető. (v) A partági terület, amely részben követi az elágazó hordalékelosztó medret és ahhoz szorosan kapcsolódik. (vi) A hordalékelosztó medrek, rá

jellemző faágszerűen elágazó megjelenéssel, amely DNy-i irányból fejlődik ki. A mederalakulat 5km hosszan és körülbelül 0.5km szélességben szeli át a torkolati zátonytestet.

Az azonosított fáciesek a Szőreg-1 rezervoár üledékes felhalmozódásának két fő fázisát körvonalazzák: az első fázis két diszkrét kifejlődésű (DNy-i és ÉK-i) torkolati zátonytest megjelenését rögzíti, míg a második a torkolati zátonytesteket felépítő üledéket szállító elosztó medrek megjelenését körvonalazza.

A Száva medencebeli mintaterületen a neurális háló alapú csoportosítás segítségével négy klaszter fácies lett azonosítva. A klaszter fáciesek, mint pontszerű információk szintén az IK alapján lettek laterálisan kiterjesztve a kőzettestből kiválasztott több horizontális felszínében. Egyes horizontokban a C_3 és C_4 jelzésű klaszter összevontan lett kiterjesztve, mint porozitás és homoktartalom szerinti főcsoportok. Ennek oka a kis adatsűrűség egyes szintekben. Az összevonásra az lehetőséget is adott, hogy a varianciaanalízis alapján ez a két csoport relatíve homogén együttest tud képezni a teljes minta tekintetében.

A négy klaszter fáciest a következő kőzettípusokkal lehetett definiálni: (i) A medence aljzati agyagmárgák, alacsony porozitással, amely a nyugodtvízi szedimentációhoz kötődnek. (ii) Az alacsony sűrűségű turbidites áramlatok által felhalmozott homokkövek agyagmárga és aleuritos betelepülésekkel. Ezek az üledékek az elágazó mederalakulatos lobok köztes területeit jellemzik. (iii) Az alacsony sűrűségű turbidites áramlatok által felhalmozott homokkövek aleuritos betelepülésekkel, amelyek a lobszerű üledékekhez köthetők. Ez a klaszter fácies a passzív lobokat tárja fel. (iv) Tiszta homokkövek, amelyek a turbidites medreket és a mederalakulatos lobok centrális részeit definiálják. A medrek elnyújtott ú.n. cipőfűző, illetve faágszerűen elágazó geometriát mutatnak.

Az üledékes környezet a jellemzett fáciesek alapján lett azonosítva. A klasztereket jellemző litológiai típusok eloszlása alapján elmondható, hogy a vizsgált üledékes környezet egy homokban gazdag mélyvízi üledékes rendszer tagjai közé sorolható. Mivel az azonosított fáciesek alapján elágazó, de partgátnélküli mederalakulatos lobokat lehetett azonosítani, így a fáciesek egy homokban gazdag mélyvízi hordalékkúp rendszer középső (suprafan) részét írják le.

A mintaterületen kiválasztott alsó rezervoár ennek a hordalékkúp rendszernek a progradációs fázisait öleli fel. A rendszer térbeli változását a vertikális sorozatban méterről-méterre (az egyes laterális felszínek mentén) nyomon lehetett követni a klaszter fáciesek eloszlása alapján. A hordalékkúp rendszer ÉNy-i irányból progradál DK-felé. Az IK

eredményeként kapott térképek jól mutatták, hogy mindeközben laterálisan eltolódnak a kifejlődött lobok és kialakulnak passzív lobterületek is.

A dolgozatban közölt módszer a fácies elemzéshez és üledékes környezetek jellemzéshez kíván új eljárási lehetőséget bemutatni, amely lehetővé teszi továbbá a különböző objektum alapú modellek készítését, illetve azokhoz tréning képek kialakítását. A neurális háló által lementett súlyok pedig lehetőséget biztosítanak arra, hogy hasonló üledékes környezetek feldolgozásánál a fáciesek azonosítása neurális háló alapú osztályozással történjen.

ACKNOWLEDGEMENTS

I would like to express my thanks to the Hungarian Oil and Gas Company (MOL Group) and the Geology and Reservoir Management of Croatian Oil and Gas Company (INA Plc.) for permission to publish this PhD research. Many thanks also to the Department of Geology and Palaeontology, University of Szeged, Hungary for enabling my PhD research and support.

Many thanks to my mentor János Geiger and my colleagues (especially Kata Náfrádi) for their suggestions and comments.

Special thanks to István Cziczer, Imre Magyar and János Geiger who aroused my interest in a scientific and research career with their valuable and very interesting conversations.

Many thanks to my Croatian friends and colleagues for their special support during my professional practise in Zagreb.

Last but not least, very special thanks to my family and friends, as well.

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

ALTRICHTER, M., HORVÁTH G.; PATAKI, B.; STRAUSZ, GY; TAKÁCS, G.; VALYON, J. (2003): Neural networks (Neurális hálózatok), ed.: Horváth, G., Panem Publisher, p. 447

ANDERBERG, M. (1973): Cluster analysis for applications, Academic Press in New York, p. 359

BÉRCZI, I. (1971): Statistical interpretation of grain size distribution, in: New results of the depositional petrology. (Szemcseeloszlás vizsgálatok statisztikus kiértékelése, in: Az üledékes pertrológia újabb eredményei) Proceeding of the Hungarian Geological Society (Magyar Földtani Társulat), pp. 59-115

BÉRCZI, I. (1988): Preliminary sedimentological investigations of a Neogene depression in the great Hungarian Plain–in: Royden and Horváth (ed): The Pannonian basin-a study in basin evolution, AAPG Memoir Vol. 45, pp. 107-116

BHATT, A. and HELLE, H. B. (2002): Determination of facies from well logs using modular neural networks, Petroleum Geoscience, Vol. 8, pp. 217-228

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

BOCK, H. (1989): Probabilistic models in cluster analysis. Computation Statistics and Data Analysis, Vol. 23, no. 1, pp.5-28

BUSBY-SPERA, C. (1985): A Sand-rich submarine fan in the Lower Mesozoic Mineral King Caldera Complex, Sierra Nevada, California. J. Sedimentological Petr. Vol.55, pp. 376-391

CHANN, M. A., & DOTT, R. J. (1983): Shelf and deep sea sedimentation in Eocene Forearc Basin, Western Oregon - fan or non fan? AAPG Bulletin, Vol. 67, pp. 2100-2116

CHERKASSKY, MULIER, (1998): Learning from data: Concepts, theory, and methods, John Wiley and Sons, Inc., p 441

DELFINER, P.; PEYRET, O.; SERRA, O. (1987): Automatic determination of lithology from well logs by statistical analysis. SPE Formation Evaluation, Vol. 2, pp. 303-310

DEUTSCH, C.V., JOURNEL, A.G. (1992): GSLIB: Geostatistical Software Library and User's Guide. Oxford University Press, Oxford, New York, (2nd ed. 1998), 340

DING, C. and HE, X. (2004): K-means clustering via principal component analysis, Proceedings of the 21st international Conference on Machine Learning, Banff, Canada, (ICML-2004), pp. 225-232

DUDA, R. O.; HART, P. E. (1973): Pattern classification and scene analysis, John Wiley and Sons, Inc., p. 512

DUDA, R. O.; HART, P. E.; STORK, D. G. (2000): Pattern classification, 2nd edition, John Wiley and Sons, Inc., p. 680

EVERITT, B. (1993): Cluster analysis, 3rd edition, Halsted Press, New York, p. 170

FANG, J. H.; CHEN, H. C.; SHULTZ, A. W.; MAHMOUD, W. (1992): Computer-aided well correlation, AAPG Bulletin, Vol. 76, pp. 307-317

FAUSETT, L. (1994): Fundamentals of Neural Networks – Architectures, Algorithms and Applications, Prentice Hall, Englewood Cliffs, NJ., p. 470

FELDHAUSEN, P. H.; ALI, S, A. (1974): A multivariate approach to sedimentary environmental analysis, Gulf Coast Association of Geology Societies Proc., Vol. 24, pp. 314-320

FOLK, R. L. (1954): The distinction between grain size and mineral composition in sedimentary rock nomenclature, Journal of Geology Vol. 62, pp. 344-359

FOLK, R. L.; WARD, W. C. (1957): Brazos River bar: a study in the significance of grain size parameters, Journal of Sedimentary Petrology, Vol. 27, pp. 3-26

FRIEDMAN, G. M. (1961): Distinction between dune, beach and river sand from their textural characteristics, Journal of Sedimentary Petrology, Vol. 31, pp. 514-529

FRIEDMAN, G. M. (1962): On sorting, sorting coefficients and the log-normality of the grain size distribution of sandstones, Journal of Geology Vol. 70, pp. 737-753

GALLOWAY, W.E. (1989): Genetic stratigraphic sequence in Basin Analysis I: Architecture and genesis of flooding-surface bounded depositional Units, AAPG Bullentin, Vol. 73, no. 2, pp. 125-142

GAN, G.; MA, C.; WU, J. (2007): Data Clustering: Theory, Algorithms, and Applications, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pennsylvania, p. 466

GEDEON, T. D.; WOND, P. M.; TAMHANE, D.; LIN, T. (2003): Extending the use of linguistic petrographical descriptions to characterise core porosity. In: NIKRAVESH, M.; ZADEH, L. A.; AMINZADEH, F. (2003): Soft computing and intelligent data analysis in oil exploration, Elsevier (Developments in Pertoleium Science, Vol. 51), p.755

GEIGER, J. (1982): Grain size distribution and depositional environments (Szemcseeloszlás és felhalmozódási környezet), PhD dissertation, University of Szeged, Doctoral school of geosciences, Department of geology and paleontology, p. 104

GEIGER, J. (1986): Textural and morphogenetic study of sedimentary sandstone bodies (Üledékes homokkőtestek szöveti és morfológiai vizsgálata), Földtani Közlöny, Hungarian Geological Society, Vol. 116, No. 3, pp.249-266

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

GEIGER J. (2004): A pannóniai Újfalui („Törteli") Formációban levő Algyő-delta fejlődéstörténete II.: Az Algyő-delta deltasíkjának üledékszerkezeti jegyei (Depositional history of the Pannonian Algyő-delta (Újfalu Formation). -Part two: Sedimentary structures in the Algyő delta) - Földtani Közlöny, Vol. 134/1, pp. 55-73

GEIGER J.-KOMLÓSI J. (1996): Szedimentológiai geomatematikai 3D modellező rendszer törmelékes CH-tárolókban (Sedimentological geomathematical 3D model system in clastic CH-reservoirs), Kőolaj és Földgáz, Vol. 1996/2. pp. 53-81

GILL, D.; SHAMRONI, A.; FLIGELMAN, H. (1993): Numerical zonation of log suites and log facies recognition by multivariate clusrting, AAPG, Bulletin, Vol. 77, pp. 1781-1791

GOOVAERTS P. (1997): Geostatistics for natural resources evaluation. New York: Oxford University Press; 1997. 259–368

GRIMM, E. C. (1987): CONISS: A fortran 77 program for stratigraphically contrained cluster analysis by the method of incremental sum of squares, Computer and Geoscience, Vol. 13, pp. 13-35

HARTIGAN, J. (1975): Clustering Algorithms, John Wiley and Sons, Inc. New York, NY, USA p. 351

HARTIGAN, J.; WONG M. (1979): A K-means clustering algorithm, Applied Statistics, Vol. 28, pp. 100-108

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. (2001): Elements of statistical learning: "Elements of statistical learning: data mining, inference and prediction", Springer-Verlag, New York, p. 539

HAYKIN, S. (1994): Neural Networks: a Comprehensive foundation. Macmillan, The knowledge engineering review, New York, Vol. 13, no. 4, pp. 409 – 412

HAYKIN, S. (1999): Neural Networks: A comprehensive foundation, 2nd edition, Prentice-Hall International, p. 842

HEBB, D. O. (1949): "The Organization of Behaviour": A neuropsychological theory, John Wiley and Sons, New York, p. 378

HELLER, P. L. and DICKINSON, W. R. (1985): Submarine Ramp Facies Model for Delta-Fed Sand-Rich Turbidite Systems, AAPG Bulletin, Vol. 69. pp. 960-976

HERNITZ, Z., DUREKOVIĆ, M., CRNIČKI J. (1996): Production characteristics and reservoir quality at the Ivanic oil Field (Croatia) Predicted by Machine Learning System, Geologia Croatica, Vol. 49, no. 2, pp. 237-242

HILL, T., LEWICKI, P. (2005): Statistics: Methods and Applications: A Comprehensive Reference for Science, Industry, and Data Mining, StatSoft, Inc., 1st ed., p. 800

HOPFIELD, J. J. (1982): "Neurons networks and physical systems with emergent collective computational abilities", Proc. National Academi Science, Vol. 79, no. 8. pp. 2554-2558

JAIN, A.; DUBES, R. (1988): Algorithms for Clustering Data. Englewood Cliffs, NJ: Prentice-Hall, p. 320

JOLLIFFE, T. I. (2002): Principal component analysis, Springer-Verlag, New York, 2nd ed., p. 487

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

KIM, Y.K. and RA, J.B., (1991): 'Weight value initialization for improving training speed in the backpropagation network', International Joint Conference on Neural Networks (IJCNN'91), Singapore, Vol. 3, pp.2396–2401

KLEVERLAAN, K. (1989): Three distinctive feed-lobe systems within one tome slide of the Tortonian Tabernas fan, S.E. Spain. Sedimentology, Vol 36, pp. 25-46

KLOVAN, J. E. (1966): The use of factor analysis in determining depositional environments by size analysis, Mustand Island, Texas, Journal of Sedimentary Petrology, Vol. 28, pp. 211-226

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

KOHONEN, T. (1990): Self-organization map, Proceeding of the IEEE, Vol 78, no.9, pp. 1464-1480

KOHONEN, T. (2001): Self Organized Maps, 3rd edition, Springer-Verlag, Berlin, p. 501

KRUMBEIN, W. C.; ABERDEEN, E. (1937): The sediments of Barataria Bay, Journal of Sedimentary Petrology, Vol. 7, pp. 3-17

LAMPINEN, T., LAURIKKALA, M., KOIVISTO, H., 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, 15-27

LINK, M. H. and WELTON, J. E. (1982): Sedimentology and reservoir potential of Matilija Sandstone an Eocene sand-rich deep sea fan and shallow marine complex, California, AAPG Bulletine, Vol. 66. pp. 1514-1534

LORR M. (1983): Cluster analysis for social scientists, The Jossey-Bass Social and Behavioral Science Series, San Francisco, CA: Jossey Bass, p. 233

MALVIĆ, T., VELIĆ, J., 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, Vol. 58, no. 1, pp. 73-85

MALVIĆ, T., VELIĆ, J. (2011): Neogene Tectonics in Croatian Part of the Pannonian Basin and Reflectance in Hydrocarbon Accumulations - in New Frontiers in Tectonic Research: At the Midst of Plate Convergence, ed. Schattner, U., InTech, Rijeka, pp. 215-238

MCCULLOCH, W. S.; PITTS, W. H. (1943): A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics, Vol. 5 pp. 115-133

MILLER, R.L, KHAN, J.S. (1962): Statistical analysis in the geological sciences. John Wiley and Sons Inc. 164-144

MINSKY, M.; PAPERT, S. (1969): Perceptrons: An introduction to computational geometry MIT Press, Cambridge MA, p. 258

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

MURTAGH, F.; HERNÁNDEZ-PAJARES, M. (1995): The Kohonen self-organizing feature map method: an assessment", Journal of Classification, Vol. 12, pp. 165–190

NORMARK, W. R. (1978): Fan valleys, channels and depositional lobes on modern submarine fans: Characters for the recognition of sandy turbidite environments. AAPG Bulletin, *Vol. 62*, 912-931

NOVAK ZELENIKA, K.; VELIĆ, J.; MALVIĆ, T. (2012): Application of geostatistics in description of turbiditic depositional environments, case study Kloštar Field, Sava Depression - in New horizons in central european geomathematics, geostatistics and geoinformatics, ed. Geiger, J.; Pál-Molnár, E. Malvić, T, GeoLitera Publishing House, Szeged. pp. 61-72

OLEA, R. (1999): Geostatistics for Engineers and Earth Scientists, Boston: Kluwer Academic, 303

PASSEGA, R. (1964): Grain size representation by cM pattern as a geological tool, Journal of Sedimentary Petrology, Vol. 34, pp. 830-847

PASSEGA, R. (1972): Sediment sorting related to basin mobility and environment, AAPG Bulletin, Vol 56, pp. 2440-2450

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

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

READING, H.G., RICHARDS, M. (1994): Turbidite systems in deep-water basin margins classified by grain size and feeder system: AAPG Bulletin, Vol. 78, 792-822

RÉVÉSZ, I. (1980): Az Algyő-2 telep földtani felépítése, üledékföldtani heterogenitása és ősföldrajzi viszonyai [Hydrocarbon deposits Algyő-2: geological structure, sedimentological heterogeneity and paleogeographic features – in Hungarian] - Földtani Közlöny, Vol.110, pp. 512–539

RÉVÉSZ, I. (1982): Az algyői Maros-Szőreg szénhidrogéntelepek üledékföldtani modellje – egy fosszilis delta fejlődéstörténete (Depositional model of Maros-Szőreg CH reservoirs in the Algyő Field - evolution of delta) – Kőolaj és Földgáza, Vol 115, pp. 176-177

RICHARDS, M., and BOWMAN, M. (1998): Submarine fans and related depositional systems II: variability in reservoir architecture and wireline log character. *Marine and Petroleum Geology*, *15*, pp. 821-839

RICHARDS, M., BOWMAN, M., READING, H. (1998): Submarine!fan systems I: characterization and stratigraphic prediction, Marine and Petroleum Geology, Vol15, pp. 689-717

RITTER, H.; MARTINETZ, T.; SCHULTE, K. (1992): Neural competition and Self-Organizing Maps - An introduction, Addison-Wesley, New York, p. 293

ROGERS, S. J.; FAND, J. H.; FARR, C. L.; STANLEY, D. A. (1992): Determination of lithology from well logs using a neural network, AAPG Bulletin, Vol. 76, pp. 731-739

ROJAS, R. (1996): Neural networks: A systematic Introduction, Springer-Verlag Berlin Heidelberg, p. 502

ROSENBLATT, F. (1957): The perceptron: A perceiving and recognizing automaton. - Technical report 85 -460 -1, Project PARA, Cornell Aeronautical Lab.

ROSENBLATT, F. (1958): The perceptron: A probabilistic model for information storage and organization in the brain. – Psychological Review, Vol. 65, pp. 386 - 408.

RUMELHART, D.E.; HINTON, G.E.; WILLIAMS, R.J. (1986): “Learning internal representations by error propagation,” in D. E. Rumelhart and J. L. McCleland, eds.

Parallel Distributed Processing: Explorations in the Microstructure Cognition, 1. MIT Press. pp. 318-362

SAFTIĆ, B., VELIĆ, J., SZTANÓ, Ó., JUHÁSZ, GY., 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, Vol. 56, no. 1, pp. 101-122

SAHU, B. K. (1964): Depositional mechanisms from the grain size analysis of clastic sediments, Journal of Sedimentary Petrology, Vol. 34, pp. 73-83

SEBŐK-SZILÁGYI, SZ., (2011): A Szőreg-1 telep gázsapkát tartalmazó teleprészének szedimentológiai modellezése, doktori értekezés (Sedimentological model of gas cap reservoir unit in Szőreg-1, doctoral dissertation), University of Szeged, Szeged, Hungary, p. 126

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

SELLEY, R., C. (1970): Studies of sequence in sediments using a simple mathematical device. Q. J. geol. Soc. London 125, pp. 557-581

SERRA, O. (1986): Fundamentals of well-log interpretation. Developments in Petroleum Science, 15a,b. Elsevier, Amsterdam, p. 684

SERRA, O. (1985): Sedimentary environments from wireline logs: Houston, Schlumberger, p. 211

ŠIMON, J. (1980): Prilog stratigrafiji u taložnom sustavu pješčanih rezervoara Sava-grupe naslaga mlađeg tercijara u Panonskom bazenu sjeverne Hrvatske. PhD dissertation, University of Zagreb, p. 66

SMITH, R. D. A.(1995): Complex bedding geometries in proximal deposits of the Catelnuova Member, Rochetta Formation, Tertiary Piedmont Basin, NW Italy. In: Pickering, K. T.; Hiscott, R. N.; Kenyon, N. H.; Ricci-Lucchi, F. and Sith R. D. A. (Eds): Atlas of deepwater environments: Architectueal style in turbidite system, London: Chapmnan and Hall. pp. 244-249

SPENCER, D. W. (1963): The interpretation of grain size distribution curves of clastic sediments, Journal of Sedimentary Petrology, Vol. 33, pp. 180-190

TALAŚKA, T. and DŁUGOSZ, R. (2008): Initialization Mechanism in Kohonen Neural Network Implemented in CMOS Technology, European Symposium on Artificial Neural Networks (ESANN), pp. 337-342

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

UDDEN, J. A. (1914): Mechanical composition of clastic sediments, Bulletin of the Geological Society of America, Vol. 25, pp. 655-744

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

VARFIS, A.; VERSINO, C. (1992): Clustering of Socio-Economic Data with Kohonen Maps, IDG VSP, Ispra, Italy, pp. 813-833

VISHER, G. S. (1967): The relation of grain size to sedimentary process, AAPG Bulletin. Vol. 51, p. 484

VISHER, G. S. (1969): Grain size distribution and depositional processes, Journal of Sedimentary Petrology, Vol. 39, pp. 1074-1106

VRBANAC, B. (1996): Paleostrukturne i sedimentološke analize gornjopanonskih naslaga formacije Ivanić Grad u Savskoj depresiji. PhD dissertation, Faculty of Natural Sciences, University of Zagreb, p. 121

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

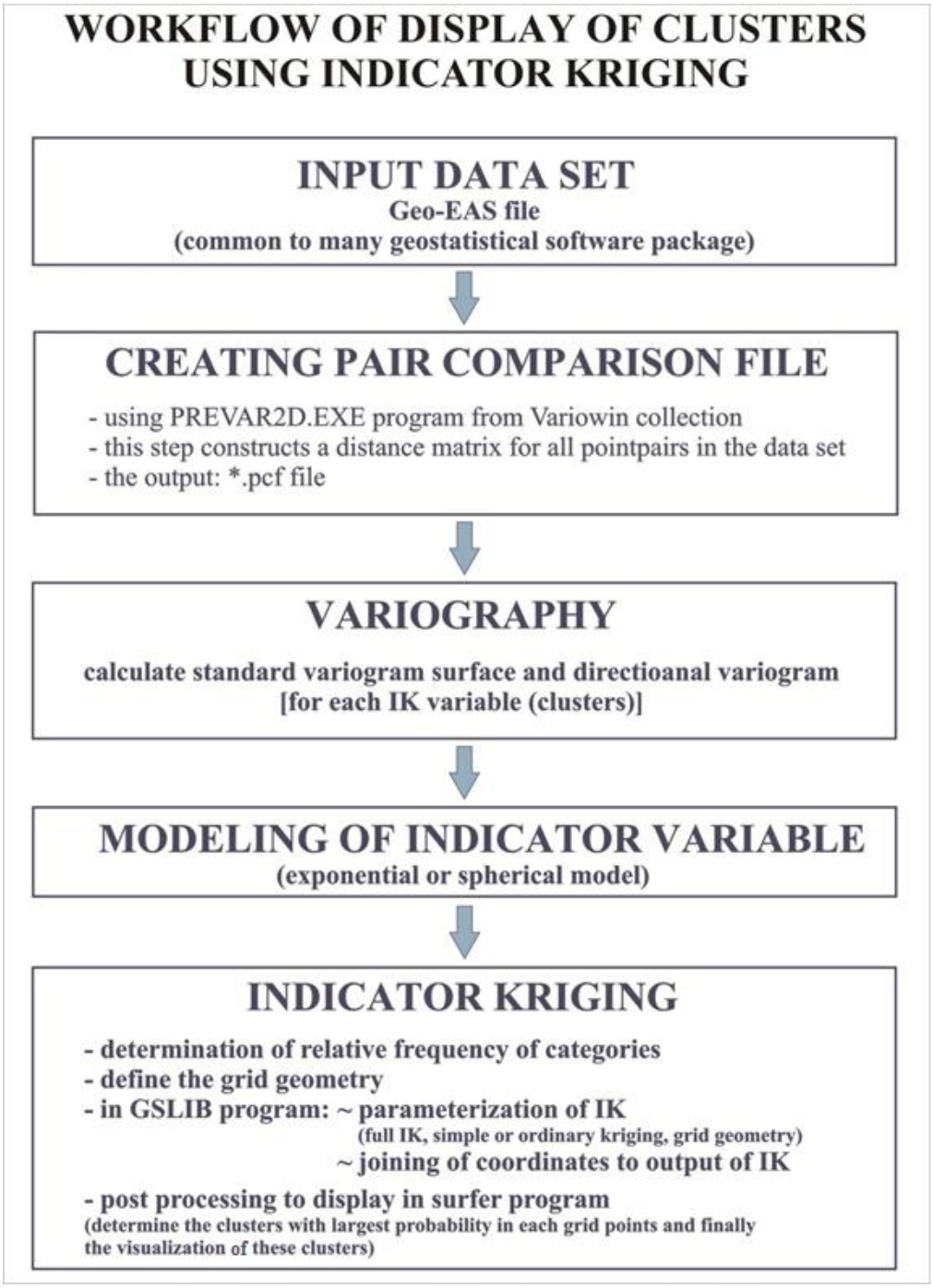
WARREN, S. S. (1994): Neural Networks and Statistical Models, 19th Annual SAS Users Group International conference, April 1994, USA

WIDROW, B. AND HOFF, M. E. (1960): Adaptive Switching Circuits, IRE WESCON… Convention Record, Vol.4, pp. 96-104

WOLFF, M. and PELISSIER-COMBESCURE, J. (1982): FACIOLOG: automatic electro-facies determination, SPWLA Annual Logging Symposium, pp. FF, 6-9

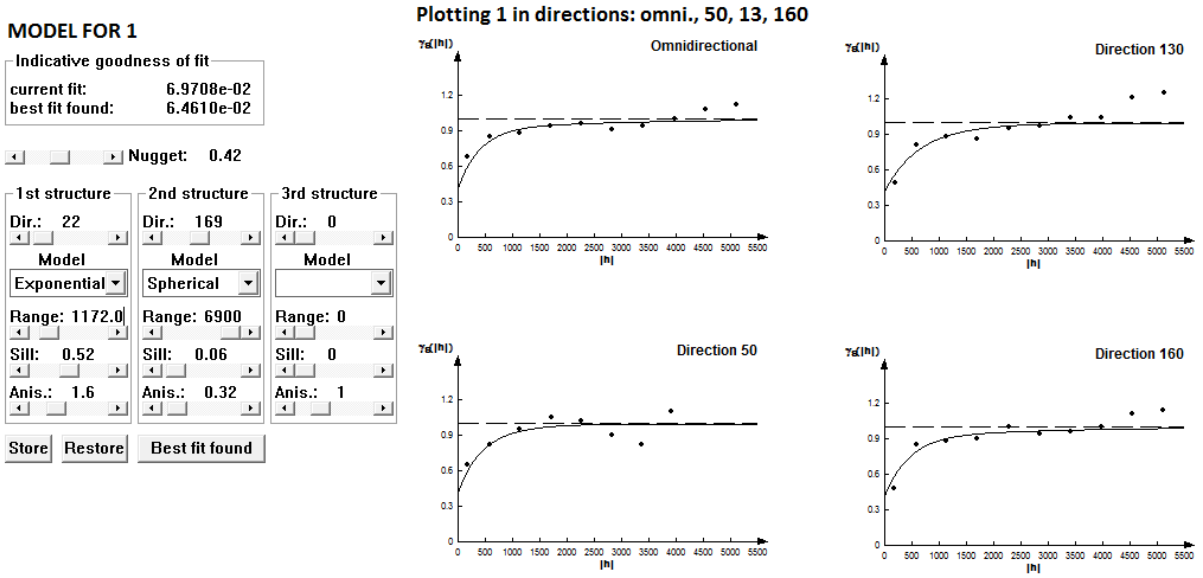
APPENDIX

Appendix of chapter 4.4. (EXTENDING CLUSTERS AS POINT-LIKE RESULTS BASED ON INDICATOR KRIGING)

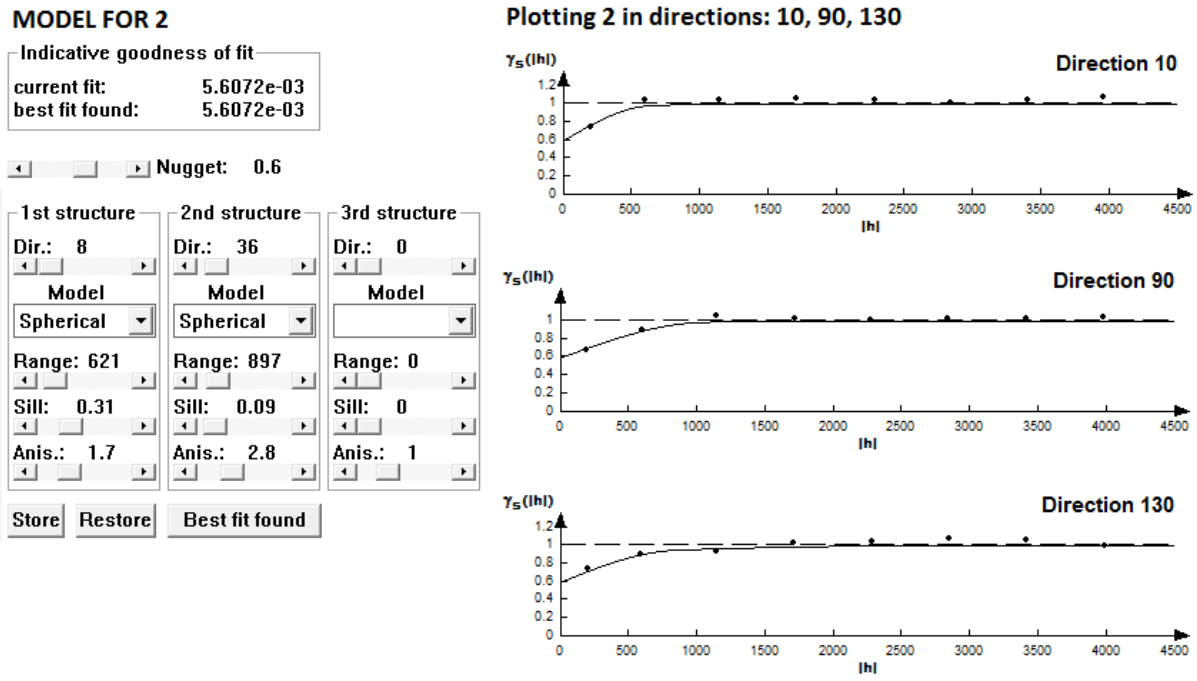


Appendix 1: The workflow for display of clusters which is based on extending the point-like results into the plan using indicator kriging

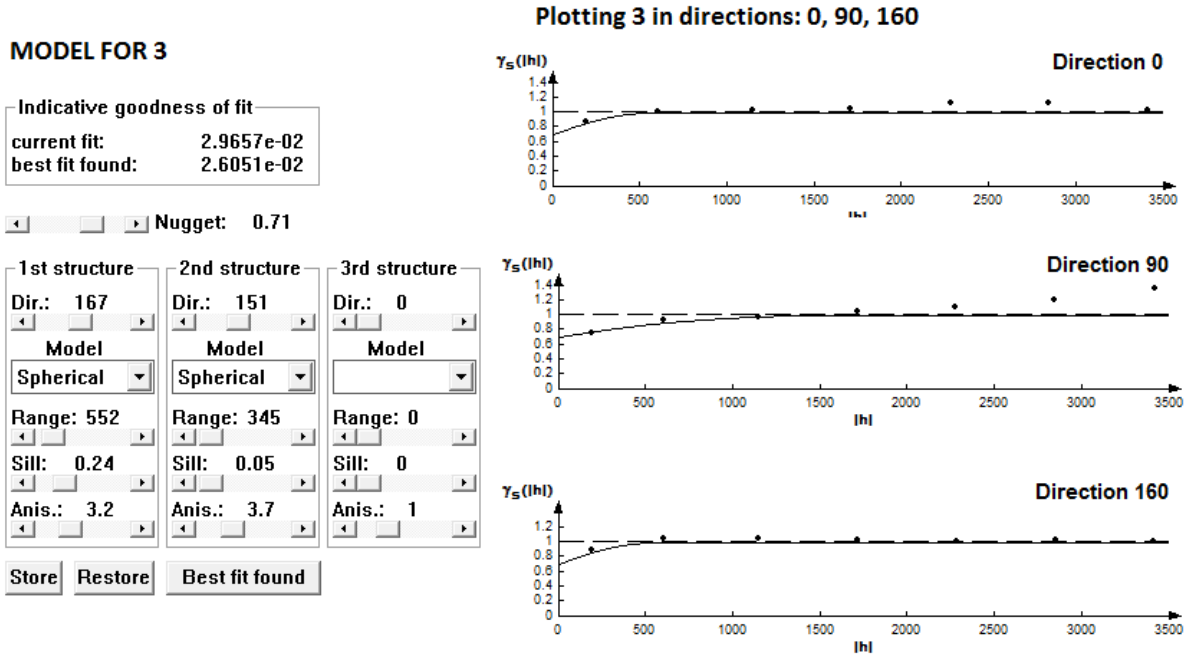
Appendix of chapter 5.1.4.4. (Preparation of input data for IK process)



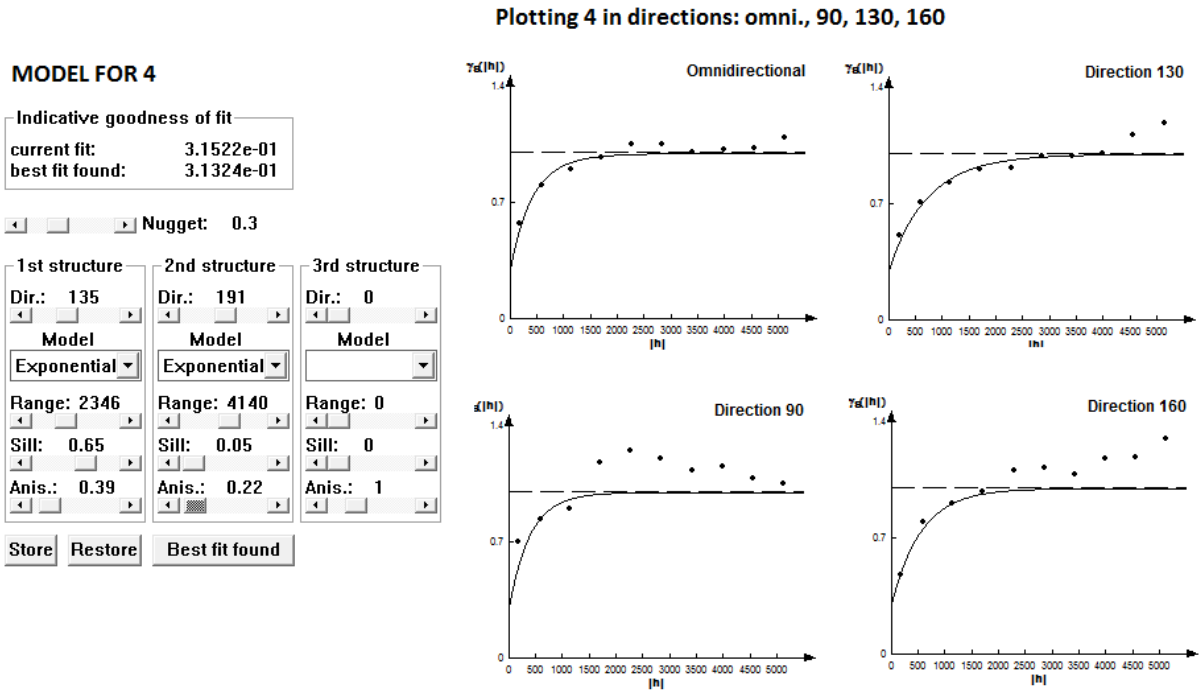
Appendix 2: Model for cluster C_1 in the first surface in Szőreg-1 reservoir



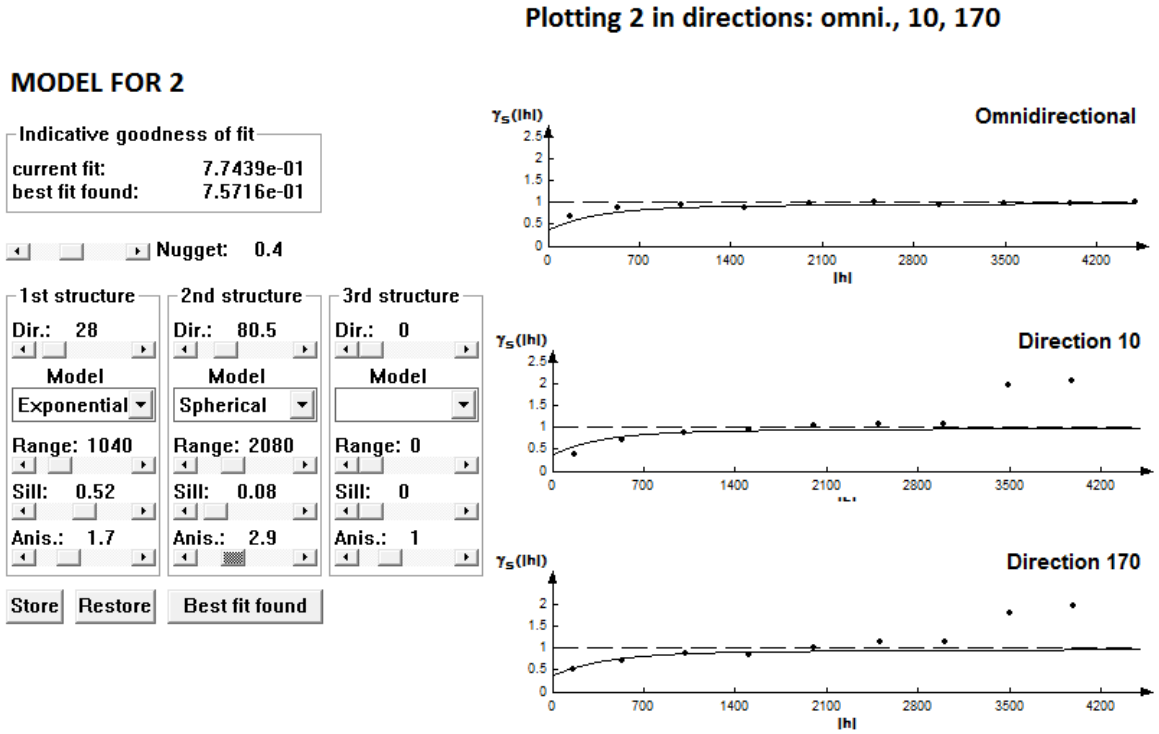
Appendix 3: Model for cluster C_2 in the first surface in Szőreg-1 reservoir



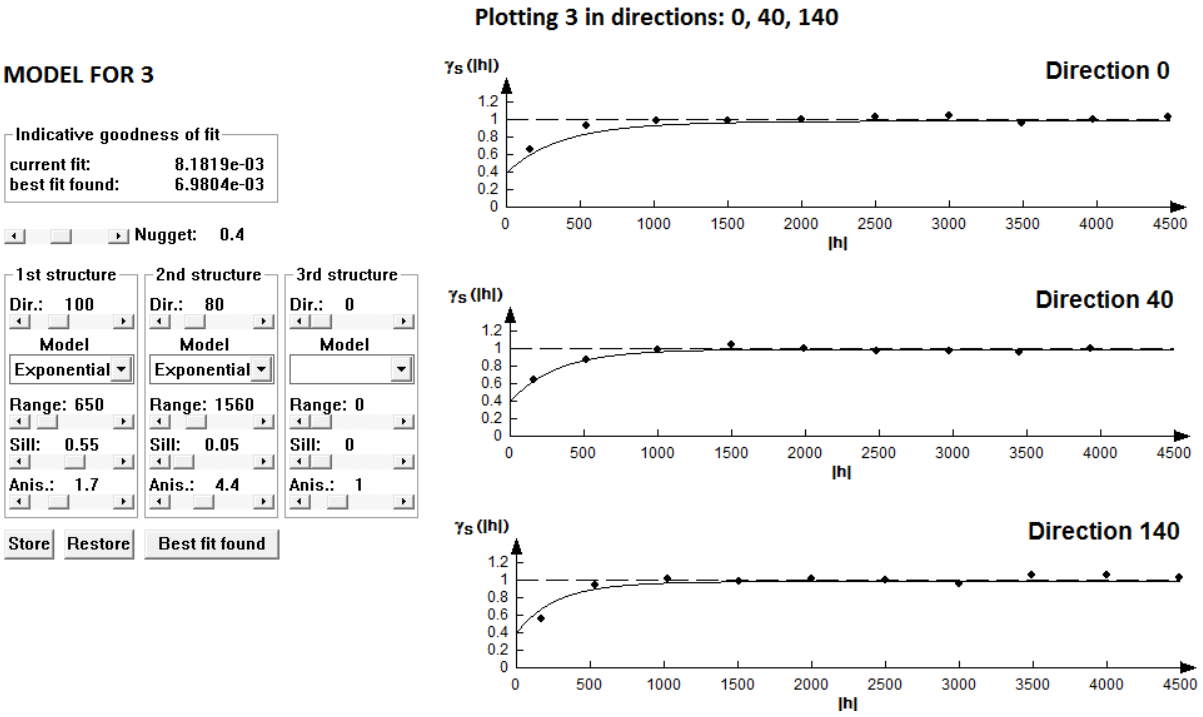
Appendix 4: Model for cluster C_3 in the first surface in Szőreg-1 reservoir



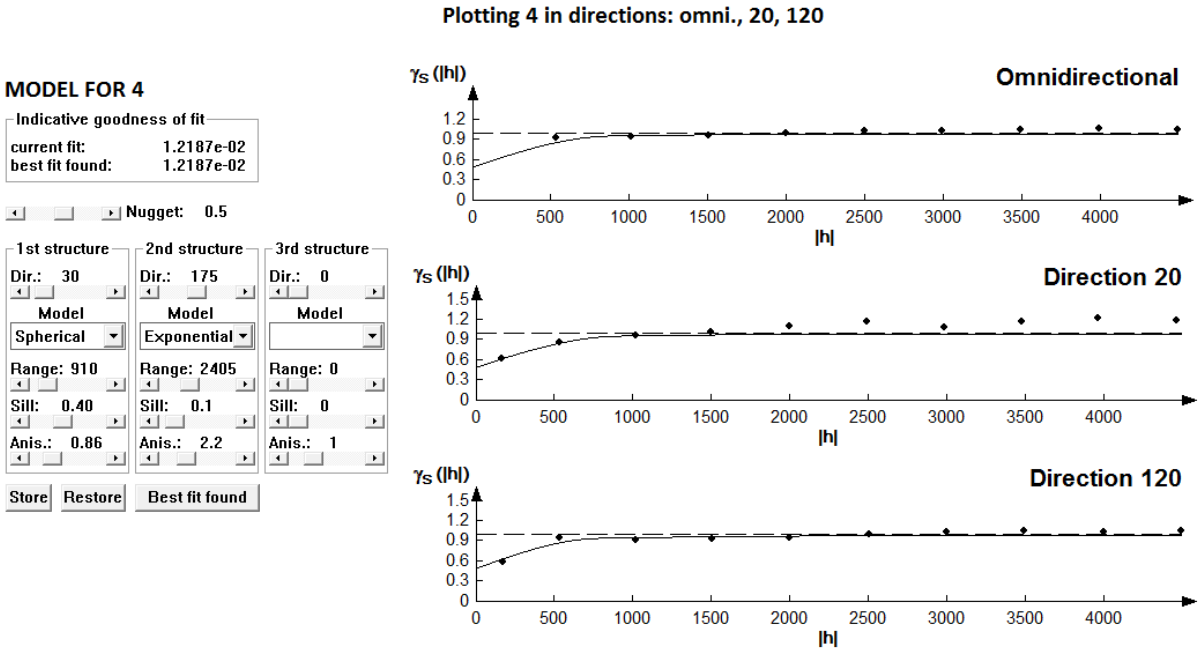
Appendix 5: Model for cluster C_4 in the first surface in Szőreg-1 reservoir



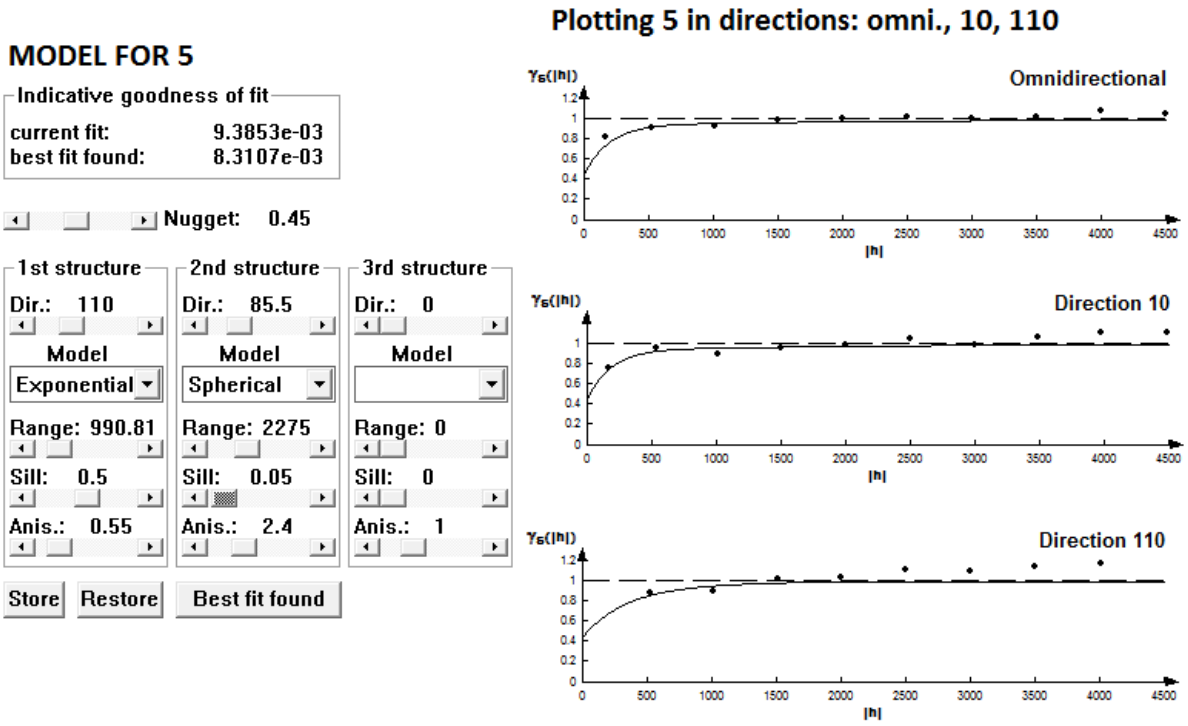
Appendix 6: Model for cluster C_2 in the second surface in Szőreg-1 reservoir



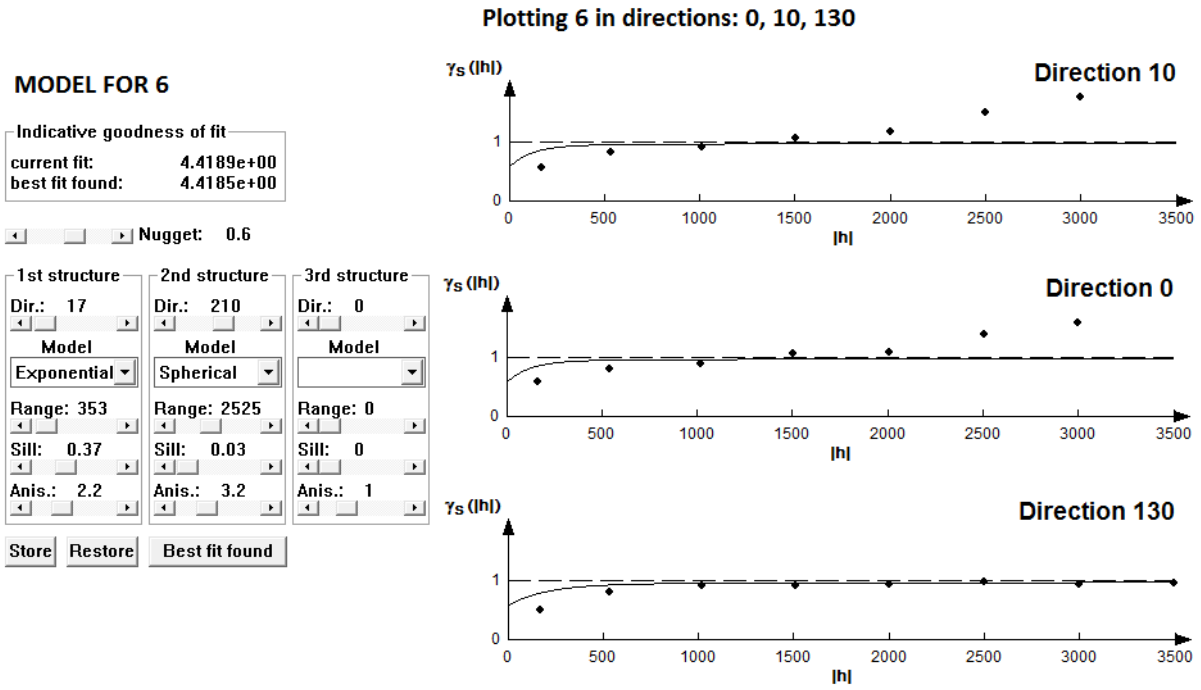
Appendix 7: Model for cluster C_3 in the second surface in Szőreg-1 reservoir



Appendix 8: Model for cluster C_4 in the second surface in Szőreg-1 reservoir

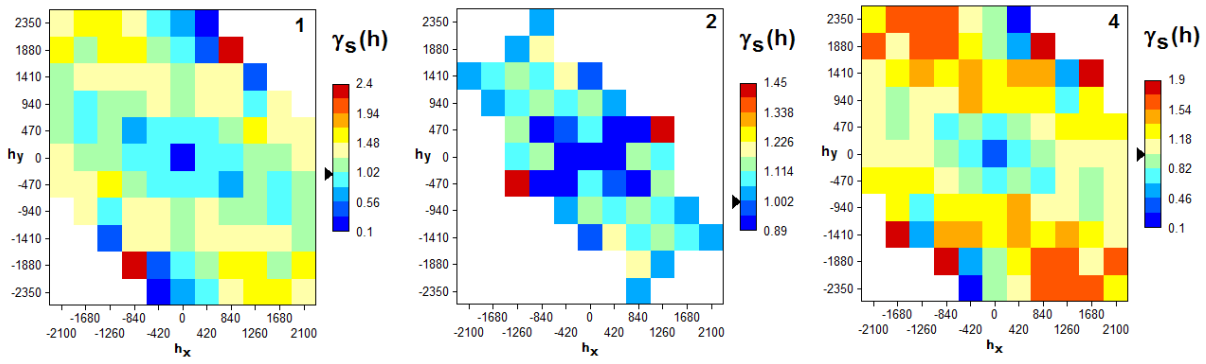


Appendix 9: Model for cluster C_5 in the second surface in Szőreg-1 reservoir

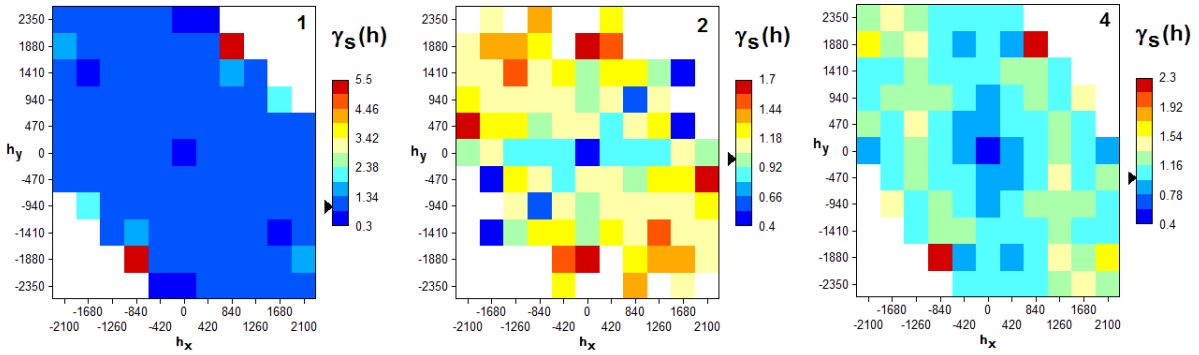


Appendix 10: Model for cluster C_6 in the second surface in Szőreg-1 reservoir

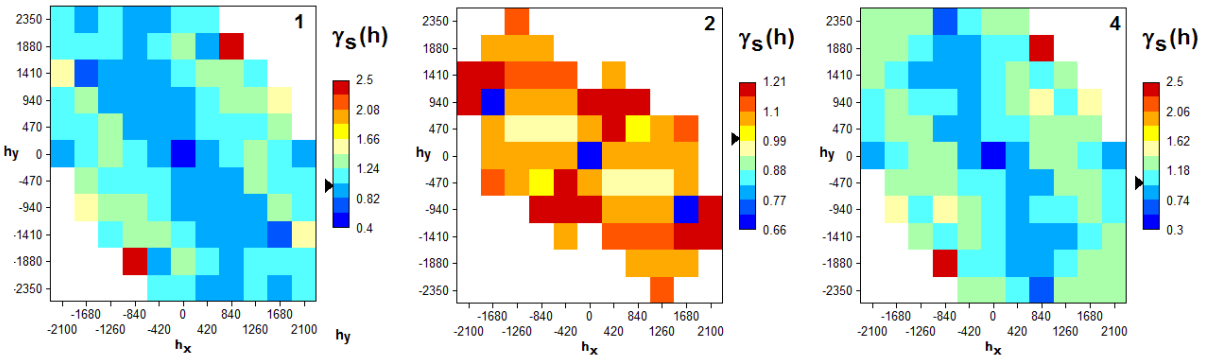
Appendix of chapter 5.2.4.4. (Preparation of input data for IK mapping process)



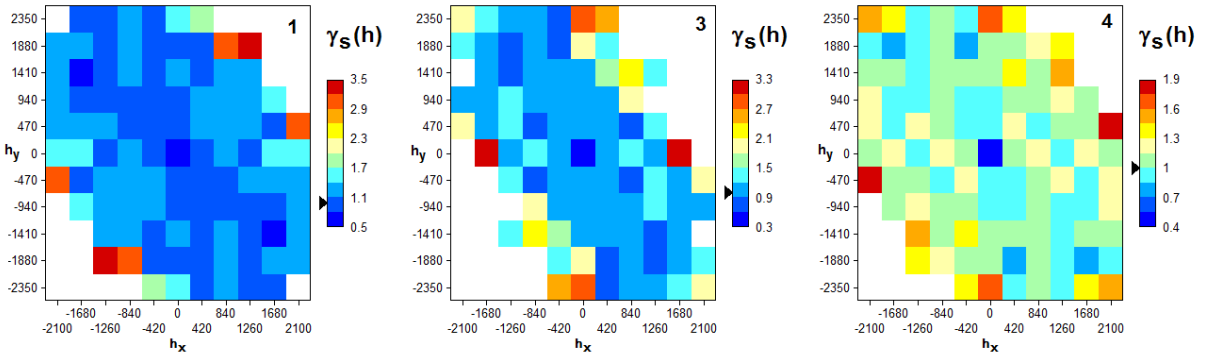
Appendix 11: Variogram surfaces in the lateral surface 13m below the top (in lower reservoir in Sava Filed)



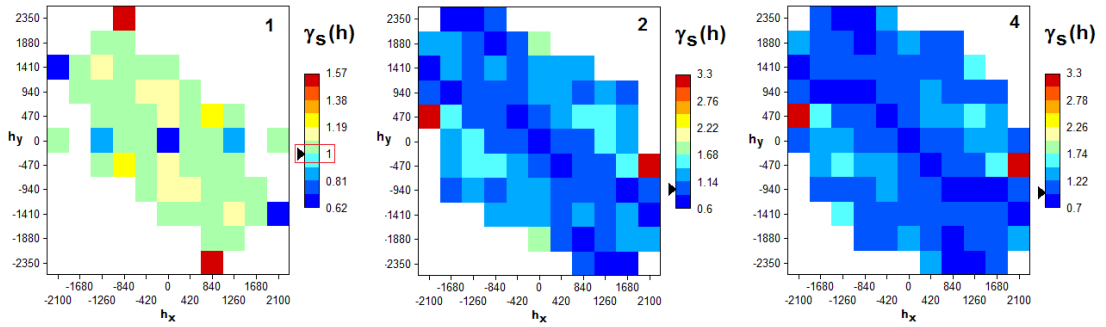
Appendix 12: Variogram surfaces in the lateral surface 11m from the top (in lower reservoir in Sava Filed)



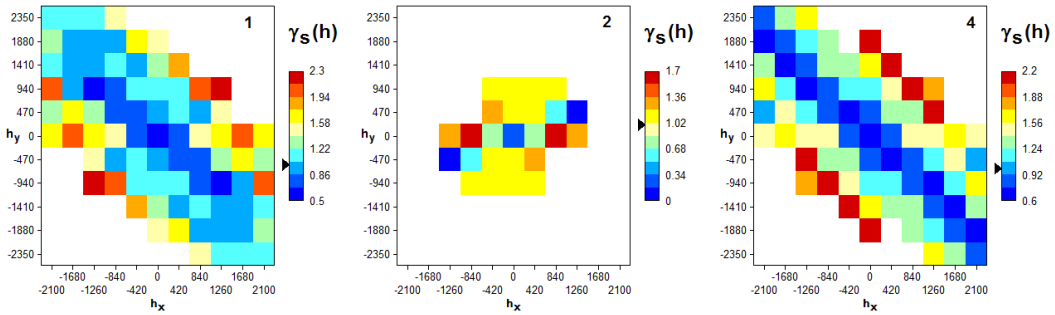
Appendix 13: Variogram surfaces in the lateral surface 10m from the top (in lower reservoir in Sava Filed)



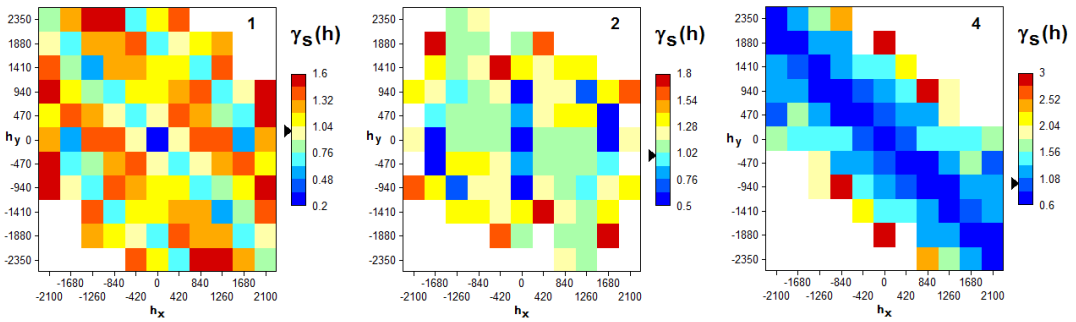
Appendix 14: Variogram surfaces in the lateral surface 7m from the top (in lower reservoir in Sava Filed)



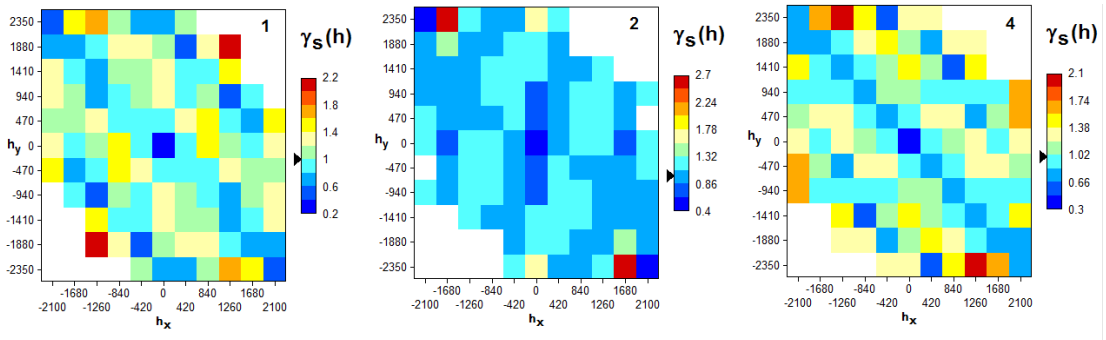
Appendix 15: Variogram surfaces in the lateral surface 4m from the top (in lower reservoir in Sava Filed)



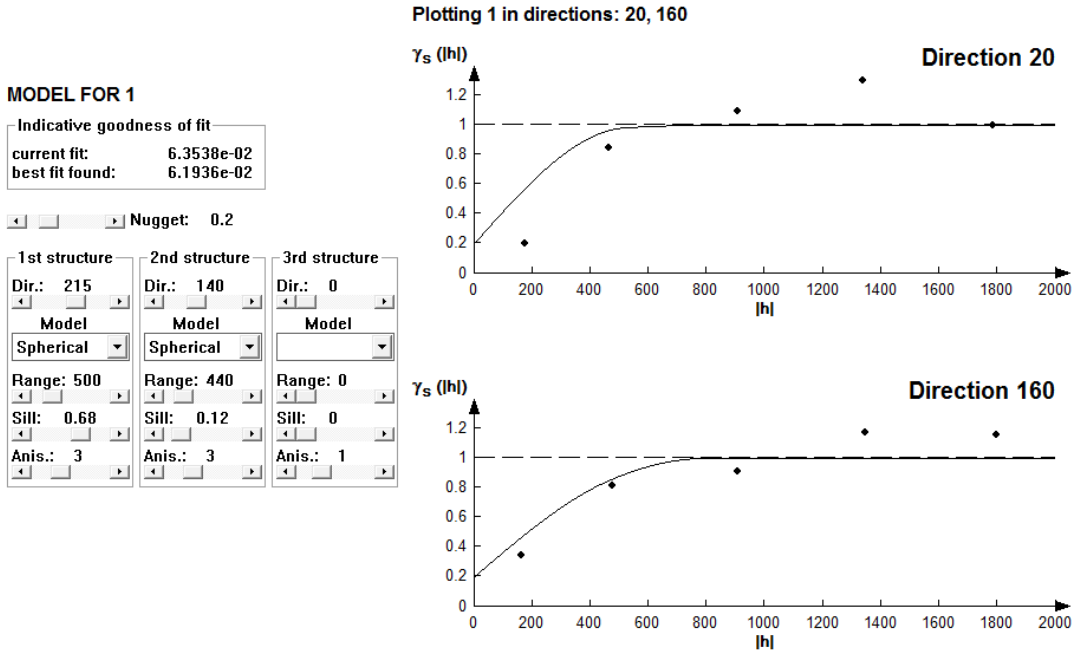
Appendix 16: Variogram surfaces in the lateral surface 9m from the top (in upper reservoir in Sava Filed)



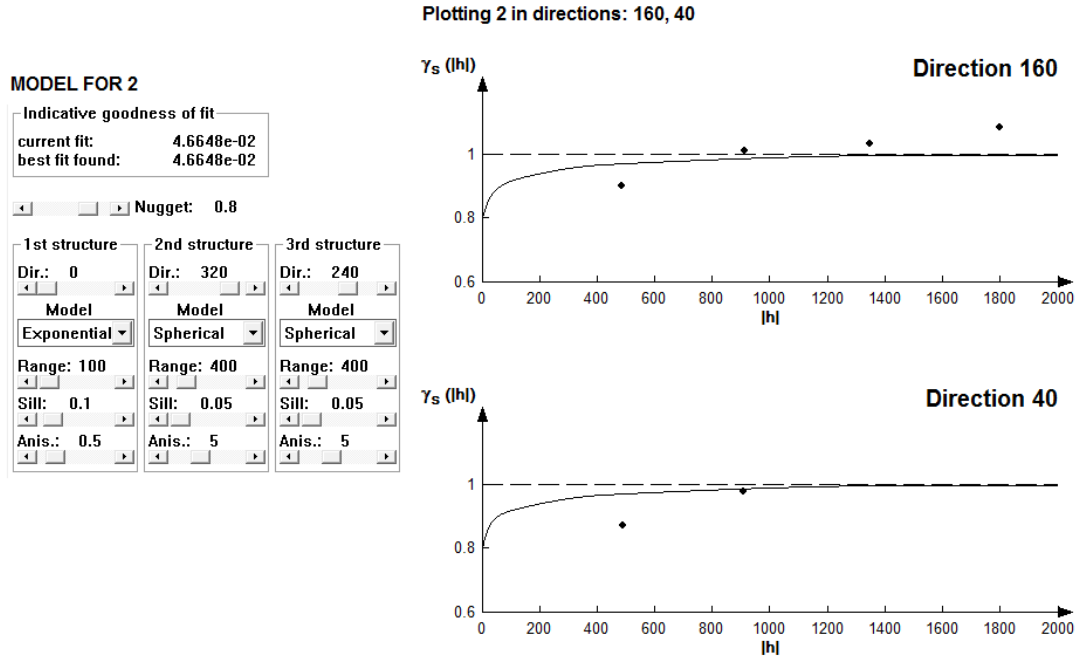
Appendix 17: Variogram surfaces in the lateral surface 6m from the top (in upper reservoir in Sava Filed)



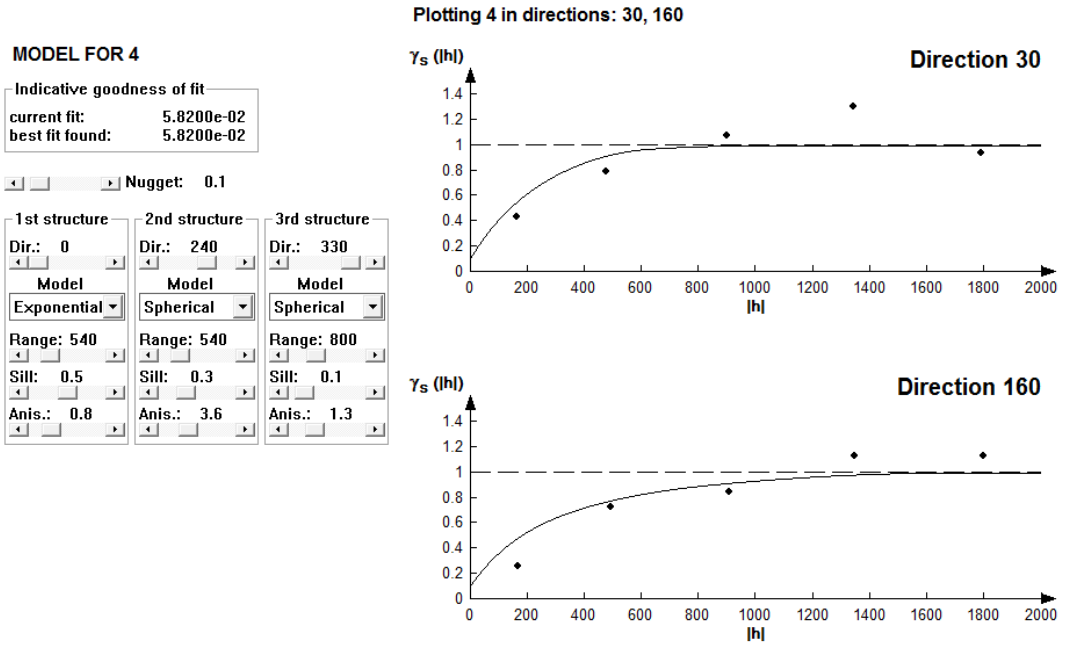
Appendix 18: Variogram surfaces in the lateral surface 3m from the top (in upper reservoir in Sava Filed)



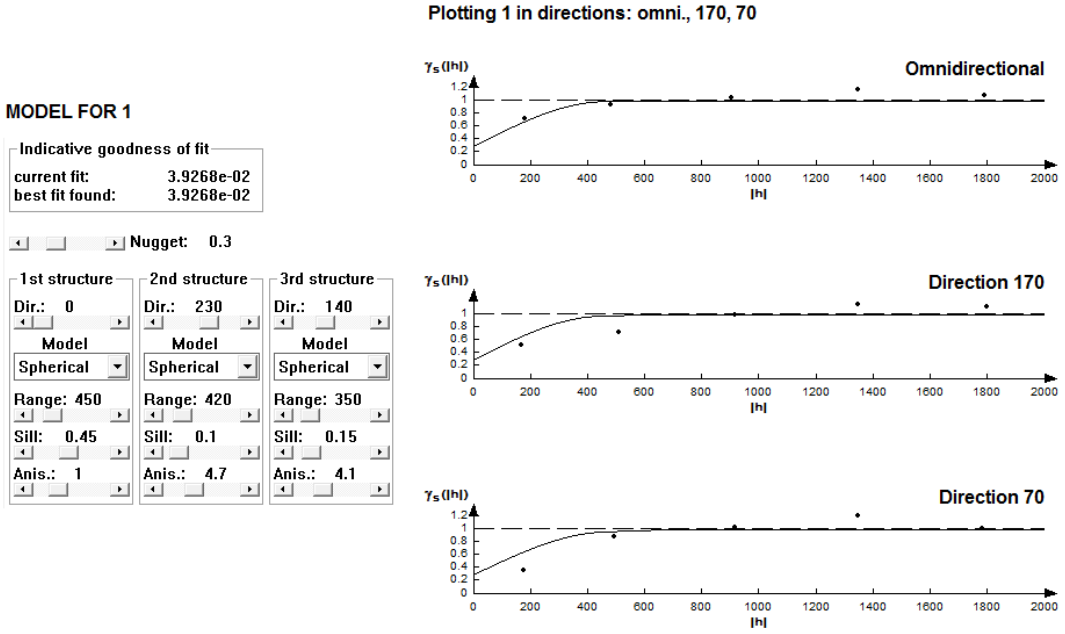
Appendix 19: Model for cluster C_1 in the lateral surface 13m from the top (in lower reservoir of Sava Field)



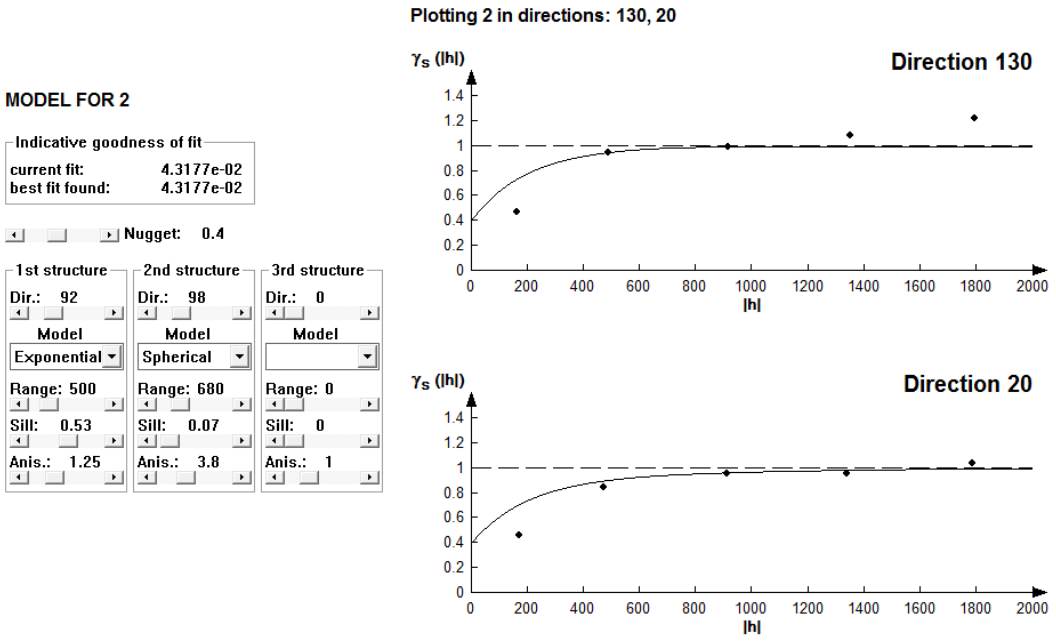
Appendix 20: Model for cluster C_2 in the lateral surface 13m from the top (in lower reservoir of Sava Field)



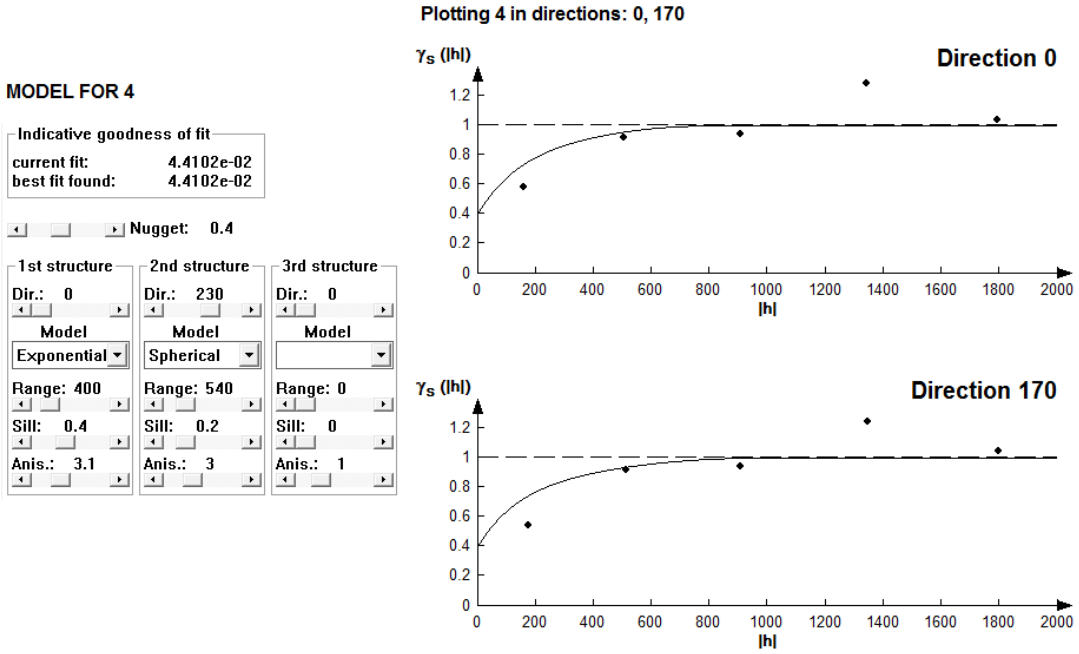
Appendix 21: Model for cluster C_4m (pooled C_3 and C_4) in the lateral surface 13m from the top (in lower reservoir of Sava Field)



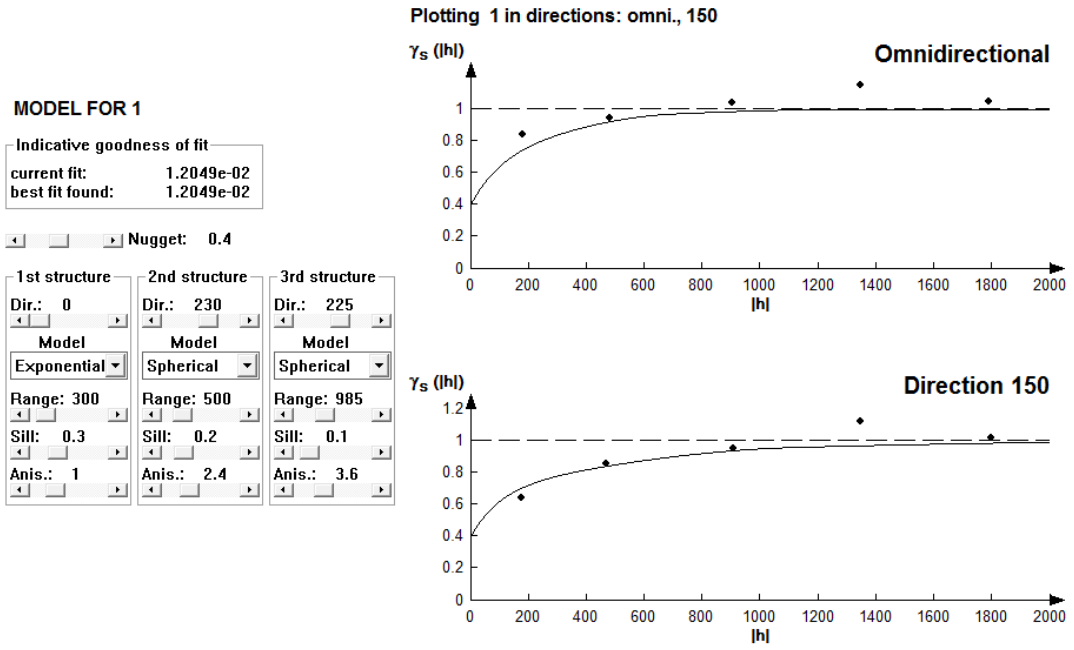
Appendix 22: Model for cluster C_1 in the lateral surface 11m from the top (in lower reservoir of Sava Field)



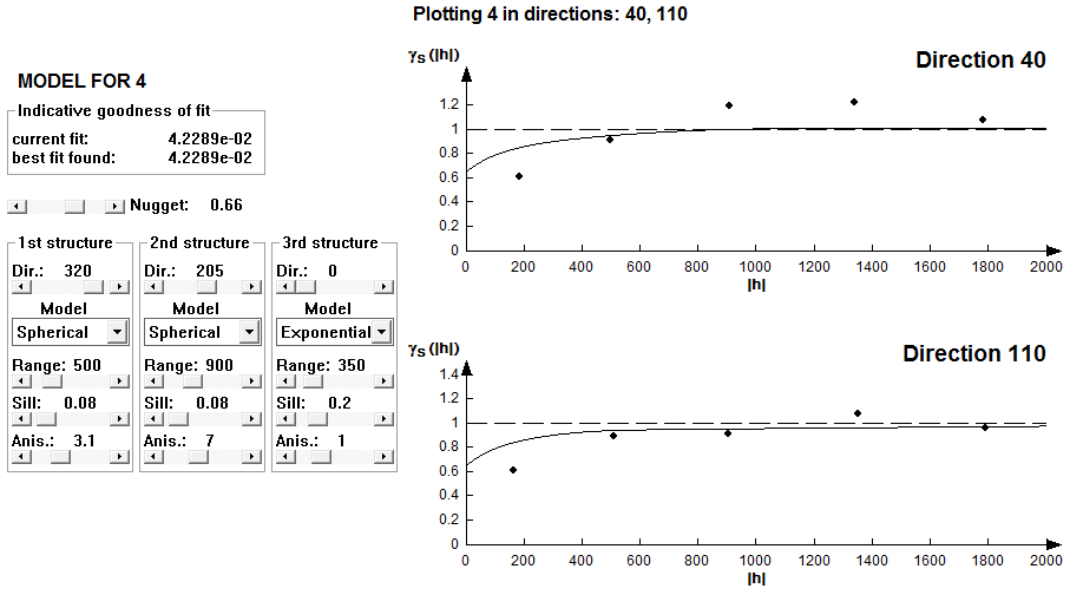
Appendix 23: Model for cluster C_2 in the lateral surface 11m from the top (in lower reservoir of Sava Field)



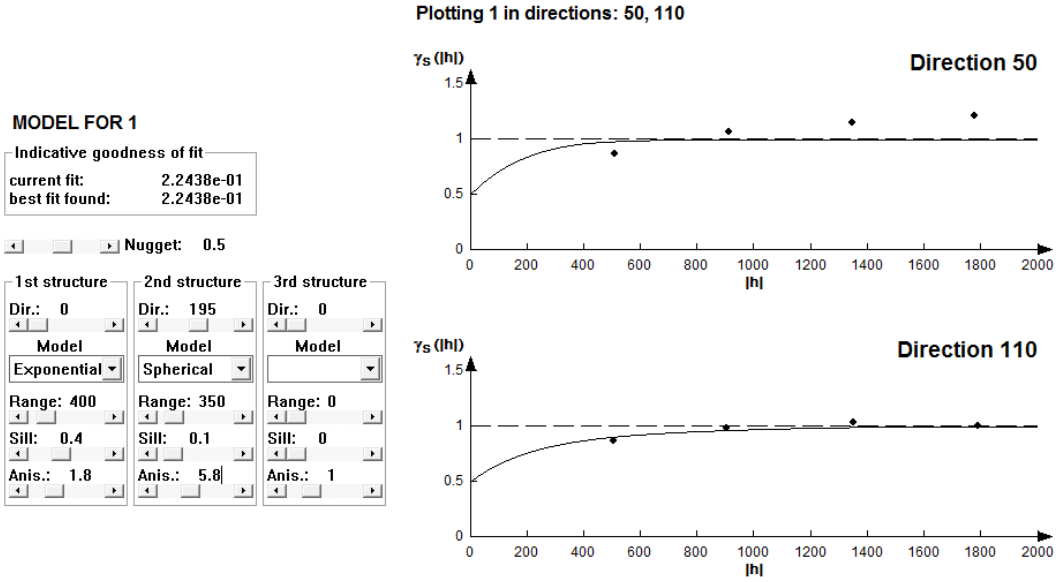
Appendix 24: Model for cluster C_4m (pooled C_3 and C_4) in the lateral surface 11m from the top (in lower reservoir of Sava Field)



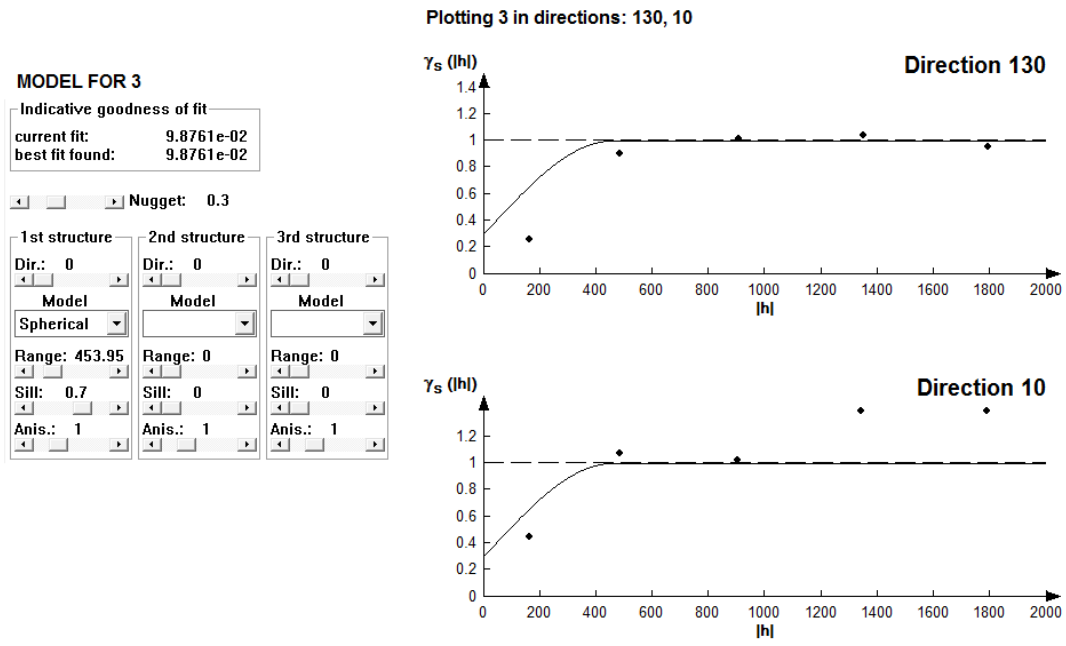
Appendix 25: Model for cluster C_1 in the lateral surface 10m from the top (in lower reservoir of Sava Field)



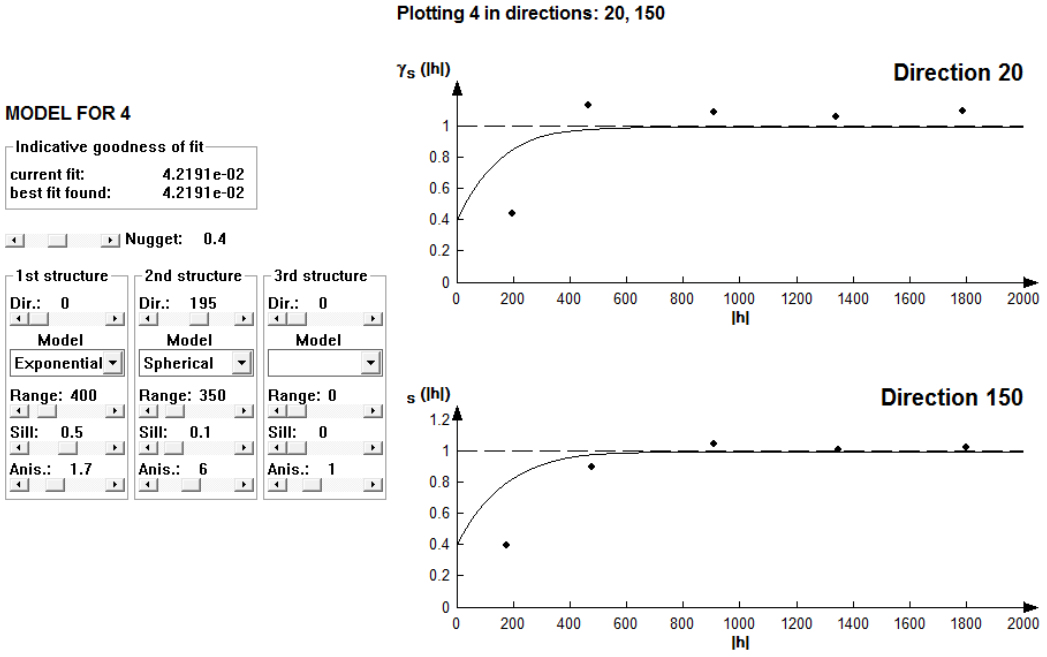
Appendix 26: Model for cluster C_4m (pooled C_3 and C_4) in the lateral surface 10m from the top (in lower reservoir of Sava Field)



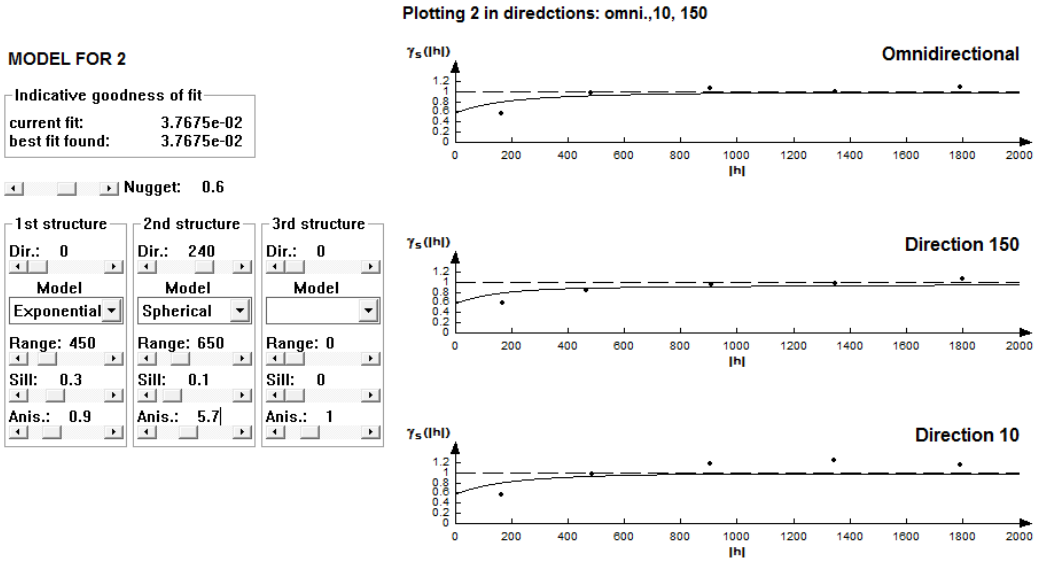
Appendix 27: Model for cluster C_1 in the lateral surface 7m from the top (in lower reservoir of Sava Field)



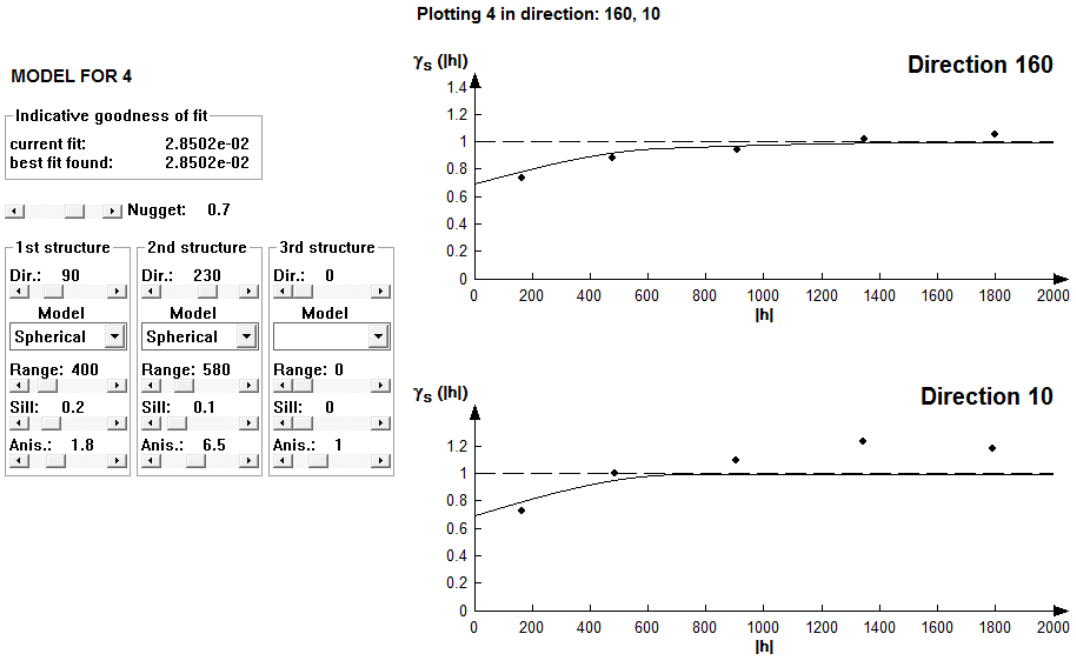
Appendix 28: Model for cluster C_3 in the lateral surface 7m from the top (in lower reservoir of Sava Field)



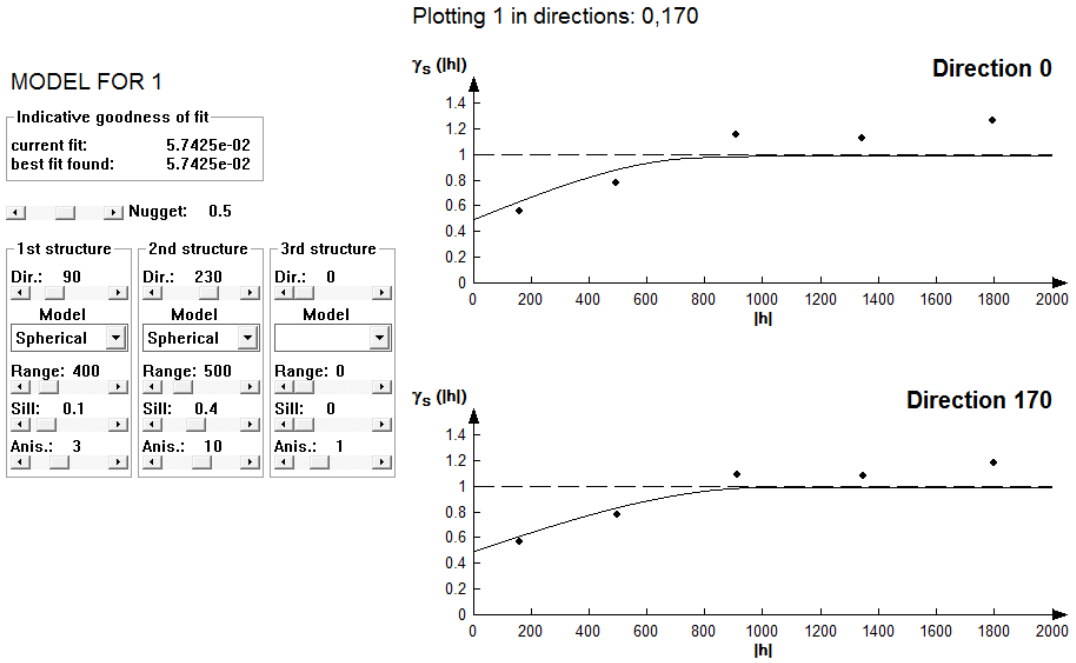
Appendix 29: Model for cluster C_4 in the lateral surface 7m from the top (in lower reservoir of Sava Field)



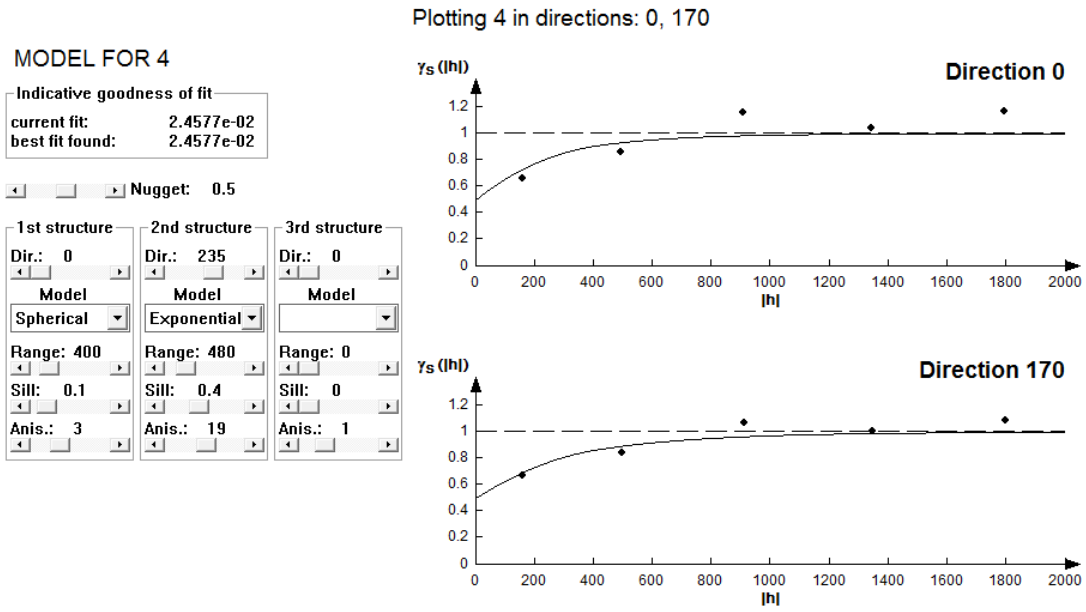
Appendix 30: Model for cluster C_2 in the lateral surface 4m from the top (in lower reservoir of Sava Field)



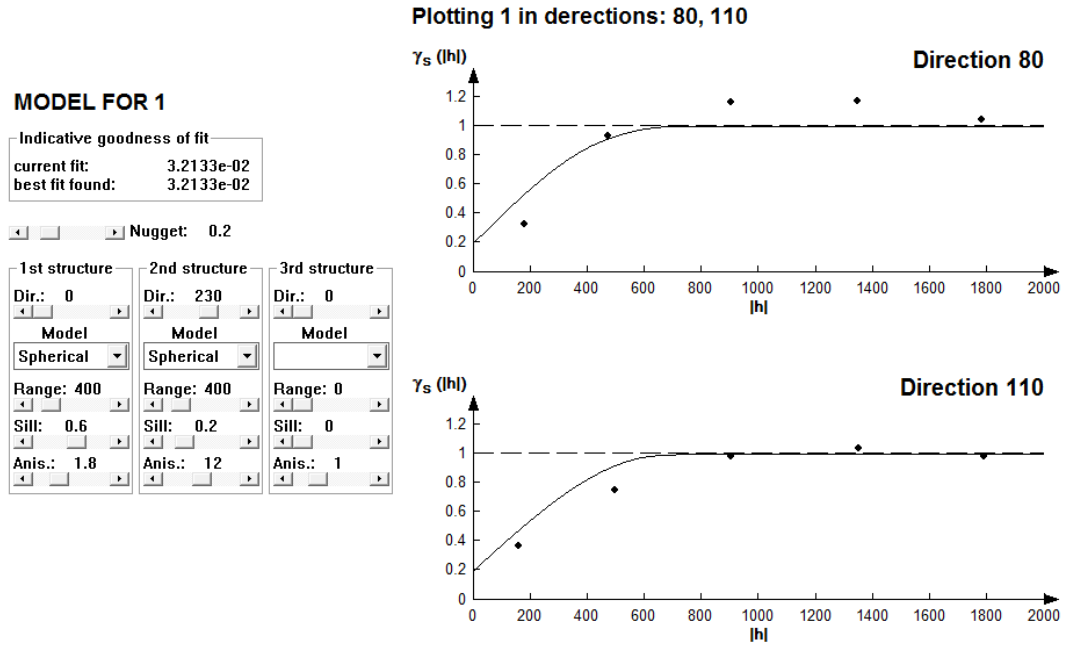
Appendix 31: Model for cluster C_4m (pooled C_3 and C_4) in the lateral surface 4m from the top (in lower reservoir of Sava Field)



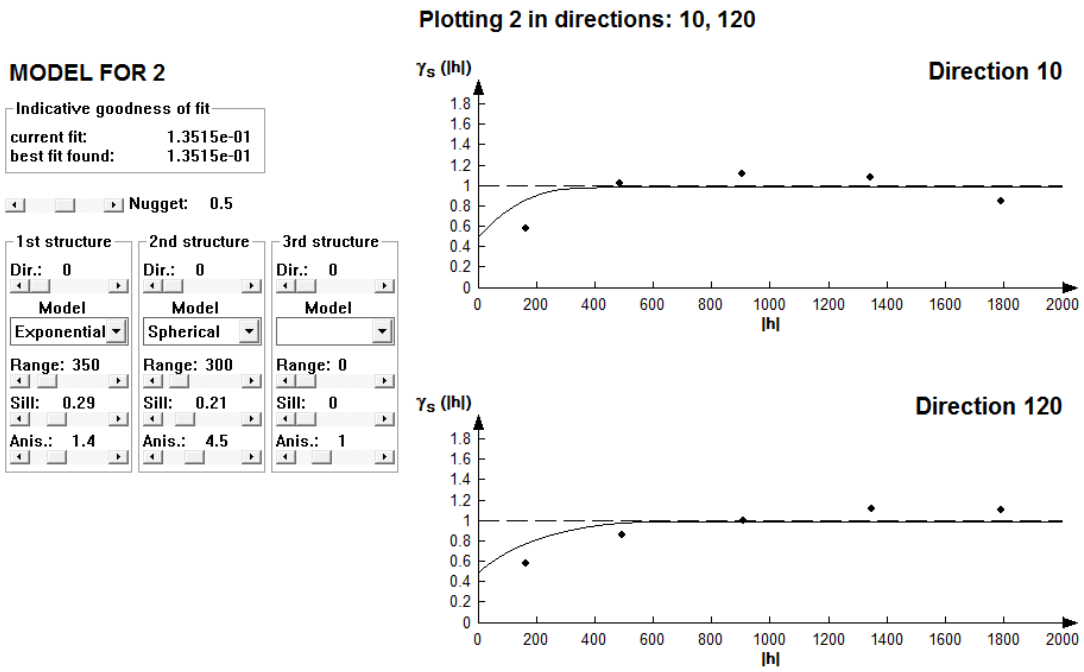
Appendix 32: Model for cluster C_1 in the lateral surface 9m from the top (in upper reservoir of Sava Field)



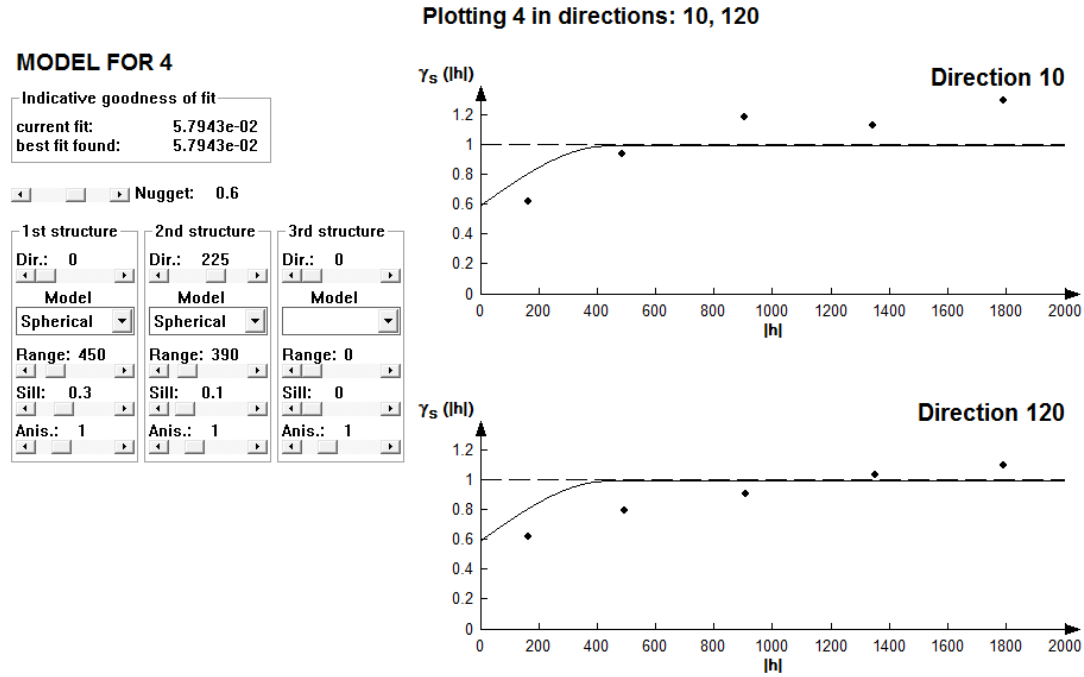
Appendix 33: Model for cluster C_4m (pooled C_3 and C-4) in the lateral surface 9m from the top (in upper reservoir of Sava Field)



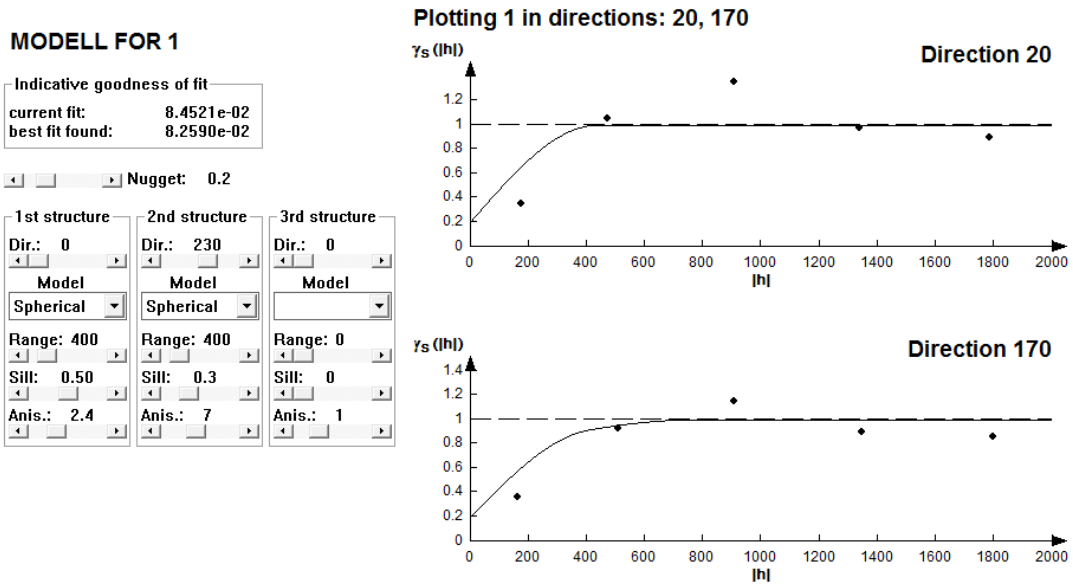
Appendix 34: Model for cluster C_1 in the lateral surface 6m from the top (in upper reservoir of Sava Field)



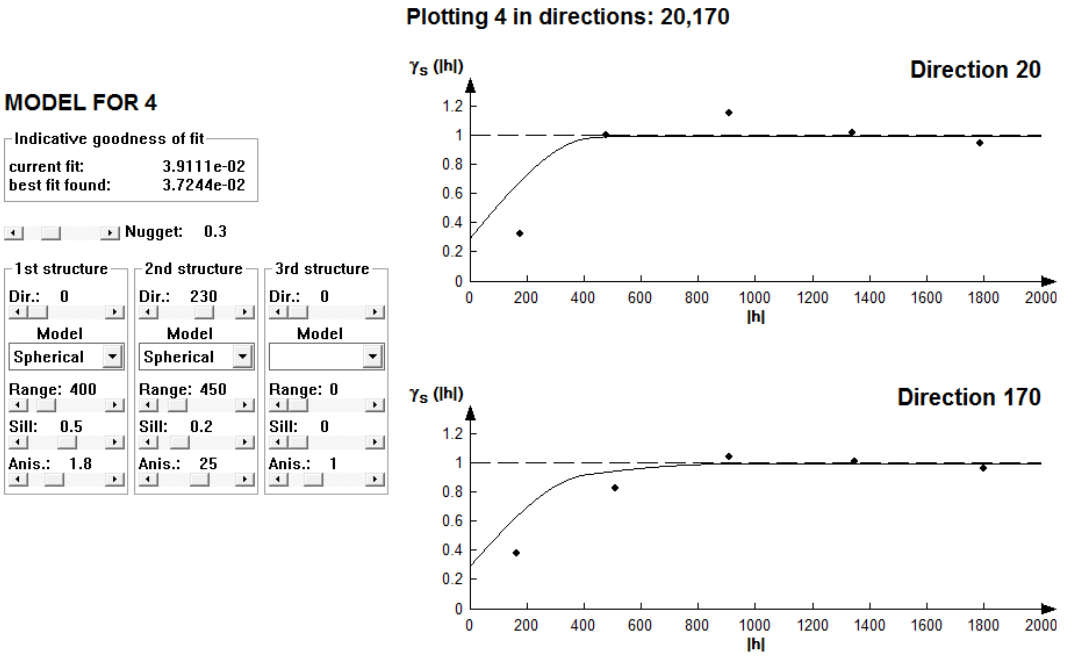
Appendix 35: Model for cluster C_2 in the lateral surface 6m from the top (in upper reservoir of Sava Field)



Appendix 36: Model for cluster C_4m (pooled C_3 and C_4) in the lateral surface 6m from the top (in upper reservoir of Sava Field)



Appendix 37: Model for cluster C_1 in the lateral surface 3m from the top (in upper reservoir of Sava Field)



Appendix 38: Model for cluster C_4m (pooled C_3 and C_4) in the lateral surface 3m from the top (in upper reservoir of Sava Field)

Appendix 39: Table of the model parameter for probability estimation by IK (in lateral surface 13m from the top of lower reservoir in Sava Field)

cluster code	structure	type of variogram	hMax	hMin	cc	ang.	anis.	hVert.	ng.
C_1	**first**	spherical	500	166.65	0.68	235	3	1	0.2
	second	spherical	440	146.65	0.12	310	3	1	
C_2	**first**	exponential	33	16.5	0.1	90	0.5	1	0.8
	second	spherical	400	80	0.05	130	5	1	
	third	spherical	400	80	0.05	210	5	1	
C4_m	**first**	exponential	180	144	0.5	90	0.8	1	0.1
	second	spherical	540	150	0.3	210	3.6	1	
	third	spherical	800	615.4	0.1	120	1.3	1	

Appendix 40: Table of the model parameter for probability estimation by IK (in lateral surface 11m from the top of lower reservoir in Sava Field)

cluster code	structure	type of variogram	hMax	hMin	cc	ang.	anis.	hVert.	ng.
C_1	**first**	spherical	450	450	0.45	90	1	1	0.3
	second	spherical	420	90	0.1	22	4.6	1	
	third	spherical	350	85	0.15	310	4.1	1	
C_2	**first**	spherical	500	400	0.53	358	1.25	1	0.4
	second	spherical	680	179	0.07	352	3.8	1	
C_4m	**first**	exponential	133.33	43	0.4	90	3.1	1	0.4
	second	spherical	540	180	0.2	220	3	1	

Appendix 41: Table of the model parameter for probability estimation by IK (in lateral surface 10m from the top of lower reservoir in Sava Field)

cluster code	structure	type of variogram	hMax	hMin	cc	ang.	anis.	hVert.	ng.
C_1	**first**	exponential	100	100	0.3	90	1	1	0.4
	second	spherical	500	208.3	0.2	220	2.4	1	
	third	spherical	985	273.611	0.1	225	3.6	1	
C_4m	**first**	exponential	500	161.29	0.08	130	3.1	1	0.66
	second	exponential	900	128.6	0.08	245	7	1	
	third	spherical	116.66	116.66	0.18	90	1	1	

Appendix 42: Table of the model parameter for probability estimation by IK (in lateral surface 7m from the top of lower reservoir in Sava Field)

cluster code	structure	type of variogram	hMax	hMin	cc	ang.	anis.	hVert.	ng.
C_1	**first**	exponential	133.33	74.1	0.4	90	1.8	1	0.5
	second	spherical	350	60.34	0.1	255	5.8	1	
C_3	**first**	spherical	453.95	453.95	0.7	90	1	1	0.3
C_4	**first**	exponential	133.33	78.43	0.5	90	1.7	1	0.4
	second	spherical	350	58.33	0.1	255	6	1	

Appendix 43: Table of the model parameter for probability estimation by IK (in lateral surface 4m from the top of lower reservoir in Sava Field)

cluster code	structure	type of variogram	hMax	hMin	cc	ang.	anis.	hVert.	ng.
C_2	**first**	exponential	150	135	0.3	90	0.9	1	0.6
	second	spherical	650	114	0.1	210	5.7	1	
C_4	**first**	spherical	400	222.22	0.2	0	1.8	1	0.7
	second	spherical	580	89.23	0.1	220	6.5	1	

Appendix 44: Table of the model parameter for probability estimation by IK (in lateral surface 9m from the top of upper reservoir in Sava Field)

cluster code	structure	type of variogram	hMax	hMin	cc	ang.	anis.	hVert.	ng.
C_1	**first**	spherical	400	133.33	0.1	0	3	1	0.5
	second	spherical	500	50	0.4	220	10	1	
C_4m	**first**	spherical	400	133.3	0.1	90	3	1	0.5
	second	exponential	160	8.5	0.4	215	19	1	

Appendix 45: Table of the model parameter for probability estimation by IK (in lateral surface 6m from the top of upper reservoir in Sava Field)

cluster code	structure	type of variogram	hMax	hMin	cc	ang.	anis.	hVert.	ng.
C_1	**first**	spherical	400	222.22	0.6	90	1.8	1	0.2
	second	spherical	400	33.33	0.2	220	12	1	
C_2	**first**	exponential	116.66	83.328	0.29	90	1.4	1	0.6
	second	spherical	300	66.66	0.21	90	4.5	1	
C_4m	**first**	spherical	450	321.43	0.3	90	1.4	1	0.5
	second	spherical	390	22.94	0.1	225	17	1	

Appendix 46: Table of the model parameter for probability estimation by IK (in lateral surface 3m from the top of upper reservoir in Sava Field)

cluster code	structure	type of variogram	hMax	hMin	cc	ang.	anis.	hVert.	ng.
C_1	**first**	spherical	400	0.66	0.5	90	2.4	1	0.2
	second	spherical	400	57.14	0.3	220	7	1	
C_2	**first**	exponential	93	23.93	0.35	90	3.9	1	0.4
	second	spherical	300	71.43	0.25	110	4.2	1	
C_4m	**first**	spherical	400	222.22	90.5	90	1.8	1	0.3
	second	spherical	450	18	0.2	220	25	1	

8