



# A Comparison of Artificial Neural Networks and Ordinary Kriging depth maps of the Lower and Upper Pannonian stage border in the Bjelovar Subdepression, Northern Croatia

Marijan Šapina<sup>1</sup>

<sup>1</sup> Pavičeva 93, 31400 Đakovo, mag.ing.petrol.

## Abstract

Computerised mapping of subsurface strata is possible with a wide range of methods and techniques, such as geostatistical interpolation and stochastic simulations, but also with geomathematical methods. Geomathematical methods are, for example, the use of statistics in geology and the use of artificial neural networks. Artificial neural networks are primarily used in the case of flawed data and data that is in a non-linear relation. The set hypothesis of successful mapping of depth data using this original artificial neural network algorithm is confirmed using statistical analysis and comparison with geostatistical interpolation methods. The algorithm is made in „R“, an open source statistical computing software, and is used on the mapping of depth of the e-log marker „Rs5“ in the Bjelovar Subdepression, Northern Croatia, that is the border between the Lower and Upper Pannonian stages in the Croatian part of the Pannonian Basin System. The neural network architecture that produced the best responses is a network with two hidden layers, with 10 and 6 neurons, respectively. A backpropagation algorithm is used. Two methods were compared by cross-validation and the neural network produced a mean squared error as 16294.5, and Ordinary Kriging produced 14638.35.

## Keywords

artificial neural networks, Bjelovar Subdepression, Croatia, Neogene, mapping, Ordinary Kriging.

## 1. Introduction

The main objective of this paper is to prove that artificial neural networks (ANN) can be used for the mapping of any geological variable (in this case depth) as successfully as geostatistical interpolation methods. The data is processed and mapping is made using an original “R” source code. Since the data can be flawed or in a non-linear relationship, ANN can unite the data into one complex dataset. The characteristic of this method is that it simulates the learning process of human beings by training and optimizing parameters in a number of repetitions. The used geostatistical interpolation technique was Ordinary Kriging. The two methods were applied on depth mapping of the e-log marker “Rs5”. The e-log marker “Rs5” represents the border of the Moslavačka Gora Formation (Lower, Middle Miocene and Lower Pannonian sediments) and the Ivanić-Grad Formation (Upper Pannonian sediments), deposited inside the Bjelovar Subdepression, i.e. southwest of the Drava Depression (see **Figure 1**). This border had been previously regionally mapped by several authors (e.g. **Malvić, 2003**; **Malvić, 2011**; **Špelić et al., 2014**). The e-log marker “Rs5” as mapped in **Malvić (2011)** is presented in **Figure 2**.

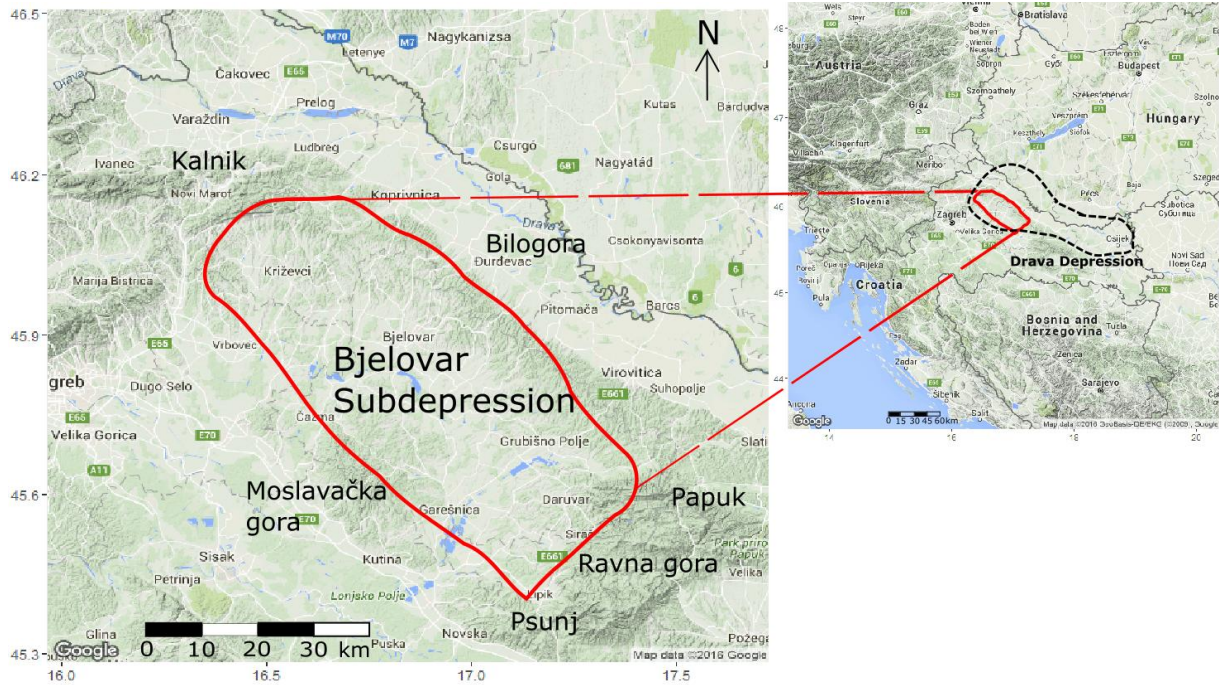


Figure 1: Geographical location of the Bjelovar Subdepression in the Drava Depression

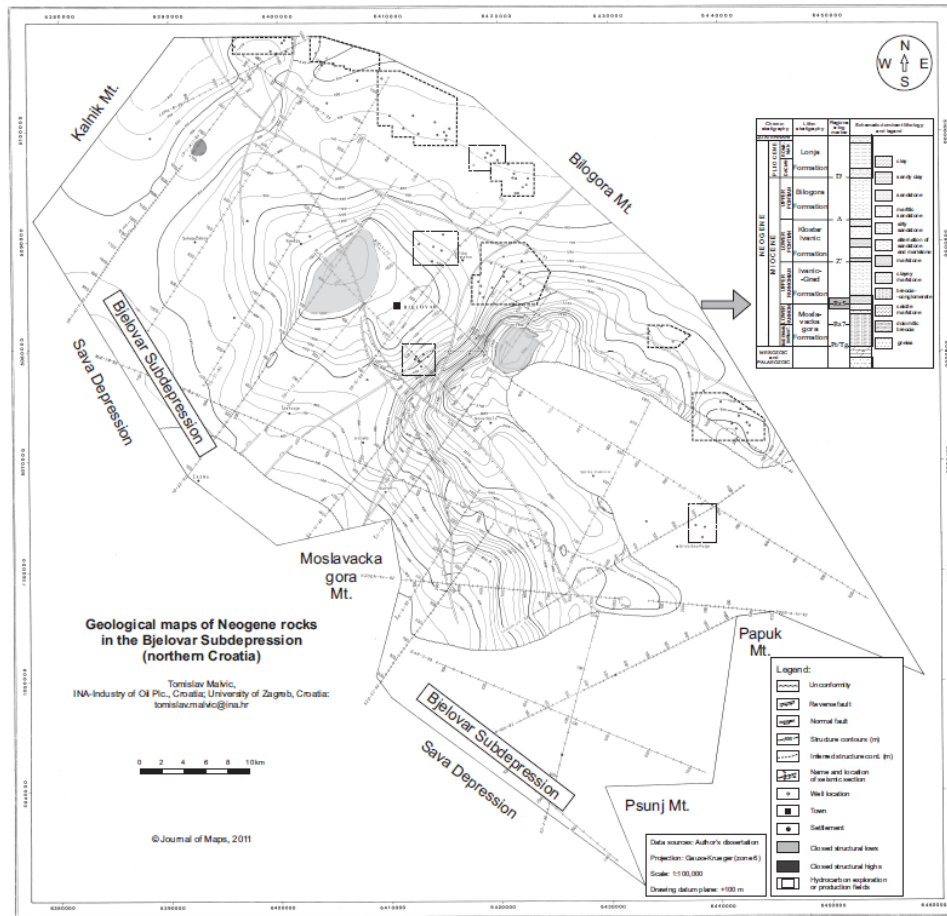


Figure 2: Structural map of the e-log marker “Rs5” (Malvić, 2011)

## 2. Basic geography and geology of the mapped area

The Croatian part of the Pannonian Basin System is located in the southwest of the geological macrounit called the Pannonian Basin System. Existing macrounits, i.e. depressions, in the Croatian part of the Pannonian Basin are the Mura, Drava, Sava and Slavonija-Srijem Depressions.

The Drava Depression is located in the northeastern part of Croatia and also extends to Hungary (see Figure 1). The total area is approximately 12000 km<sup>2</sup>, out of which 9100 km<sup>2</sup> is in Croatia (e.g., **Malvić & Cvetković, 2013**). The Bjelovar Subdepression, as a part of the Drava Depression, covers about 2900 km<sup>2</sup> (**Malvić, 2003**) in the Southwest. Geographically, it is surrounded with Kalnik Mountain in the Northwest, Bilogora Mountain in the Northeast, Papuk Mountain, Ravna Gora Mountain and Psunj Mountain in the East and Moslavačka Gora Mountain in the South (see **Figure 1**).

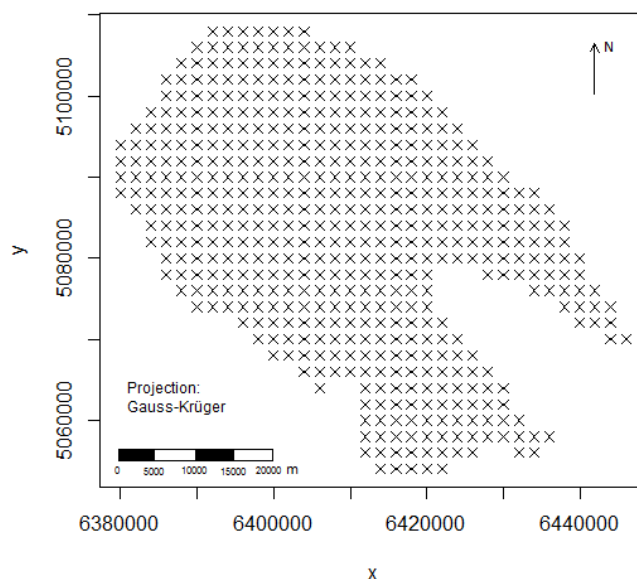
Lithology of the Drava Depression consists mostly of Neogene and Quaternary rocks and deposits. The total thickness of the deposits can be more than 7000 m in the middle part of the depression (**Velić, 2007**). Besides sedimentary rocks, volcanic Middle Miocene rocks can be found, along with Lower Miocene fluvial and lacustric sediments. Exploratory drilling in the Bjelovar Subdepression has revealed rocks that have been systematized in two lithological and chronostratigraphical groups. First, the younger, consists of Neogene-Quaternary rocks, and the second, the older, consists of Mesozoic and Palaeozoic magmatites, metamorphites and carbonates (e.g., **Malvić, 2003**).

Following the lithostratigraphic division (**Šimon, 1980**), the Lower and Middle Miocene rocks belong to the Moslavačka Gora Formation, which is further divided into the Mosti Member (of Badenian, Lower Miocene and Sarmatian age) and the Križevci Member (of Lower Pannonian age). The border with the Palaeozoic and Mesozoic rocks on the bottom is defined with the e-log border "Tg" and on the top with the Ivanić-Grad Formation with the "Rs5" e-log marker (see **Figure 2**). The next, younger, lithostratigraphic unit of the formation rank in the Drava Depression is named the Ivanić-Grad Formation. It borders with the Moslavačka Gora Formation on the top, with the e-log marker "Rs5" (analysed here), and with the Kloštar Ivanić Formation on the bottom. The E-log marker "Z" defines the bottom border. The Ivanić-Grad Formation is further divided into the Lipovac Marl member and the Zagreb Member or laterally equivalent Okoli Sandstones (e.g., **Šimon, 1980; Malvić, 2003**, see **Figure 2**). Lower Pontian sediments comprise the Kloštar Ivanić Formation. It borders with the Ivanić-Grad Formation on the bottom and the Bilogora Formation on the top with the „Δ“ e-log marker. The units of the member rank are the Lepsić Marl, followed by the Poljana Sandstones, the Graberje Marl, the Pepelana Sandstones and the Cabuna Marl (e.g., **Šimon, 1980; Malvić, 2003**, see **Figure 2**). The Bilogora Formation is of Upper Pontian age, and is not further divided into lithostratigraphic units of lower rank as the other older formations. It borders on the bottom with the Kloštar Ivanić Formation and on the top with the Lonja Formation with the e-log marker "D". The youngest deposits are defined as the Lonja Formation. Its age is defined as approximately Pliocene (Dacian and Romanian) and Quaternary. The border on the bottom is with the Bilogora Formation and on the top it is defined with the present terrain.

## 3. Mapping of the e-log marker "Rs5" using Artificial Neural Networks

The e-log "Rs5" marker has large significance in the interpretation of the Neogene depositional environment in the Croatian part of the Pannonian Basin System. Its regional character made it easy to recognise and map through almost the entire area of Northern Croatia. Also, as a chronostratigraphical border (Lower and Upper Pannonian) it could be considered as an approximate border between brackish and fresh-water lake environments as remnants of the Central Paratethys. This is the reason why the mapping methods were frequently tested on that regional marker bed, and the newest such evaluation of approach and mapping results was done with a neural network algorithm (ANN).

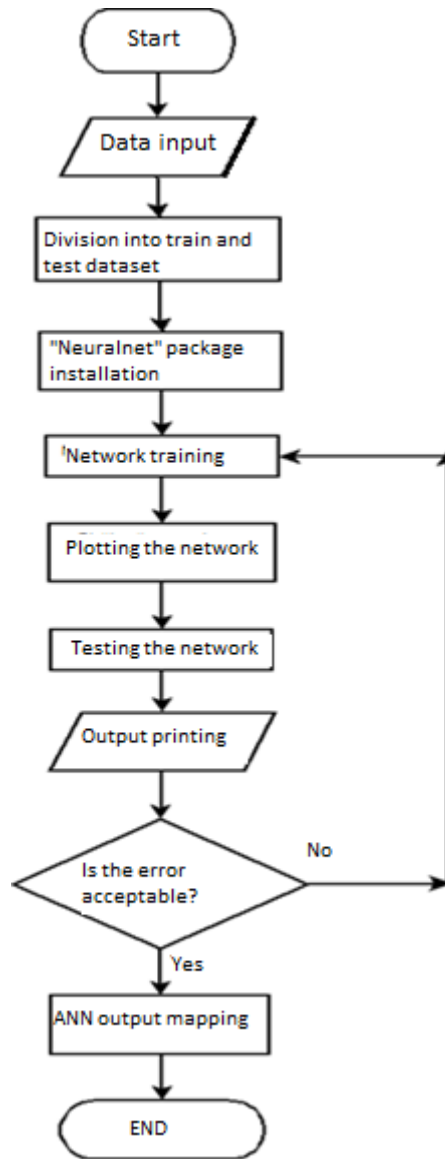
Since the ANN algorithm needs a large dataset for the learning process, the e-log marker "Rs5" depth data is obtained from **Malvić (2003)** and **Malvić (2011)**. The dataset can be obtained if new values for ANN are collected using a regular grid (e.g., using the methodology described in **Špelić et al., 2014**), which consists of  $x$  and  $y$  coordinates and depth value. **Figure 3** depicts the map with the position of every point that has depth value associated.



**Figure 3:** Location map of input values across the regular grid used on the existing regional map of the EL-marker “Rs5” in the Bjelovar Subdepression

The neural mapping of subsurface data in the Croatian part of the Pannonian Basin System had been published extensively for different structures and data types. One of the first attempts of neural network usage in the geological analysis of the Croatian geological systems was the clastic facies prediction in the Okoli Field in the Sava Depression (Malvić, 2006). The porosity prediction in the Beničanici Field in the Eastern part of the Drava Depression (e.g. Malvić & Cvetković, 2013a; Malvić & Prskalo, 2007) using artificial neural networks gave better results than the previously interpolated geostatistical porosity maps (Malvić & Đureković, 2003). The data obtained from well logs can be very useful in subsurface mapping. Such examples are the reservoir lithology and hydrocarbon saturation prediction (Cvetković et al., 2009) and the mapping of the Pliocene, Pleistocene and Holocene lithofacies mapping in the Sava Depression (Cvetković et al., 2012). Many of those examples are summarised in Malvić & Cvetković (2009) and Malvić & Cvetković (2013a).

Figure 4 shows the analysis flow chart of the ANN algorithm. The first step of successful mapping using the ANN algorithm is to gather and pre-process the dataset. To get the best output, i.e. the error as low as possible, Gauss-Krüger coordinates have been converted to *relative* coordinates, which didn't change the spatial relationship, but enabled easier handling with coordinate values, while the depth values were logarithmized with the base of 10 and numerically diminished, but the relationship remained the same.



**Figure 4:** Flow chart of the ANN algorithm

After pre-processing, the dataset is divided into two datasets, training and testing (validation), respectively. The training dataset consists of 75 % randomly chosen data from the entire set. The next step involves optimizing neural network parameters, such as threshold, number of repetitions (iterations) for the training process, starting weights, error function, algorithm type, etc. (e.g., **Günther & Fritsch, 2010; Malvić et al., 2008**). After optimization, the network calculates outputs. The new dataset with minimal total error will be used as final output and mapped. The new dataset is an artificial one with  $x$  and  $y$  coordinates from the existing dataset increased by 1 (i.e.,  $x_i+1$ ,  $y_i+1$ ) to get a larger number of points on the map and larger resolution. So, the resulting map includes twice as much point data as the map shown in Figure 3. When the depths are predicted with respect to the coordinates, the outputs of the old and the new datasets are combined and the map is made using the “ggplot2” package (see **Figure 5**). The final map consists of 1024 cells. The best network is chosen with regards to the resulting lowest total error. The chosen network has two hidden layers with 10 and 6 neurons, respectively. **Table 1** shows randomly chosen network outputs and a comparison with the original data. The value of the most overrated data is placed in the coordinates  $x=6431993$  and  $y=5059986$ . Instead of 900 m, the algorithm has predicted 2089.34 m. The most underrated data is the one placed in coordinates  $x=6429993$  and  $y=5081986$ . Instead of 3060 m it has predicted 2086.28 m.

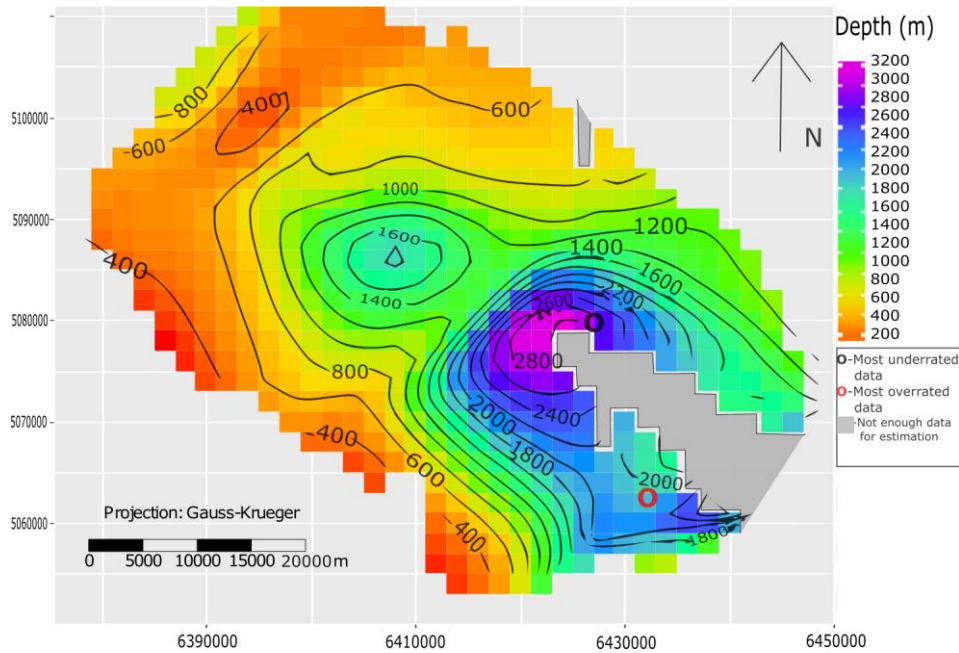


Figure 5: E-log border “Rs5” depth map made using the ANN algorithm

Table 1: Randomly chosen ANN input and output data

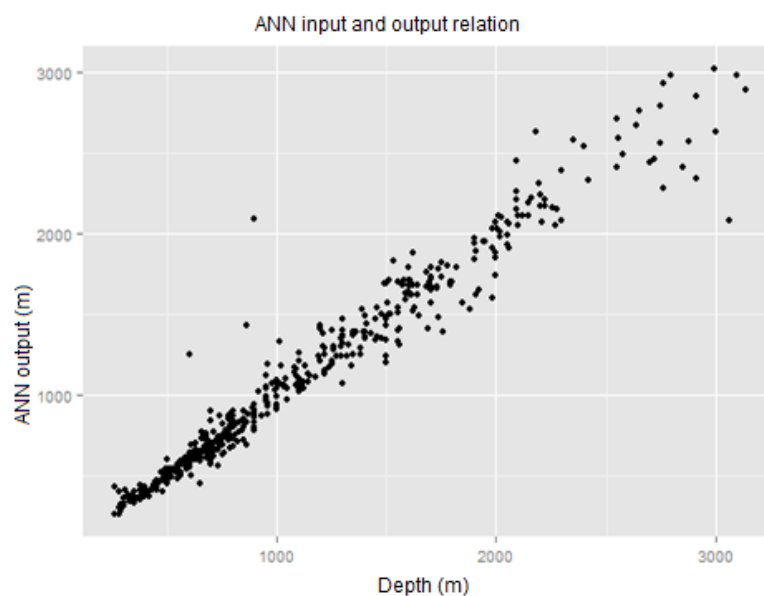
Relative coordinates <i>x</i>	Relative coordinates <i>y</i>	Depth (m)	Neural network output (m)
25	19	960	935.0805
24	20	850	804.0181
11	20	920	1023.6280
0	17	395	388.9871
1	20	520	521.9115
13	26	480	526.1691
2	14	290	289.4170
5	15	510	495.0765

Table 2 shows the results of reliability and correlation analysis. Based on the reliability analysis i.e. comparison of the output to the input data, ANN algorithm has an extremely high reliability coefficient of 0.99 with 99% confidence interval in the range from 0.988 to 0.991 ( $p < 0.001$ ). Using the correlation analysis, it is visible that there is a statistically significant ( $p < 0.001$ ) positive correlation with the correlation coefficient of 0.91 with a 99% confidence interval in the range from 0.895 to 0.92.

**Table 2:** Reliability and correlation analysis results

		99 %-C.I.		p
$\alpha^*$	0.99	0.988	0.991	<0.001
$\tau^\dagger$	0.91	0.895	0.92	<0.001
*Cronbach alfa, †Kendall tau				

The diagram in Figure 6 presents the relation of input and output data. Clear consistency can be seen in the lower values, but there is some dispersal in the higher values. Some outliers are visible, but thorough data analysis showed that they are a consequence of extreme changes in depth values on small distances and the ANN algorithm could not adapt to them. High reliability and correlation are indicators of the excellence of this algorithm and of a potential application of neural networks in strata mapping in an analysed area.

**Figure 6:** Input and output relation diagram

#### 4. Interpolation of depths of the e-log marker “Rs<sub>5</sub>” using the Ordinary Kriging technique

Geostatistical interpolation is carried out using the Ordinary Kriging and variogram analysis preceded the interpolation. The numerous Kriging interpolations have been done over the last decade for the Croatian part of the Pannonian Basin System. Some of the relevant geostatistical interpolations of the fields in the Drava Depression are the porosity maps of the Beničanci field, and the Eastern part of the Drava Depression, were made using Kriging and inverse distance (Malvić, 2008a; Malvić & Đureković, 2003) and in the Sava Depression, porosity maps were made for the Ivanić field (Malvić, 2008b) and the Kloštar field (Balić et al., 2008).

##### 4.1. Variogram analysis

A variogram that will be used in further interpolation has been made in the “Variowin” program (Pannatier, 1996). The first step was defining parameters with respect to the input dataset, and then the experimental variogram was made. The experimental variogram is presented in **Figure 7**.

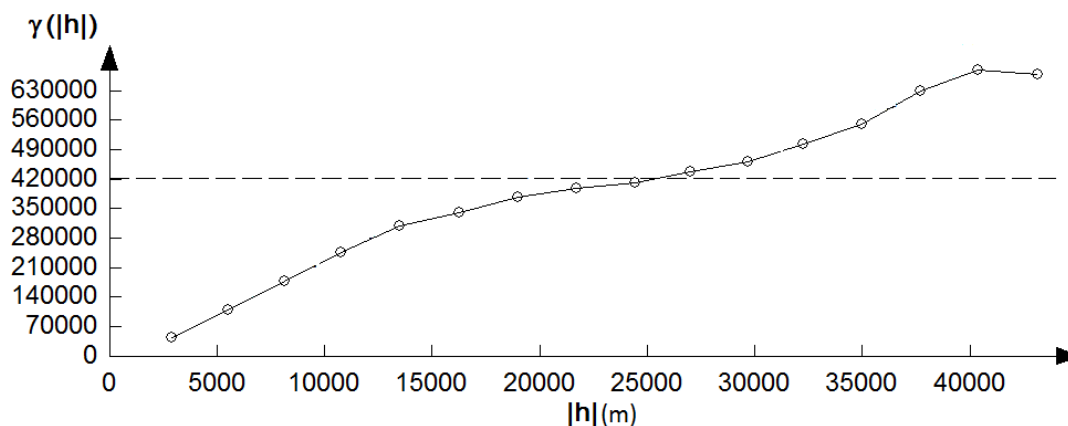


Figure 7: Experimental variogram with the line connecting calculated values at variogram classes (lags)

The next step is the approximation of the experimental variogram using one of the theoretical models (spherical, exponential, Gaussian) to reach the Kriging approach as the best linear interpolator (e.g. **Hohn, 1988; Malvić, 2008; Journel & Huijbregts, 1978; Krige, 1951; Liebhold et al., 1993**). This variogram has been approximated with the spherical model (see **Figure 8**) defined by sill 425,514 and the range 24,684 m.

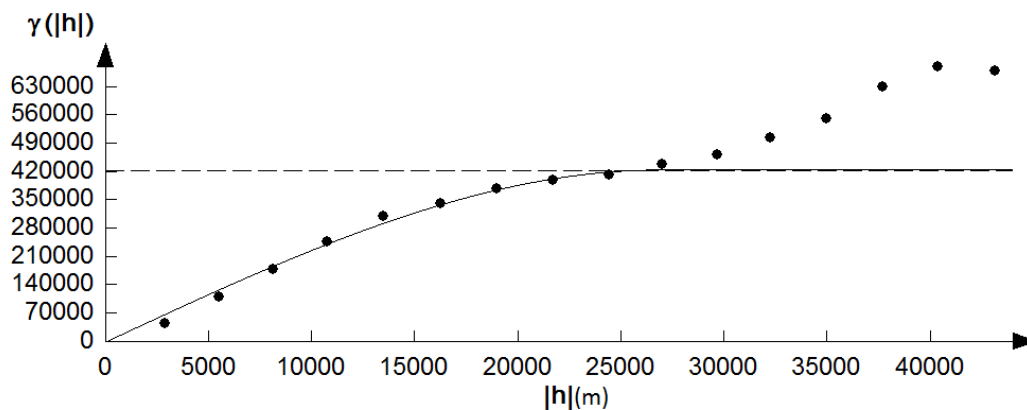


Figure 8: Approximation of the variogram with spherical theoretical model

#### 4.2. Interpolation using Ordinary Kriging

The map was made in Surfer 9, and the input variables were coordinates of the e-log marker “Rs5” ( $x$  and  $y$ ) and the depth values. The variogram value is the most important input Ordinary Kriging equations. The two methods (Kriging and ANN) were compared using cross-validation (e.g., **Davis, 1987**). Their mean squared errors (MSE) have been compared and the most underrated and overrated data is presented.

The result of mapping using Ordinary Kriging is presented in **Figure 9**. The MSE of Ordinary Kriging is 14,638.35. The most overrated data is placed in coordinates  $x=6431992$  and  $y=5059986$ . The predicted depth is 2161.50 instead of 900 m. The most underrated data is placed in coordinates  $x=6429993$  and  $y=5081986$ . Instead of 3060 m it has predicted 2271.39 m.



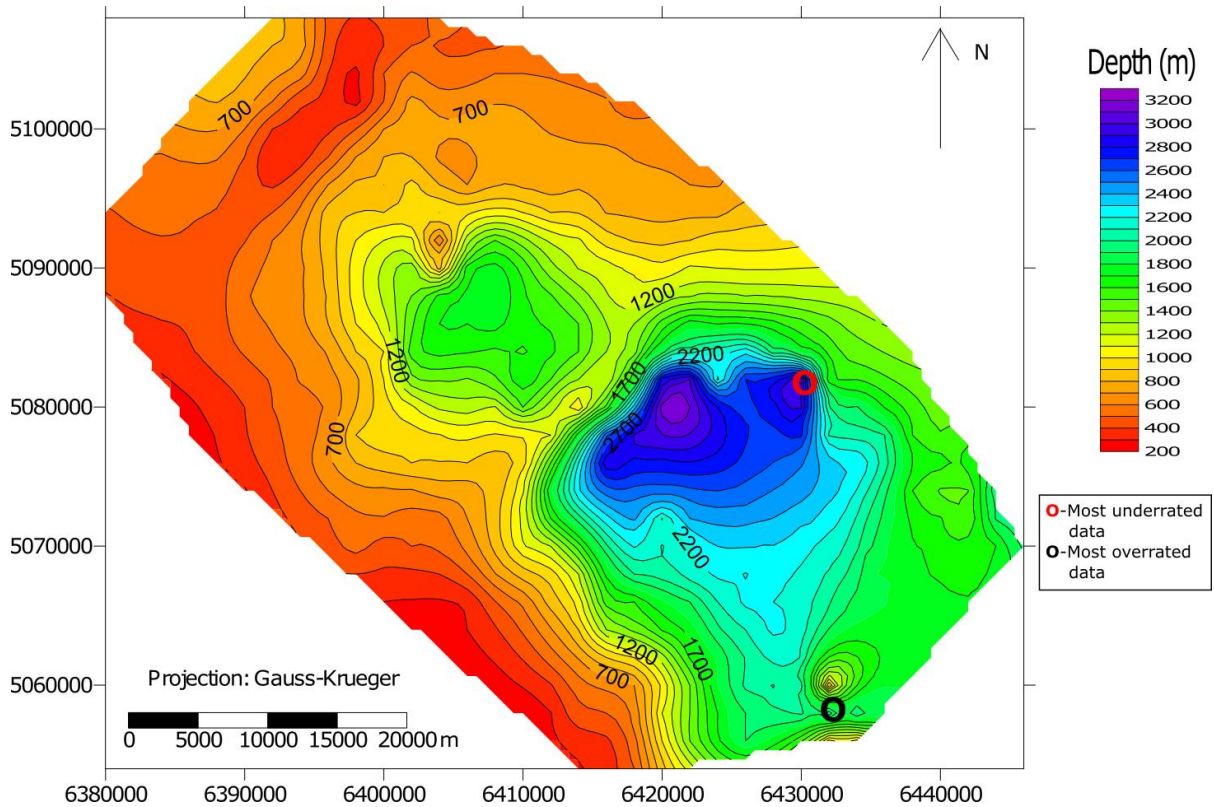


Figure 9. E-log border „Rs5“ depth map made using Ordinary Kriging

## 5. Discussion and conclusion

In this paper, I confirmed the hypothesis of successful mapping using ANN. The mapping is based on depth mapping of the e-log marker “Rs5” in the Bjelovar Subdepression, as part of the Drava Depression in the Croatian part of the Pannonian Basin System. The best results are obtained by ANN architecture of two hidden layers with 10 and 6 neurons, respectively. A multilayer perceptron has been used with every layer fully connected to the next one. A backpropagation algorithm has been used.

The input dataset that have been used were coordinates converted from the Gauss-Krüger coordinate system to *relative* coordinates so that better output is gained and the error is smaller. The depth values range from 260 to 3140 and are diminished by logarithmizing with the same purpose, to diminish the error. If using datasets marked with a linear relationship, neural networks should be compared with other, often successful, linear mapping methods. The most well-known of them is Kriging. Although intended for analyses of non-linear relationships, here the ANN in comparison with them gives good results which are supported by reliability and correlation analysis.

The ANN algorithm applied here could probably give valid output in mapping any other geological parameter (porosity, permeability, saturation, depth) in the area of the Bjelovar Subdepression. Of course, it needs to be trained on valid datasets and in that case the error could be low. The largest problem in successful network building is (1) the dataset editing and (2) network parameter optimization. Both can be time-consuming. Here, the largest error in network training resulted from highly different depth values on relatively small horizontal distances (a few dozen meters) that mostly came from a location placed on opposite fault walls (footing and hanging).

The ANN method used 511 input values of the original dataset and another 511 data from the new dataset ( $x+I, y+I$ ). The Ordinary Kriging method used the original dataset with 511 data. The absolute value of the most overrated and the most underrated value in the ANN method is 2163.06, and in Ordinary Kriging 2050.11. The ANN contains a relatively small amount of data in comparison with Kriging, where the additional data is obtained with interpolation. Because of

that, the ANN has a somewhat larger MSE (16294.5 in respect to 14638.35 in Ordinary Kriging). The absolute value of the most overrated and the most underrated value in the ANN method is 2163.06, and in Ordinary Kriging 2050.11. Due to a different way of mapping (cell estimation vs. stratoisohypses interpolation) the neural map is grainy. Both problems can probably be solved with an increase in the number of data cells in neural networks. In that case, there would be more data on a smaller distance, which would decrease the difference in adjacent values and a “transitional” trend can be observed more easily by ANN. Such network would “learn” and in spite of increasing the overall MSE, it would be smaller than that in Ordinary Kriging. Here, I show that the application of ANN in the mapping of depth of the Neogene geological strata in the Bjelovar Subdepression can be a reasonable approach. It is recommended as probably the best approach if the data (e.g., geological variable and depth) is not strongly linearly related, i.e. linear correlation is not high and/or significant.

## 6. 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.
- Cvetković, M., Velić, J., Malvić, T. (2009). Application of neural networks in petroleum reservoir lithology and saturation prediction. *Geologia Croatica: journal of the Croatian Geological Survey and the Croatian Geological Society*. 62, 2, 115-121.
- Cvetković, M., Velić, J., Malvić, T. (2012). Application of Artificial Neural Networks on Well Log Data for Lithofacies Mapping of Pliocene, Pleistocene and Holocene Sediments in Sava Depression, Pannonian Basin, Croatia. XIth International Conference on Geoinformatics – Theoretical and Applied Aspects, Kiev, Ukraine 14-17. 5. 2012. Kiev, Ukraine – Geoinformatics 2012 Kyiv: EAGE, 2012, 1-5.
- Davis, B (1987). Uses and Abuses of Cross Validation in Geostatistic. *Mathematical Geology*, 19, 3, 241-248.
- Günther, F., Fritsch, S. (2010). Neuralnet: Training of neural networks. *The R Journal*. 2, 1, 30-38.
- Hohn, M. E. (1988). *Geostatistics and Petroleum Geology*. Van Nostrand Reinhold, New York, 400 pp.
- Journel, A. G., Huijbregts, C. J. (1978). *Mining Geostatistics*. Academic Press, London, UK, 600 pp.
- Krige, D. G. (1951). A Statistical Approach to Some Basic Mine Valuation Problems on the Witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, 52, 119-139.
- Liebhold, A. M., Rossi, R. E., Kemp, W. P. (1993). Geostatistics and Geographic Information System in Applied Insect Ecology. *Annual Review of Entomology*, 38, 303-327
- Malvić, T. (2003). Naftnogeološki odnosi i vjerojatnost pronalaska ugljikovodika u Bjelovarskoj uleknini [Oil-geological relations and probability of discovering new hydrocarbon reserves in the Bjelovar Sag], Ph.D. Dissertation, University of Zagreb, Faculty of Mining, Geology and Petroleum Engineering, Zagreb, Croatia, 123 pp.
- Malvić, T., Đureković, M. (2003). Application of methods: Inverse distance weighting, ordinary kriging and collocated cokriging in porosity evaluation, and comparison of results on the Beničanci and Stari Gradac fields in Croatia. *Nafta*. 54, 9, 331-340.
- Malvić, T. (2006). Predviđanje klastičnih facijesa neuronskom mrežom (primjer polja Okoli) (Clastic facies prediction using neural networks (case study from Okoli field)). *Nafta*, 57, 10, 415-431.
- Malvić, T., Prskalo, S. (2007): Using of neural network in porosity prediction (Beničanci field). XI. Congress of Hungarian Geomathematics, Geiger, Janos (ed.). Szeged: Hungarian Geological Society, Geomathematical Section of the Hungarian Geological Society.
- Malvić, T. (2008a). Kriging, cokriging or stohastical simulations, and the choice between deterministic or sequential approaches. *Geologia Croatica*. 61, 1, 37-47.
- Malvić, T. (2008c): Izrada karte poroznosti krigiranjem u pješčenjačkim ležištima, primjer iz Savske depresije (Production of a Porosity Map by Kriging in Sandstone Reservoirs, Case study from the Sava Depression). *Kartografija i geoinformacije*, 7, 9, 12-19.
- Malvić, T. (2008c). Primjena geostatistike u analizi geoloških podataka (Application of Geostatistics in Geological Data Analysis – in Croatian), INA-Industry of Oil Plc., 103 pp.
- Malvić, T. (2011). Geological maps of Neogene sediments in the Bjelovar Subdepression (northern Croatia). *Journal of Maps*, 7 (1), 304-317.
- Malvić, T., Cvetković, M., Balić, D. (2008). Geomatematički rječnik (Geomathematical dictionary). Hrvatsko geološko društvo, Zagreb.
- Malvić, T., Cvetković, M. (2009). Neuronski alati u geologiji ležišta ugljikovodika, I. izdanje (Neural tool in geology of hydrocarbon reservoirs, 1st issue). Hrvatsko geološko društvo (Geomat. odsjek), Zagreb.
- Malvić, T., Cvetković, M. (2013). Lithostratigraphic units in the Drava Depression (Croatian and Hungarian parts) - a correlation / Korelacija litostratigrafskih jedinica u Dravskoj depresiji (hrvatski i mađarski dio). *Nafta*. 64 (1), 27-33.
- Malvić, T., Cvetković, M. (2013a). Neuronski alati u geologiji ležišta ugljikovodika, II. Izdanje (Neural tool in geology of hydrocarbon reservoirs, 2nd issue). Hrvatsko geološko društvo (Geomat. odsjek), Sveučilište u Zagrebu (RGN fakultet), sveučilišni priručnik, Zagreb.
- Pannatier, Y. (1996). *VARIOWIN – Software for spatial data analysis in 2D*. New York: Springer, 91 pp.
- Šimon, J. (1980): Prilog stratigrafiji i taložnom sustavu pješčenih rezervoara Sava- grupe mlađeg tercijara u Panonskom bazenu Sjeverne Hrvatske. Disertacija, Sveučilište u Zagrebu, Rudarsko-geološko-naftni fakultet, 56+82 pp., Zagreb.
- Špelić, M., Malvić, T., Saraf, V., Zalović, M. (2014). Remapping of depth of e-log markers between Neogene basement and Lower/Upper Pannonian border in the Bjelovar Subdepression, *Journal of Maps*, DOI: 10.1080/17445647.2014.978909
- Velić, J. (2007). *Geologija ležišta nafte i plina (Geology of Oil and Gas)*. Sveučilište u Zagrebu, Rudarsko-geološko-naftni fakultet, Zagreb. 342 pp.

## Sažetak

### Usporedba karata dubina dobivenih metodom umjetnih neuronskih mreža i običnim krigiranjem u donjemu i gornjemu panonu unutar Bjelovarske subdepresije u sjevernoj Hrvatskoj

Računalno kartiranje podzemlja moguće je korištenjem velikoga broja metoda i tehnika, poput geostatističke interpolacije, stohastičkih simulacija, ali također i geometematičkih metoda, poput primjene umjetnih neuronskih mreža. Umjetne neuronske mreže primarnu primjenu imaju u slučaju manjkavih podataka te podataka koji su u nelinearnome odnosu. Statističkom analizom i usporedbom s geostatističkim interpolacijskim metodama potvrđena je hipoteza uspješnoga kartiranja dubina primjenom izvornoga algoritma umjetnih neuronskih mreža. Algoritam je izrađen u programu za statističku obradu podataka „R” i korišten je za izradu karte dubina EK markera „Rs5” unutar Bjelovarske subdepresije u sjevernoj Hrvatskoj. EK marker „Rs5” granica je donjega i gornjega panona u hrvatskome dijelu Panonskoga bazenskog sustava. Arhitektura neuronske mreže koja je dala najmanju pogrešku jest ona mreža s dvama skrivenim slojevima, gdje je prvi imao 10 neurona, a drugi 6. Korišten je algoritam s povratnim postupkom. Metode običnoga krigiranja i umjetnih neuronskih mreža uspoređene su metodom krosvalidacije, gdje je neuronska mreža dala zbroj kvadrata pogreške u iznosu od 16 294,5, a obično krigiranje 14 638,35.

#### Ključne riječi

Bjelovarska subdepresija, Hrvatska, kartiranje, obično krigiranje, umjetne neuronske mreže