

Article

Application of the bootstrap method in data analysis of Upper Miocene sandstone reservoirs of the western part of the Sava Depression

Josip Ivšinić^{1,*}, Tomislav Malvić² and Dubravka Pleše²

¹ INA-Industry of Oil Plc., Avenija Većeslava Holjevca 10, 10000 Zagreb; josip.ivsinovic@ina.hr (J.I.) (<http://orcid.org/0000-0002-7451-1677>)

² University of Zagreb, Faculty of Mining, Geology and Petroleum Engineering; tomislav.malvic@rgn.hr (T.M.) (<http://orcid.org/0000-0003-2072-9539>); dubravka.plese@rgn.hr (D.P.)

* Correspondence: josip.ivsinovic@ina.hr (J.I.)

Abstract: In deep geological analysis of data, these are input data that are few and include a small set of data. In a small set of case data, it is necessary to obtain reliable data of individual geological variables from this type of data. The paper analyzes the possibility of applying the bootstrap method on variables that are important in the exploration and production of hydrocarbons. The variables analyzed were the following: porosity and total costs of disposal formation water. The case study was made on the data of reservoir "K", field "B" located in the western part of the Sava Depression. The analysis of the results showed the possibility of applying the bootstrap method in the analysis of deep geological data with the application of three different sizes of resampling dataset.

Keywords: Sava Depression; bootstrap; geostatistics; disposal formation water; porosity

1. Introduction

The bootstrap is convenient method to support design of experiment analysis when non-Gaussian data are met i.e. sample is small or a normality test failed [1]. It has a wide application in all branches of science. Authors [2] applied the bootstrap method to determine assessing metasomatic mass and volume changes for deep crustal hydrothermal alteration of marble. [3] test a model-free bootstrap (MFB) method that can be used for any time series of data. The author's [4,5] applied bootstrap method in risk management. Bootstrap method also applied for measure agglomeration of manufacturing industries at the county level in the United States [6]. In the fishing industry, the application of the bootstrap method has applied to the example of the fish population in Iceland [7]. A bootstrap method for estimating uncertainty of water quality trends for Susquehanna River at Conowingo (U.S.A.) used by [8]. Bootstrap analyses performed for three rainfall events in the upstream of the Qingjian River basin, a sub-basin of the Yellow River [9]. The bootstrap interval data used for 24 hours' annual maximum precipitation records obtained from 21 meteorological stations in Mexico [10]. [11] also use it in production processes capability assessment. When analyzing geological data in a large number of cases, it is a small set of numbers, especially when analyzing deep geological data. So far, the "Jack-knifing" method has been successfully applied in the Croatian part of the Pannonian Basin in order to increase the sample size [12,13]. Therefore, the application of the statistical tool of the bootstrap method is very important in the analysis of the obtained geological data. These papers used as a theoretical basis for calculating bootstrap interval values for examples of costs of disposal formation water and porosity.

2. Methods

3.1. Mathematical settings of the bootstrap method

Before calculating bootstrap statistics, it is necessary to resample the data from the input data set. For nonparametric resampling, it is not necessary to know the type of distribution, model and sample size. During the resample process, the size of the input data set does not change, but within the same input data set is randomly changed with the same data from the input set (Figure 1).

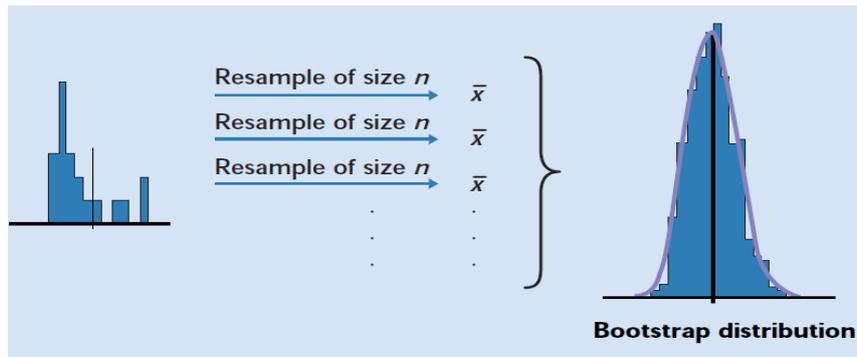


Figure 1. Creating a bootstrap data set by resampling process [14]

There are several bootstrap methods described in the literature, given the usual deep geological data the most suitable method to apply is smooth bootstrap. For each new set of data obtained by the random selection method, the arithmetic mean is calculated by [15,16]:

$$\bar{X}_i = \frac{1}{n} \sum_{i=1}^n X_i \quad (1)$$

where are: \bar{X}_i - mean sample value (after resampling), X_i - data set after resampling, n-sample size.

From the calculated mean values of individual resample data sets, a new set of numbers where obtained. From this new set of numbers, bootstrap statistics where calculated, so mathematical expression for arithmetic bootstrap means [17,18]:

$$\bar{X}_m = \frac{1}{m} \sum_{i=1}^m \bar{X}_i \quad (2)$$

where are: \bar{X}_m - arithmetic bootstrap mean, \bar{X}_i - mean sample value (after resampling) of an individual data set, m- number of resampling data set.

Standard deviation after m of resampling's where calculated according to the expression [19]:

$$S_m = \sqrt{\frac{1}{m} \sum_{i=1}^m (\bar{X}_i - \bar{X}_m)^2} \quad (3)$$

where are: S_m - standard deviation of bootstrap, \bar{X}_m - arithmetic bootstrap mean, \bar{X}_i - mean sample value (after resampling) of an individual data set, m- number of resampling data set.

By calculating the mean and standard deviation of the bootstrap data set, the last step is an interval estimate of the data set expectations:

$$\left\langle \bar{X}_m - z \frac{S_m}{\sqrt{m}}, \bar{X}_m + z \frac{S_m}{\sqrt{m}} \right\rangle \quad (4)$$

where are: S_m - standard deviation of bootstrap, \bar{X}_m - arithmetic bootstrap mean, z - value from the normal distribution, m - number of resampling data set. The most common reliability of the estimate of the interval is 90% or 95%. The procedure repeated from expressions (1) to (4) according to the number of resampling samples made, in order to obtain a reliable interval of data set values. The most common number of resampling data sets is 1000 [20] and 2000 [20,21] to ensure a 95 percent confidence interval of data set.

2.2. Research area and geological settings

The Sava Depression is an integral part of the Croatian Pannonian Basin System (CPBS). The subject of the analysis is field B, i.e. reservoir "L" which is located in the western part of the Sava Depression (Figure 2). The subject of the analysis is field B, i.e. reservoir "K" which is located in the western part of the Sava Depression (Figure 2).

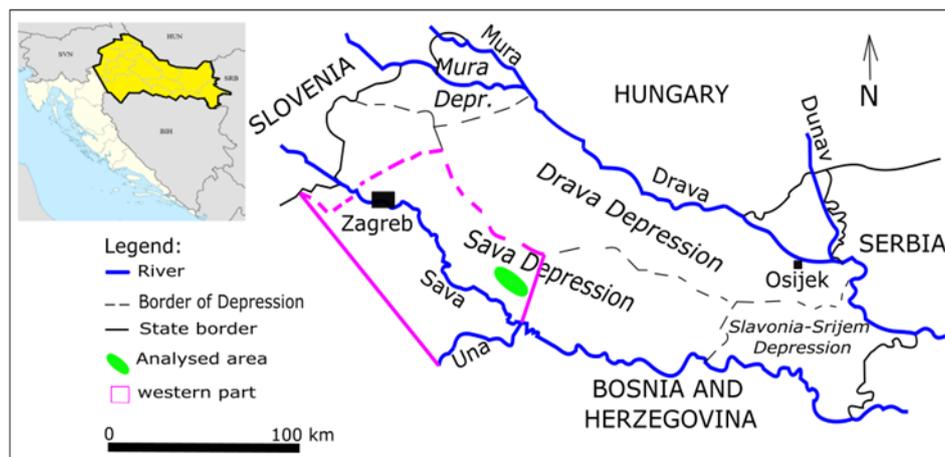


Figure 2. Geographical position of the analysed field "B" [22]

In order to understand the development of the typical sandstone reservoirs Neogene age in Northern Croatia (Croatian part of Pannonian pool system, abbreviated CPBS) is shown here, briefly, depositional model of the area. Tectonics is the cause of today's shape and diversity of deposits, and the characteristic regional provision of structures northwest - southeast. As the location of CPBS depressions was always on the edge of the Pannonian basin system (PBS) area, their area was mostly covered by a shallow marine and lake environment, often isolated from larger bodies of water in the central part of the PBS. Clastic sedimentary environments are dominant throughout the CPBS area, from the Baden to the Lower Pontian. At that time, extremely large amounts of sandy and silty detritus were transmitted by turbidite currents to the area. In periods when such activity of turbidite currents did not exist, mainly different species of marl were deposited, both in the marine and lake environment. Therefore, marls are considered rocks of a "quiet" environment, i.e. an environment of

"low" energy. The consequence of such developments is today's hydrocarbon deposits discovered in Upper Pannonian and Lower Pontian sandstones (Figure 3).

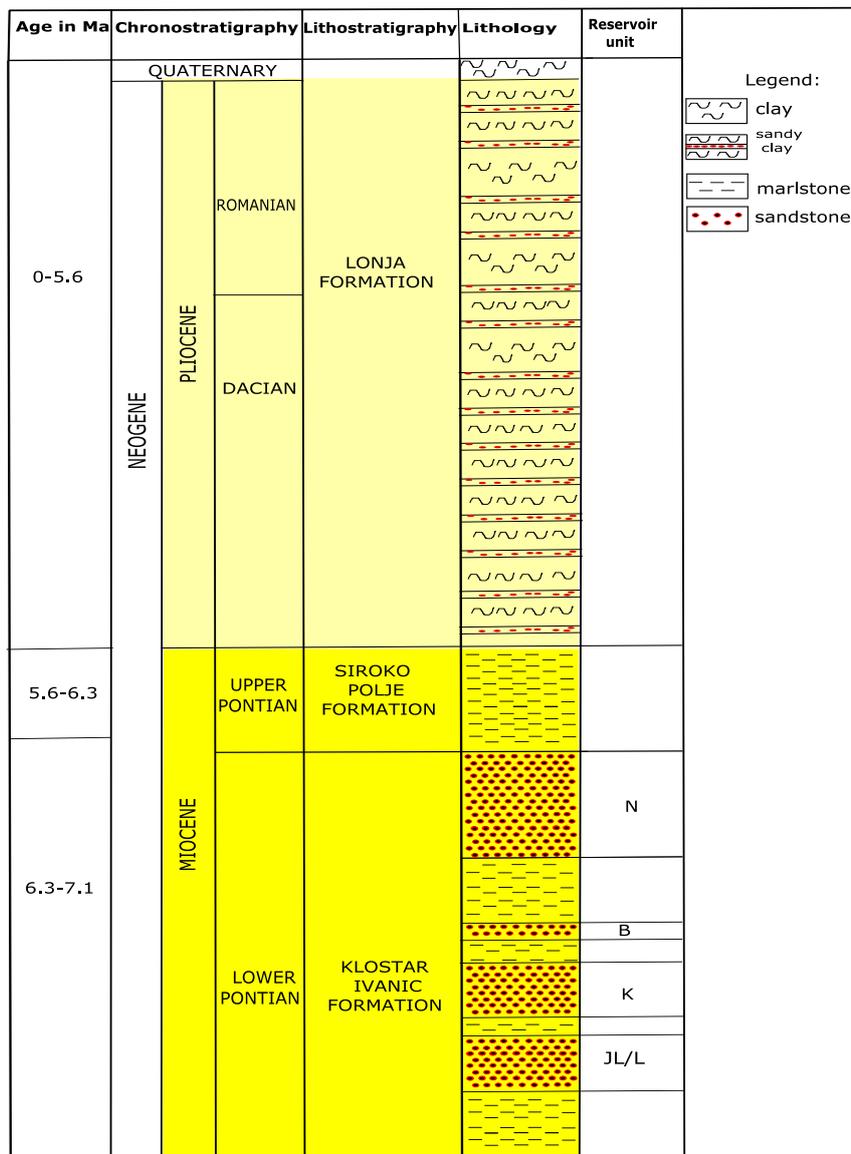


Figure 3. Geological column of the analysed field "B" [23]

All reservoirs are of turbidite origin and are therefore lithological constructed of several sandstone lithofacies. Most of the wells drilled only rocks up to and including the Klostár Ivanić formation, because hydrocarbon reservoirs have been proven within it, and only a few have drilled deeper reservoirs. Therefore, a typical geological column is limited only to the Lower Pontian reservoirs.

Development drilling activities in field "B" began in the 1960s. The analysed data on porosity were obtained during field development drilling phase, while data on the costs of disposal formation water were obtained from the production data.

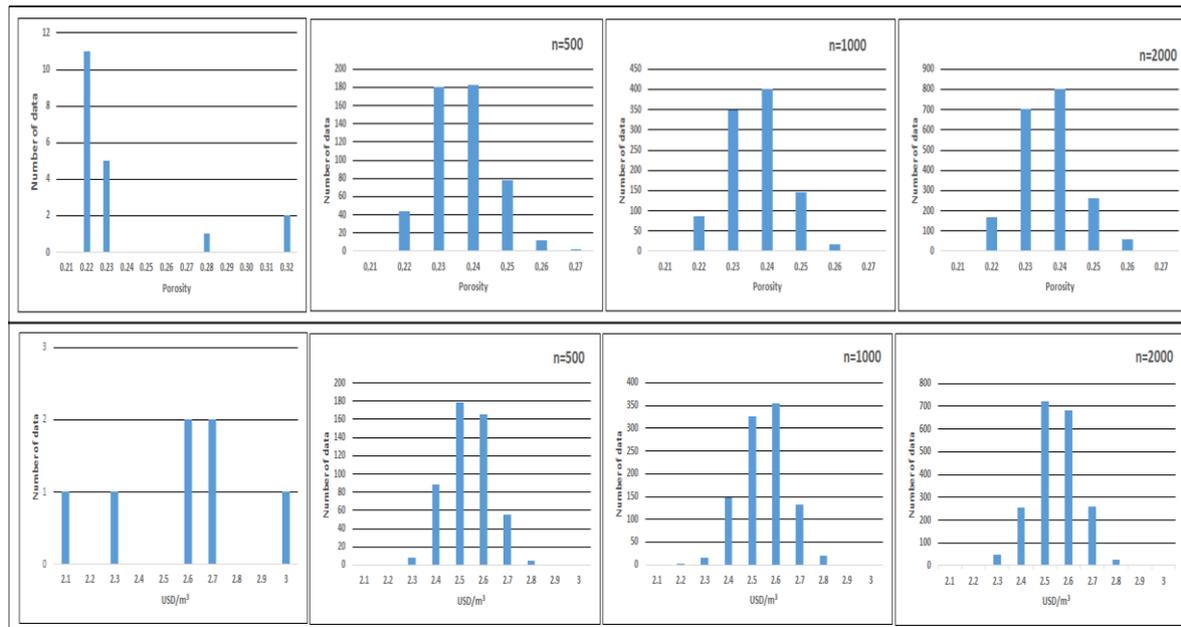
3. Results and discussion

Data on the costs of disposal formation water were processed and taken from papers [24-26], while the value of porosity of reservoir "K" taken from paper [27]. The basic statistics of the input data are shown in Table 1.

Table 1. Basic statistics of porosity input data and total costs of disposal formation water

Variable	n	Min	Max	Mean	Median
Porosity	19	0.217	0.315	0.232	0.217
Cost of disposal formation water (USD/m ³)	7	2.09	2.91	2.50	2.55

The analysed variables with respect to the size of the input data entry are classified into a small dataset. What makes this data suitable for applying the bootstrap method. For these variables, histograms were obtained by the bootstrap method in the case of 500, 1000 and 2000 number of resampling data sets are shown in Figure 4.

**Figure 4.** Histograms obtained by bootstrap method for porosity and total costs of disposal formation water

As can be seen from Figure 3 for all cases of data resampling, the obtained histograms describe the normal distribution curve. The normality of the obtained data confirmed by the correlation coefficient for normality obtained from the graphic normality test (Table 2). Table 2 also shows the estimated 95% confidence interval for the porosity and total costs of disposal formation water (CDFW) variables.

Table 2. Results of the bootstrap method for porosity and total costs of disposal formation water

Variable	n	Correlation coefficient (graphic normality test)	Confidence interval (95%)
Total costs of disposal formation water	500	0.9985	< 2.3120, 2.6668 >
	1000	0.9988	< 2.3160, 2.6947 >
	2000	0.9993	< 2.3027, 2.6744 >
Porosity	500	0.9847	< 0.2187, 0.2493 >
	1000	0.9877	< 0.2182, 0.2475 >
	2000	0.9835	< 0.2182, 0.2506 >

The correlation coefficient for the normality of the obtained sets of numbers is high (porosity over 0.99, CDFW over 0.98). The differences in the correlation coefficient between three of resampling data sets are negligible because it is a matter of changing the value only in the third decimal place.

Looking at this criterion, all three cases of resampling data are satisfactory in the case of both analysed variables. In terms of the confidence interval, the results obtained for the normality correlation coefficient are very similar. For the variables CDFW and porosity, value changes are detected in the second decimal place. It can be observed that in the obtained confidence interval for both variables the values of the lower and upper intervals do not include the minimum and maximum values from the initial data set. This is quite understandable because it is also a feature of the bootstrap method because in calculating the mean value of the sample of each individual new set of numbers, there will never be a result of the mean value of the minimum or maximum of the input data set. The correct choice of resampling size for the application of the bootstrap method with respect to the confidence interval was obtained as large a range in the interval as in the case of the variables analysed in this paper for porosity ($n = 2000$) and CDFW ($n = 1000$). Because the larger the range in the confidence interval, the closer the interval value of the input data set is. Therefore, when analysing the application of the bootstrap method, it is necessary to analyse the impact of resample size on the results and their interpretability. As can be seen from this analysis of variables, 3 different resampling sizes are enough to make the right decisions when analysing data.

4. Conclusions

The following conclusions of the paper are:

- The Bootstrap method is applicable as an additional tool in the analysis of a small input set of variables.

- To apply the bootstrap method, it is sufficient to perform an analysis of three different sizes of resampling dataset.

- In the case of equality of the correlation coefficient of bootstrap samples, the larger the range of confidence intervals of the obtained samples, the more acceptable the final result of the bootstrap analysis. In case of inequality, the sample with the highest correlation coefficient and the highest confidence interval selected, in case of discrepancy; the sample with the highest correlation coefficient is taken into account.

- The porosity of the "K" deposit is 0.2182 to 0.2506 at a resampling size of 2000, while the CDFW value for field "B" is 2.31 USD/m³ - 2.69 USD/m³ at a resampling size of 1000.

- Given the amount of data that can be collected during the analysis of deep geological data, and such bootstrap data with a 95% interval are good information in analysis and obtaining a clearer picture when mapping deep geological variables.

- Applying the bootstrap method provides a safer insight into the uncertainty of the reservoir volume and the possible costs of its production.

Author Contributions: Conceptualization, J.I., T.M. and D.P.; methodology, J.I. and T.M.; writing-original draft preparation, J.I., T. M. and D. P.; writing-review and editing, J.I., T.M. and D.P.; All authors have read and agreed to the published version of the manuscript.

Funding: This research (analyses) was partially supported with the project "Mathematical methods in geology V" (led by T. Malvić). Funds were given from the University of Zagreb, Faculty of Mining, Geology and Petroleum Engineering, for the 2020 year.

Acknowledgments: Some data from the technical documentation of INA Plc. was used and presented in some maps and tables. This was collected during doctoral research by J. Ivšinović.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Dwornicka, R.; Radek, N.; Pietraszek, J. The bootstrap method as a tool to improve the design of experiments. *System Safety: Human-Technical Facility-Environment* **2019**, *1*, 724-729. doi: <https://doi.org/10.2478/czoto-2019-0092>
2. Ague, J. J.; van Haren, J. L. M. Assessing Metasomatic Mass and Volume Changes Using the Bootstrap, with Application to Deep Crustal Hydrothermal Alteration of Marble. *Economic Geology* **1996**, *91*, 1169-1182. doi:10.2113/gsecongeo.91.7.1169

3. Cai, Y.; Davies, N. A Simple Bootstrap Method for Time Series. *Communications in Statistics - Simulation and Computation* **2012**, *41*, 621-631. DOI:10.1080/03610918.2011.598988
4. Čorkalo, Š. Comparison of value at risk approaches on a stock portfolio. *Croatian Operational Research Review (CRORR)* **2011**, *2*, 81-90.
5. Valášková, K.; Spuchlákova, E.; Adamko P. Non-parametric Bootstrap Method in Risk Management. *Procedia Economics and Finance* **2015**, *24*, 701-709. doi: [https://doi.org/10.1016/S2212-5671\(15\)00678-4](https://doi.org/10.1016/S2212-5671(15)00678-4)
6. Zheng, T. Measuring Agglomeration Using the Standardized Location Quotient with a Bootstrap Method. *Journal of Regional Analysis and Policy* **2013**, *43*, 186-197.
7. Elvarsson, B.; Taylor, L.; Trenkel, V.; Kupca, V.; Stefansson, G. A bootstrap method for estimating bias and variance in statistical fisheries modelling frameworks using highly disparate datasets. *African Journal of Marine Science* **2014**, *36*, 99–110. doi:10.2989/1814232x.2014.897253
8. Hirsch, R. M.; Archfield, S. A.; De Cicco, L. A. A bootstrap method for estimating uncertainty of water quality trends. *Environmental Modelling & Software* **2015**, *73*, 148–166. doi:10.1016/j.envsoft.2015.07.017
9. Zhang, A.; Shi, H.; Li, T.; Fu, X. Analysis of the Influence of Rainfall Spatial Uncertainty on Hydrological Simulations Using the Bootstrap Method. *Atmosphere* **2018**, *9*, 71. <https://doi.org/10.3390/atmos9020071>
10. Flowers-Cano, R.S.; Ortiz-Gómez, R.; León-Jiménez, J.E.; López Rivera, R.; Perera Cruz, L.A. Comparison of Bootstrap Confidence Intervals Using Monte Carlo Simulations. *Water* **2018**, *10*, 166. doi:10.3390/w10020166
11. Stoma, P.; Stoma, M.; Dudziak, A.; Caban, J. Bootstrap Analysis of the Production Processes Capability Assessment. *Appl. Sci.* **2019**, *9*, 5360. <https://doi.org/10.3390/app9245360>
12. Malvić, T.; Bastaić, B. Reducing variogram uncertainties using the 'jack-knifing' method, a case study of the Stari Gradac – Barcs-Nyugat field. *Bulletin of Hungarian Geological Society (Foltani Kozlony)* **2008**, *138*, 165-174.
13. Malvić, T.; Ivšinović, J.; Velić, J.; Rajić, R. Kriging with a Small Number of Data Points Supported by Jack-Knifing, a Case Study in the Sava Depression (Northern Croatia). *Geosciences* **2019**, *9*, 36.
14. Hesterberg, T.; Monaghan, S.; Moore, D. S.; Clipson, A.; Epstein, R. *Bootstrap methods and permutation tests*. W. H. Freeman and Company: New York, USA, 2003; 84 p.
15. da Silva, C. M.; Amaral, A.; Amaral, R. S.; Santos, J. A.; Vieira, J. W. Application of bootstrap method for evaluating discrepant levels of Radium-226 in forage palm (*Opuntia* spp). *Revista Brasileira de Biometria* **2007**, *25*, 109-114.
16. Ivšinović, J. Statistička obrada troškova odvajanja pridobivene slojne vode iz pješčenjačkih ležišta Savske depresije (Statistical analysis of the costs of separating the produced formation water from a sandstone reservoir of Sava Depression). In Proceedings of the Matematičke metode i nazivlje u geologiji 2016, Zagreb, Croatia, 29. October 2016. (in Croatian)
17. Zoubir, A. M.; Iskander, D. R. Bootstrap Methods and Applications: A Tutorial for the Signal Processing Practitioner. *IEEE - Signal Processing Magazine* **2007**, *24*, 10-19.
18. Chang, J.; Hall, P. Double-bootstrap methods that use a single double-bootstrap simulation. *Biometrika* **2015**, *102*, 203–214. doi:10.1093/biomet/asu060
19. Horowitz, J. L. Bootstrap Methods in Econometrics. *Annual Review of Economics* **2019**, *11*, 193-224. doi:10.1146/annurev-economics-080218-025651
20. Grunkemeier, G. L.; Wu, Y. Bootstrap resampling methods: something for nothing? *The Annals of Thoracic Surgery* **2004**, *77*, 1142–1144. doi:10.1016/j.athoracsur.2004.01.005
21. Carpenter, J.; Bithell, J. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Statistics in Medicine* **2000**, *19*, 1141-1164.
22. Ivšinović, J.; Malvić, T.; Velić, J.; Sremac, J. Geological Probability of Success (POS), case study in the Late Miocene structures of the western part of the Sava Depression, Croatia. *Arabian Journal of Geosciences* **2020**, *13* (714), 1-12 doi:10.1007/s12517-020-05640-z.
23. Malvić, T.; Ivšinović, J.; Velić, J.; Sremac, J.; Barudžija, U. Increasing Efficiency of Field Water Re-Injection during Water-Flooding in Mature Hydrocarbon Reservoirs: A Case Study from the Sava Depression, Northern Croatia. *Sustainability* **2020**, *12*, 786.
24. Ivšinović, J.; Dekanić, I. The basics of model for marginal testing of costs for disposal of extracted formation water. *Rudarsko-geološko-naftni zbornik* **2015**, *30*, 85-100 doi:10.17794/rgn.2015.2.7.
25. Ivšinović, J. The analysis of water injection systems in sandstone hydrocarbon reservoirs, case study from western part of the Sava Depression. *Rudarsko-geološko-naftni zbornik* **2017**, *32*, 17-24 doi:10.17794/rgn2017.1.3.

26. Ivšinović, J. The Cost Analysis of the Separation of Produced Formation Water from the Hydrocarbon Reservoir Using the Example of the Upper Miocene Sandstone Deposits of the Sava Depression. *Rudarsko-geološko-naftni zbornik* **2018**, 33, 35-43 doi:10.17794/rgn.2018.1.5.
27. Ivšinović, J.; Malvić, T. Application of the radial basis function interpolation method in selected reservoirs of the Croatian part of the Pannonian Basin System. *Mining of mineral deposits* **2020**, 14, 37-42 doi:10.33271/mining14.03.037.