Rudarsko-geološko-naftni zbornik	Vol. 29	str. 9-16	Zagreb, 2014.

UDK 550.8
UDC 550.8

Prethodno priopćenje
Preliminary communication

Language/Jezik:Engleski/English

ZONAL ESTIMATION AND INTERPOLATION AS SIMULTANEOUS APPROACHES IN THE CASE OF SMALL INPUT DATA SET (ŠANDROVAC FIELD, NORTHERN CROATIA)

ZONALNA PROCJENA I INTERPOLACIJA KAO ISTOVREMENI PRISTUPI U SLUČAJU MALOG ULAZNOG BROJA PODATAKA (PRIMJER POLJA ŠANDROVAC, SJEVERNA HRVATSKA)

IVANA MESIĆ KIŠ[1,2,] TOMISLAV MALVIĆ[3,4]

[1]Elementary School “Kralj Tomislav”, Matice hrvatske 1, 31500 Našice
[2]University of Zagreb, Faculty of Science, Department of Geology, Horvatovac 95, 10000 Zagreb, doctoral student
[3]University of Zagreb, Faculty of Mining, Geology and Petroleum Engineering, Department of Geology and Geological Engineering, Pierottijeva 6, 10000 Zagreb, HR
[4]INA Plc., Šubićeva 29, 10000 Zagreb, HR

AUTHOR’S AFFILIATIONS:

Ivana MESIĆ KIŠ, PhD student in Geology, University of Zagreb, Faculty of Science
Tomislav MALVIĆ, Associate Professor in Geology, University of Zagreb, Faculty of Mining, Geology and Petroleum Engineering; senior expert, INA Plc., Zagreb; scientific advisor

Ključne riječi: točkasti podatci, Thiessenovi poligoni, deklasterizacija, kriging, Bjelovarska subdepresija

Sažetak

Analiziran je prostor Bjelovarske subdepresije, osobito polje Šandrovac koje se nalazi u sjevernom dijelu. U ovom primjeru kao ulazni podatci uporabljene su dubine EK-markera Z', tj. granice panona i ponta. One su statistički analizirane na razini cijele subdepresije iz 497 podataka očitanih iz pravilne mreže s ćelijama veličine 1x1 km kojom je prekrivena postojeća paleostrukturna karta. Nadalje, odabrano je 18 bušotinskih smjestišta unutar polja Šandrovac gdje su karotažom određene dubine markera (primjer malog ulaznog broja podataka). I oni su očitani izravno sa spomenute karte te kartirani jednom od deklasterizacijskih metoda, tj. metodom Thiessenovih poligona ili kriginga. Zaključeno je kada kartiranje uključuje značaje lokalne nesigurnosti te mali broj podataka, opravdano je dubinsko kartiranje na oba prikazana načina te usporedba rješenja.

Key words: point data, Thiessen polygons, declustering, Kriging, Bjelovar Subdepression

Abstract

The Bjelovar Subdepression area in Northern Croatia was analysed, especially the Šandrovac Field that is located in the northern part of the subdepression. In this example, e-log depth marker Z’, i.e. the Pannonian and Pontian boundary, was used as an input data. The data were statistically analysed for the entire subdepression from 497 data readings from the regular grid with cell size of 1x1 km that covers the existing palaeostructural map. Then is selected 18 well data within the Šandrovac Field where e-log markers are recognised (an example of a small number of data). They are also read directly for given structural map and mapped using one of the declustering methods known as Thiessen polygon method or Kriging. It is concluded when the mapping includes small number of data, and consequently local uncertainties, the subsurface mapping need to be done on both ways and maps compared.

1. Introduction

Data in earth science applications are rarely collected uniformly over the site of interest since many factors can restrict the access to particular geographical locations, and the available data may be production locations that are not chosen with the goal of informing spatial statistics. The acceptable set of data for analysis is the one with statistics of input values which is representative and can be approximated preferably by a normal (Gaussian) distribution. However, most of natural process and data set does not follow normal distribution. Many of the statistical methods and stochastical models that are usually applied require the assumption that a variable or variables are normally distributed. When a variable is not normally distributed, transformed variable is created and tested for normality. Because of a limited number of data, the aim is to achieve such a set that is marked by second order stationarity, a set with known median, but also mean and variance. It can be considered representative for the observed spatial variable, which is also the statistical „population“. When assessing new values of sets with systematic errors and clustering, they can be corrected by assigning the weighting to each measured value (e.g., **Dubois and Saisana, 2002**), and their calculation is fundamental for every interpolation of algorithm or its estimation.

Kriging is a method of interpolation that gives good estimation if the data locations are fairly dense (i.e. more than 20 data) and relatively uniformly distributed throughout the study area. It helps to compensate for the effects of data clustering (although Kriging is one of the most successful methods for interpolation of clustered data), assigning individual points within a cluster less weight than isolated data points (e.g., **Malvić, 2008**).

The ”conventional” statistics often do not provide reliable solutions to the problem of constructing representative spatial distributions. In stationary sequences (joint probability distribution is invariant over time), extreme values can occur in clusters. The first step in making inferences is identifying clusters in the data, a procedure known as declustering. Based on the configuration of data, declustering assigns weights to the available data set. Clustered data are given less weight and data in sparsely sampled areas are given more weight. The data values are unaltered; they are simply given more or less influence based on their spacing (**Oy Leuangthong et al., 2008**). However, in all Kriging techniques except the Simple Kriging, the sum of all weights is standardised to 1.

Geometrical declustering methods generally used in geostatistics are: the cell-declustering method and the polygonal method. When estimating the exhaustive (whole sample space) mean, a weighted linear combination of the available sample values is used.

The cell-declustering method was first proposed by **Journel (1983)** and a modified version was later introduced by **Deutsch (1989)**. The method uses a moving window system that splits the study area into rectangular cells, and each measurement gets a weight that is inversely proportional to the number of points that are in the same window (**Dubois and Saisana, 2002**).

In the polygonal method, polygon of influence (known as Thiessen or Voronoi polygon, e.g., **Chow, 1964**; **Boots, 1987**) is constructed in such a way that its geometry will include all data points that are closer to the sample compared to any other measurements. As a result, the estimated global mean, F (x), of a data set is defined by:

$$F(x) = \frac{\sum_{i=1}^{n} w_i \times x_i}{\sum_{i=1}^{n} w_i} \qquad \textbf{Eq. 1}$$

where the weights w_i are defined by the surfaces of the polygons (**Isaaks and Srivastava, 1989**). Therefore, isolated points will have larger polygons than clustered points. The external borders of the polygons are usually delimited by a convex hull. Programs called geographic information systems (GIS), allow the use of an additional information, and consequently more efficient determination of the mapped area's edges (**Dubois and Saisana, 2002**). Since the 1990s, Voronoi diagram has received increased attention as a spatial data model in GIS.

The purpose of using the Thiessen polygons can be shown in three cases: (a) when having an extremely large amount of data according to the range of analysed area, (b) with a very small number of data when the calculation of interpolation algorithm is doubtful, or (c) where the data are irregularly spaced.

An example of small data set is processed here using data from the Bjelovar Subdepression, specifically the Šandrovac Field which is its largest hydrocarbon field in that area. The Thiessen polygon method and Ordinary Kriging are used for subsurface mapping of e-log marker Z’ depths.

2. Methodology

Polygonal declustering, known as Thiessen polygon method, is based on the construction of polygons of influence about each of the sample data. These polygons (or diagrams) of influence are described by all midpoints between each neighbouring sample data (**Chow, 1964**).

Although there are many terms for Thiessen polygons, proximal polygons are sometimes proposed as a more articulate and unbiased term. They are constructed by computing the perpendicular bisectors among all neighbouring points of the set. A set *N* of *n* labelled points in the plane is called *centroid* (**Figure 1**). Thiessen *vertices* represent a set of points *V* in the plane in a way that each $V_i \in V$ is equidistant and closest to at least three centroids. A *Thiessen edge* may be delimited by two vertices or it may be unlimited in one direction. It represents the locus of all points equidistant and closest to two centroids. A *Thiessen polygon* is defined as the locus of all points closer to a centroid $C \in N$ than to any other centroid which implies that Thiessen polygons are convex. Thiessen diagram (known as Voronoi diagram) represents a set of all polygons (**Brassel and Reif, 1979**). Thiessen polygons may be closed or open. Closed polygons are fully enclosed by Thiessen edges, while open polygons may extend indefinitely.

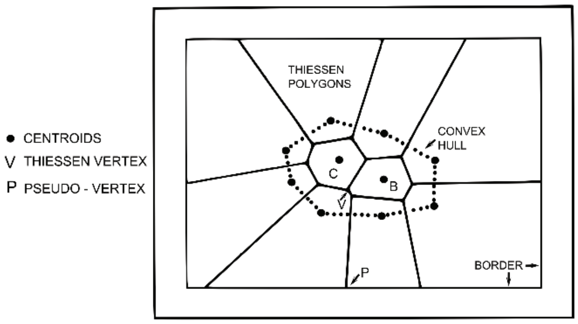

Figure 1: *Definition of terms used for Thiessen polygon computation (modified after Brassel and Reif, 1979).*

Slika 1: Izgled Thiessenovih poligona i elemenata potrebnih za njihovo definiranje (modificirano prema Brassel and Reif, 1979).

Shirriff (1993) described how to generate fractal patterns by recursively creating *Voronoi diagrams* on a set of points. This approach is useful when having abundant set of data. It is started with small set and the Voronoi diagram is drawn. Then, a denser set of points is used to compute a new Voronoi diagram inside each region of the first diagram. **Figure 2** demonstrates the procedure by using 10, 100, 1000, and 10000 points in four levels of recursion, so each polygon is subdivided into an average of l0 sub-polygons. The line thickness is half the size at each level to emphasize the higher levels.

In GIS, existing capabilities for generating Thiessen or Voronoi diagrams normally focus on the ordinary (not weighted) points. However, weighted diagrams for lines and areas are more useful in geosciences. For example, target areas in mineral exploration can be delineated depending on the size of alteration zones, and the zone of influence for active faults may be assessed based on the length of the fault segments (Dong, 2008).

Procedures mentioned above are available using ArcGIS. In this paper, SAGA GIS 2.1.0 software is used, the freeware at official SAGA GIS site. Generating Thiessen polygons and interpolation using Ordinary Kriging represents well-known procedures. Software produces Thiessen polygon network, the polygons being formed by the perpendicular bisectors of the lines joining nearby wells. The area of each polygon is determined and is used to weight the depth amount of the point in the centre of the polygon. Since the Thiessen polygonal method is flexible and straightforward, it is commonly applied in other scientific disciplines for the purpose of correcting for clustering in spatial data.

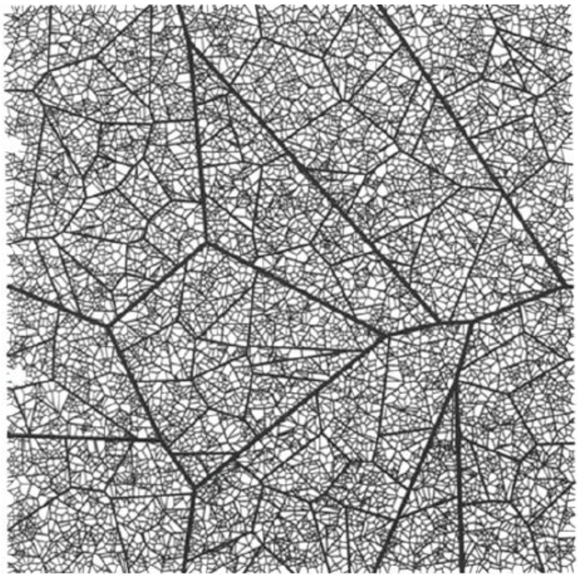

Figure 2*. A Voronoi fractal generated by subdivision by 10. The four procedure levels have 10, 100, 1000, and 10000 points (Shirriff, 1993).*

Slika 2: Voronoi fraktal dobiven podjelom s 10. Četiri razine postupka imaju 10, 100, 1000 i 10000 točaka (Shirriff, 1993).

3. Case Study

Here is given practical analyses and results of Thiessen and Kriging methods for data collected in the both analysed areas. The larger one was the Bjelovar Subdepression, i.e. existing palaeostructural maps of Neogene and Quaternary formations tops and bottoms. The smaller one was the Šandrovac Field as one of the most famous geological structure at the northern margin of subdepression.

3.1. Geography of the study area

The Šandrovac Field is located on the northern margin of the Bjelovar Subdepression (**Figure 3**). The field represents classical uplifted structure in the Croatian part of the Pannonian Basin System (abbr. CPBS), and it is geographically placed in the Bjelovar Subdepression that is a part of the Drava Depression. The present day topography is mostly influenced by the Bilogora Mt. (highest peak is 312 m), which was uplifted during the Pliocene, Pleistocene and Holocene. The field, i.e. the structure, margins cover about 39 km^2.

Figure 3. *Geographic location of the Šandrovac Field (modified after Balić et al., 2008).*

Slika 3: Geografski položaj polja Šandrovac (modificirano prema Balić et al., 2008).

3.2. Geology of the study area

The creation of the Pannonian Basin System (abbr. PBS) began about 16 million years ago in the Ottnangian when the Apulian plate started to subduct and converge under the Dinarides (e.g., **Malvić, 2003**; **Royden, 1988**). Numerous local transtensions as proto-areas of later depressions and subdepressions opened between the southern and northern boundaries of PBS (e.g., **Rögl, 1996; Malvić, 2003**). At present, the PBS area mainly covers the southern edge of the European plate and lesser part of the Internal Dinarides.

The CPBS is a part of the PBS bounded approximately by the Sava, Drava and Kupa Rivers (**Figure 3**). The 1st transtension (**Malvić and Velić, 2011**) lasted until the Middle Miocene, and resulted in the formation of elongated depressions (previously sometimes named as half-grabens) characterized by large sediment thicknesses strongly influenced by tectonics (**Pavelić , 2001**). During the Late Miocene clastic deposition continued its transition from marine to freshwater environment caused by the reduction of the sedimentary basin. At present, the CPBS occupies the entire south-western part of PBS (e.g., **Malvić, 2012**).

The opening of the Bjelovar Subdepression was a result of the activities of the main transcurrent fault system, along with other fault systems diagonal or transversal to subdepression (e.g., **Malvić, 2003, 2011**) or to the direction of the Central Drava Fault. As large geomorphological feature, the Moslavačka Mt. at present separates the Bjelovar Subdepression from the Sava Depression.

Thickness of Neogene-Quaternary sediments of the Bjelovar Subdepression rarely exceeds 3000 m, in contrast to 7000 m in the central Drava Depression zone. This is a consequence of the Bjelovar Subdepression not being situated on the main direction of transposal of materials, particularly in transpressional periods for which the sediment income was significantly lower. The maximum thickness is rarely larger than 3000 m. The 1st transtension sediments are breccias, conglomerates and sandstones. It is followed by limestone and marls that marked the transition from the 1st transtension to the 1st transpression. Monotonic change of sandstone and shale in different proportions took place in the 2nd transtension. Eventually, different weakly consolidated and unconsolidated clastics with sporadically peat and coal characterise the 2nd transpression. According to **Malvić and Velić (2011)**, such a chronological order indicates the transition of the sedimentary environment from marine (during the Badenian) to continental with smaller fluvial and lacustrine environments (during the Pleistocene and the Holocene).

3.3. Analysed datasets

The first dataset included well data of the Šandrovac Field (18 wells, an example of small data set). It was provided from doctoral thesis and an article published by **Malvić (2003, 2011)**. Field position is shown in **Figure 3**.

There are 5 e-log markers within the Bjelovar Subdepression (Rs7, Rs5, Z', Δ i D') and 1 EK-border (Tg/Pt), that were used for structural mapping (**Malvić 2003, 2011**). Those regional e-log markers are shown in **Figure 4**.

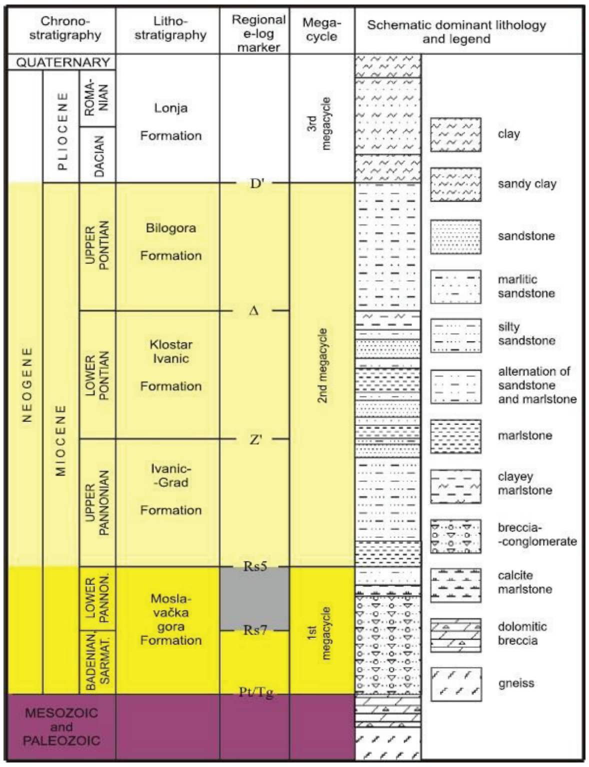

Figure 4. *Correlation between lithostratigraphic formation and depositional megacycles in the Drava Depression (Malvić, 2012).*

Slika 4: Međusobni odnosi litostratigrafskih formacija i taložnih megaciklusa u Dravskoj depresiji (Malvić, 2012).

The data are taken from the mentioned 18 wells where are previously determined the exact depths of Z' marker as border between Pannonian and Pontian. It is one of the most regional and chronostratigraphically important marker in the Drava Depression, situated approximately in the middle of the 2nd transtensional period. The observed wells are given inside the coloured field polygon on **Figure 5.**

The second dataset was the palaeostructural map of marker Z', however this time observed in the entire Bjelovar Subdepression (**Figure 5**) The map is interpolated with datum plane at +100 m.

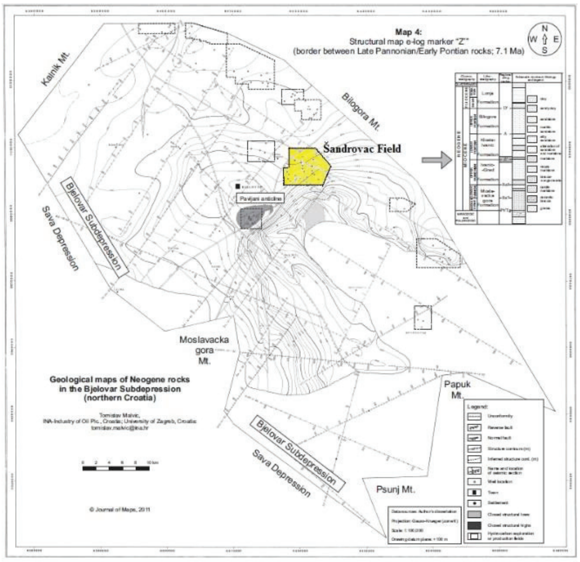

Figure 5: Structural map by e-log marker Z' of the Bjelovar Subdepression constructed by manual interpolation and drawing of faults (modified after Malvić, 2011).

Slika 5: Strukturna karta Bjelovarske subdepresije dobivena iz bušotinskih i seizmičkih podataka ručnom interpolacijom i ucrtavanjem rasjeda (modificirano prema **Malvić, 2011**).

Structural map by the e-log marker Z' was selected as an analytical source of regional data for this research. The reason for that is geological importance of the selected marker that divides the Pannonian and the Pontian stages, and both of them are typical proxies of lake deposition in the CPBS. Tectonically, they define period of 2nd transtension, when the huge quantities of clastics were transported from the Eastern Alps into CPBS (e.g., **Malvić, 2003, 2012; Malvić and Velić, 2011**). It was time of extremly large clastic detritus transport from the Eastern Alps to the CPBS and sedimentation of huge thickness of sandstones in alteration with basinal marls. At present such sequences are considered as two formations (Ivanić Grad and Kloštar-Ivanić) where thickenss of each of them can reach several hundred meters in numerous structures.

The Šandrovac Field is the largest hydrocarbon saturated structure in the Bjelovar Subdepression, with the main sandstone reservoirs in Lower Pontian sediments. Consequently, the e-log marker Z' has generally the largest significance in structural and petroleum-geological reconstruction for that field structure as well for the entire subdepression.

3.4. Analysis of data from entire subdepression

Data available from the structural map by the e-log marker Z' are collected using the regular grid with cell size of 1x1 km that covered the entire subdepression. So, it was possible to collect even 497 data for regional analysis, i.e. calculate reliable (representative) descriptive statistics of such dataset. The minimum and maximum depths are 60 and 2800 m, respectively. Median and mean are 610 and 854.92, respectively (**Figure 6**).

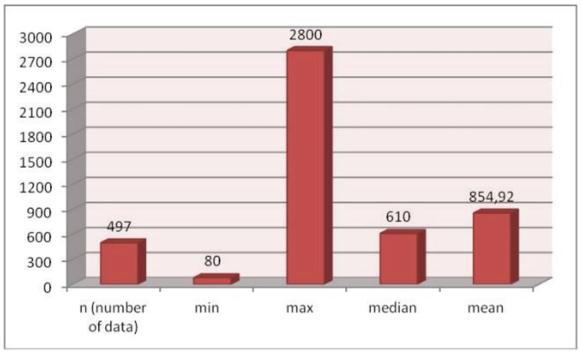

Figure 6. Statistical values obtained from structural map by e-log marker Z '.

Slika 6: Statističke vrijednosti dobivene iz strukturne karte po EK-markeru Z'.

Given the minimum and maximum range of depth values, the latter are divided into 10 classes, each 300 m in size. The interval size selection criterion was simple. The classes are of equal width, with first and last class that included dataset minimum and maximum. The number of 10 classes had been selected subjective as the number that author estimated as "the best" for visually attractive histogram. Although, with significantly lower or higher number of classes, too many or too low number of data would belong to each of them.

According to the calculation, most of the depth on the map today belongs to the class from 300 to 600 m depth. Almost 50% (49.23%) of data is at a depth up to 600 m (median is 610), i.e. into only first two classes (**Figure 7**).

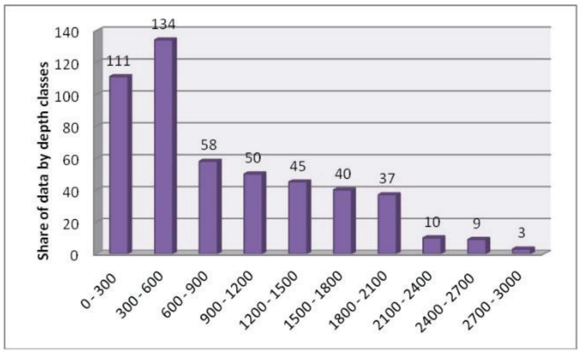

Figure 7. Statistics of depth classes obtained from structural maps by e-log marker Z'.

Slika 7: Raspodjela dubinskih razreda dobivenih iz strukturne karte po EK-markeru Z'.

3.5. Analysis of data from the Šandrovac Field

The data from the Šandrovac Field had been analysed using two mapping methods. Thiessen polygons of the Šandrovac Field (e.g., **Mesić and Medunić, 2014**) are made in SAGA GIS 2.1.0. software, using field coordinates from **Figure 5**. There were shown 18 well points, which are here used as "hard" data.

After entering the coordinates of the points, the Thiessen polygons are made. They are displayed using colours for specified depth class (**Figure 8**).

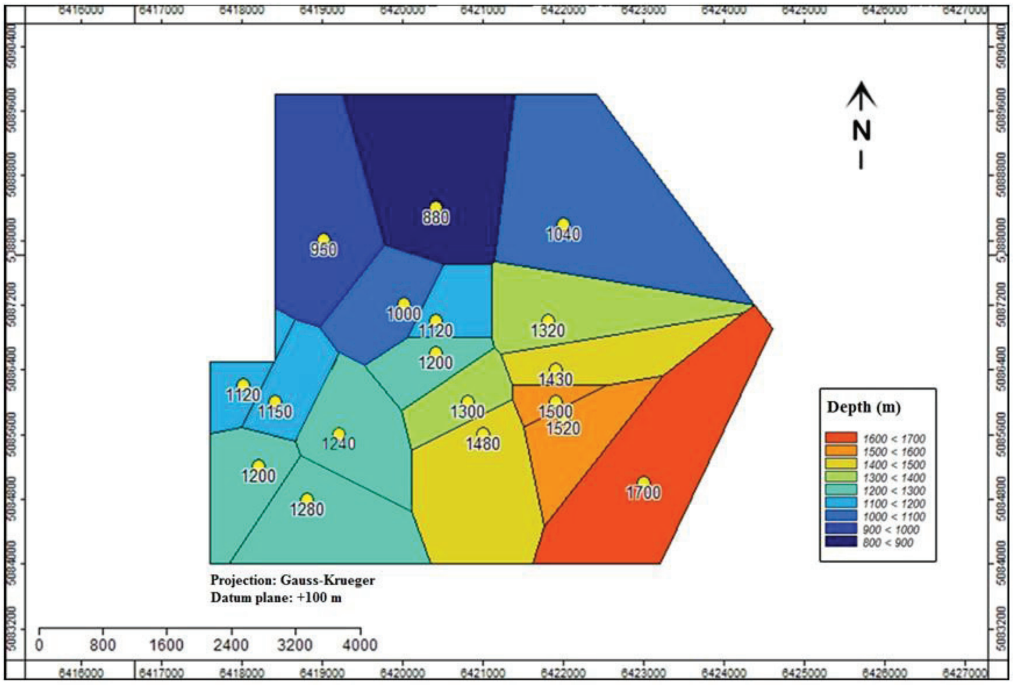

Figure 8: *Thiessen polygons in the Šandrovac Field displayed by colours corresponding to the depth values in particular points.*

Slika 8: Thiessenovi poligoni unutar polja Šandrovac prikazani bojama koje odgovaraju vrijednosti dubine u određenim točkama.

The Kriging map was created by using same software, and covering same points and approximately same area (**Figure 9**) as Thiessen polygons. In this case, due to relatively low number of points the "default" omnidirectional variogram model generated in software was used. Practically, it was not possible to decide where the variogram error would be lower with computer generated or manually fitted experimental curves. So, the simpler procedure had been selected.

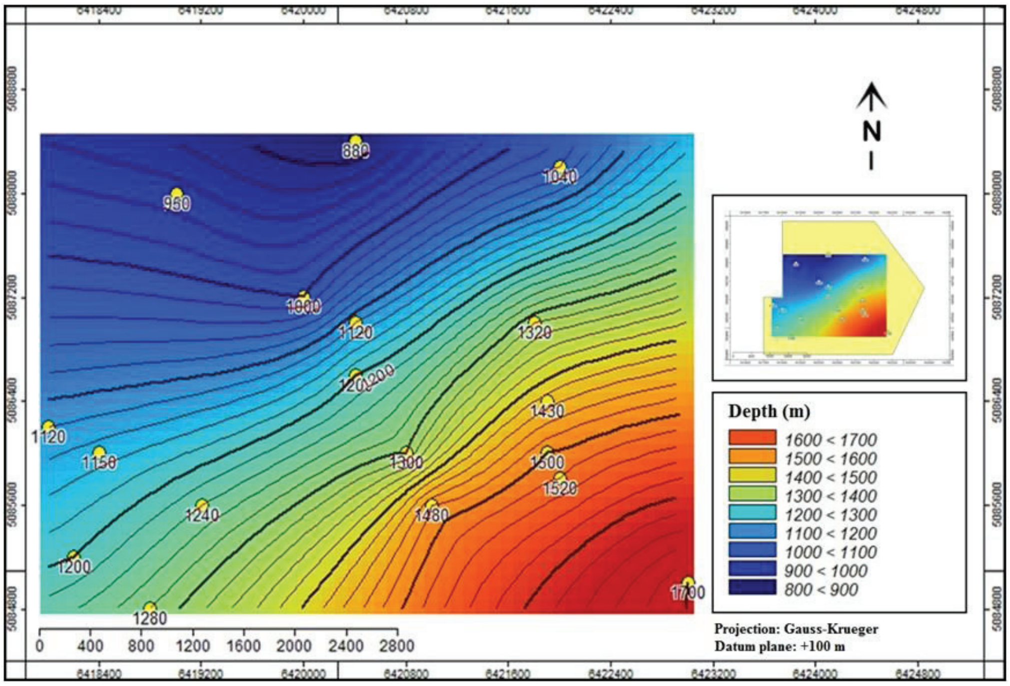

Figure 9: *Map interpolated by Kriging for the "depth" variable in the Šandrovac Field.*

Slika 9: Karta dobivena krigingom za varijablu „dubina" u polju Šandrovac.

4. Conclusion

Based on the available regional maps, it is relatively easy to establish which data points, representing e-log marker Z' depths, are located within the Šandrovac Field. Thiessen polygons map allowed a quick assessment of depth values over the large part of non-sampled areas. By means of the Thiessen polygon method a zonal assessment has been given. Accordingly, each polygon has a point value so that the maps created this way are not smoothing interpolation.

Making a depth map by the Ordinary Kriging technique resulted in a more detailed, regional, assessment of depths. Kriging provides a number of advantages over the other interpolation methods, and, in this case, where a zonal assessment had been applied, it is certain that the Kriging method provides continuous spatial depth value estimation.

However, the small number of input data is reflected in the determination of their spatial dependence, i.e. variogram model could be only assumed without any change to use reliable modelling technique. In such a case, the zonal assessment has its advantages, at least in fast obtaining the "maxima" and "minima" over the analysed area. Moreover, zonal method does not arbitrarily smooth and interpolate isolines, which is very uncertain process in the case of a small number of data. For that reason, it is recommended to consider for subsurface mapping of 20 or less data in the CPBS the application of both presented approaches, i.e. zonal estimation and deterministical interpolation. By merging their outputs it is possible to get a quick insight into the main structural feature(s) but also into the subareas where extremes are located.

5. References

BALIĆ, D., VELIĆ, J., MALVIĆ, T. (2008): Selection of the most appropriate interpolation method for sandstone reservoirs in the Kloštar oil and gas field, Geologia Croatica, 61(1), 27 - 35.

BOOTS, B.N. (1987): Modifying Thiessen Polygons, The Canadian Geographer, 31(2), 160 - 169.

BRASSEL, K.E. and REIF, D. (1979): A Procedure to Generate Thiessen Polygons, Geographical Analysis, 11(3), 289 - 303.

CHOW, V.T., 1964: Handbook of applied hydrology: A compendium of water-resources technology, McGraw-Hill, 1493 pp., New York.

DEUTSCH, C. V. (1989): DECLUS: a FORTRAN 77 program for determining optimum spatial declustering weights, Computers & Geosciences, 15(3), 325 - 332.

DONG, P. (2008): Generating and updating multiplicatively weighted Voronoi diagrams for point, line and polygon features in GIS, Computers and Geosciences, 34, 411 - 421.

DUBOIS, G. and SAISANA, M. (2002): Optimizing Spatial Declustering Weights - Comparison of Methods, Proceedings of the Annual Conference of the International Association for Mathematical Geology, Berlin-Germany, September 15-20, 2002, 479 - 548.

ISAAKS, E. and SRIVASTAVA, R. (1989): An Introduction to Applied Geostatistics, Oxford University Press Inc., 243 - 247, New York.

JOURNEL, A. G. (1983): Nonparametric estimation of spatial distributions, Mathematical Geology, 15(3), 445 - 468.

MALVIĆ, T. (2003): Naftnogeološki odnosi i vjerojatnost pronalaska novih zaliha ugljikovodika u Bjelovarskoj uleknini, Doktorska disertacija (Oil-Geological Relations and Probability of Discovering New Hydrocarbon Reserves in the Bjelovar Sag - in Croatian and English), Sveučilište u Zagrebu, Rudarsko-geološko-naftni fakultet, 123 pp.

MALVIĆ, T. (2008): Primjena geostatistike u analizi geoloških podataka, (Application of geostatistics in geological data analysis - in Croatian), INA, SD Istraživanje i proizvodnja nafte i plina, Grupa za tiskarske i grafičke poslove, 105 pp., Zagreb.

MALVIĆ, T. (2011): Geological maps of Neogen sediments in the Bjelovar Subdepression (northern Croatia), Journal of Maps, 7 (1), 304 - 317.

MALVIĆ, T. (2012): Review of Miocene shallow marine and lacustrine depositional environments in Northern Croatia, Geological Quarterly, 56(3), 493 - 504.

MALVIĆ, T. and 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. U. Schattner): 215-238.

MESIĆ, I. and MEDUNIĆ, G. (2014): Declustering of field's data located on northern margin of the Bjelovar Subdepression, Proceedings of the 6th Croatian-Hungarian and 17th Hungarian geomathematical congress "Geomathematics - from theory to practice", May 21-25, Opatija, 183 - 189.

OY LEUANGTHONG, KHAN, D.K, and DEUTSCH, C.V. (2008): Solved Problems in Geostatistics, John Wiley & Sons, Inc., 208 pp, New Jersey.

PAVELIĆ, D. (2001): Tectonostratigraphic model for the North Croatian and North Bosnian sector of the Miocene Pannonian Basin System, Basin Research, 13(3), 359-376.

RÖGL, F. (1996): Stratigraphic correlation of the Paratethys Oligocene and Miocene. Mitt. Ges. Geol. Bergbaust. Wien, 41, 65-73.

ROYDEN, L., H. (1988): Late Cenozoic Tectonics of the Pannonian Basin System, AAPG Memoir 45 (Chap. 3) - The Pannonian Basin (eds. Royden, L., H. & Horváth, F.), 27 - 48.

SHIRRIFF, K. (1993): Generating fractals from Voronoi diagrams, Comput. & Graphics, 17(2), 165 - 167.