RGNZ/MGPB

Rudarsko-geološko-naftni zbornik
(The Mining-Geology-Petroleum Engineering Bulletin)
DOI: 10.17794/rgn.2025.4.14

Original scientific paper

Comparison of the Polynomial Regression Zonal Estimation Comparison of the Inverse Distance Weighting and Ordinary Kriging Linear Interpolator: Case Study of Porosity Estimation in Neogene Sandstone Reservoirs in the Sava Depression, Croatia

Tomislav Malvić[1], Josip Ivšinović[2], Uroš Barudžija[1*], Ivana Brajnović[1], Maria Rudec[1]

[1] *University of Zagreb, Faculty of Mining, Geology and Petroleum Engineering, Pierottijeva 6, HR10000 Zagreb, Croatia*
[2] *Elementary school Pantovčak, HR-10000 Zagreb, Croatia*

Abstract

Two reservoir porosity datasets are mapped using the Polynomial Regression (PR; zonal estimation), Inverse Distance Weighting (IDW) and Ordinary Kriging (OK; both are interpolations). Data were collected from sandstone reservoirs "K" (19 samples) and "L" (25 samples), of the Lower Pontian age, and located in the Sava Depression, Croatia. Maps were compared via cross-validation (mean square error) and visual inspection of key features. Datasets are considered as small ("K" reservoir) and large ("L" reservoir). In the "K" reservoir, the MSE values are 0.000119 for IDW vs. 0.05401 for PR. In the "L" reservoir the MSE are 0.000676 (OK) vs. 0.040117 (PR). Zonal estimation obviously did not prove as primary mapping in the sets with about 20 datasets. A linear interpolator like the IDW or OK are much better choices, especially if a spatial model can be reliably modelled, like in this case for porosity characterised with normal distribution, which favoured the OK. However, zonal estimation can be a useful addon in interpretation, especially in zones where interpolation is considering less reliable, like in transitional areas.

Keywords:
sandstone reservoirs, porosity, zonal estimation, interpolation, Sava Depression, Croatia

1. Introduction

Fluid reservoir modelling is closely based on understanding spatial variations of numerous geological variables like porosity, permeability, saturation, mineral and lithological content. A better understanding of these variables improves recovery efficiency and longevity. Such data are, often, very scarce and irregularly distributed which makes it difficult to interpret them. Also, most of them are indirectly calculated which complicates their reliability. This creates a dilemma in the selection of the interpolation methods, which are practically one of the most important interpretative tools.

Here two datasets are analysed, one with 19 (reservoir "L"), and another with 25 (reservoir "K") datasets, which make them small and larger (**Malvić et al., 2019**) datasets, respectively. Analysed reservoirs are chosen in the Croatian part of the Pannonian Basin System (CPBS), i.e. in the western part of the Sava Depression (see **Figure 1**), where numerous hydrocarbon fields had been discovered, and many of them are still in production. Both are of Lower Pontian age and mostly consist of medium-grained sandstones, where porosity values are calculated (from logs and cores) as reservoir's averages in real wells and represent the most important available variable for spatial mapping. Regular geological mapping can be done with any number of data points, however dozens of them make maps reliable. In subsurface fluid reservoir's mapping, such plentiful of data is rarely available, and 10 or so are datasets that need to be handled, more or less, successful for development and recovery prediction. In such cases, methods like Kriging are less useful than mathematically simpler like Inverse Distance Weighting (IDW). Also, interpolation often, in very uncertain datasets (because of statistics or data sources), can be replaced with an even more basic approach on the zonal estimation. Here both such approaches are used as tests for the "average" reservoir in the selected area, looking for the answer regarding the most appropriate approach when choosing zonal estimation (Polynomial Regression; PR), simpler interpolation (IDW) or more advanced interpolation (Ordinary Kriging; OK).

The mentioned methods are widely used in geological researching, even in the CPBS. However, they are not limited only to geology, just the opposite. For example, application of the PR can be found in geography (**Najafzadeh & Kargar, 2019**), meteorology (**Weslati et al., 2022**), thermodynamics (**Bal, 2022**), mechanical engineering (**Trebuňa et al., 2016; Wu & Zhang, 2021**), astronomy and astrophysics (**Jiménez-López, 2021**), civil

* Corresponding author: Uroš Barudžija
e-mail address: uros.barudzija@rgn.unizg.hr
Received: 1 April 2025. Accepted: 7 May 2025.
Available online: 27 August 2025

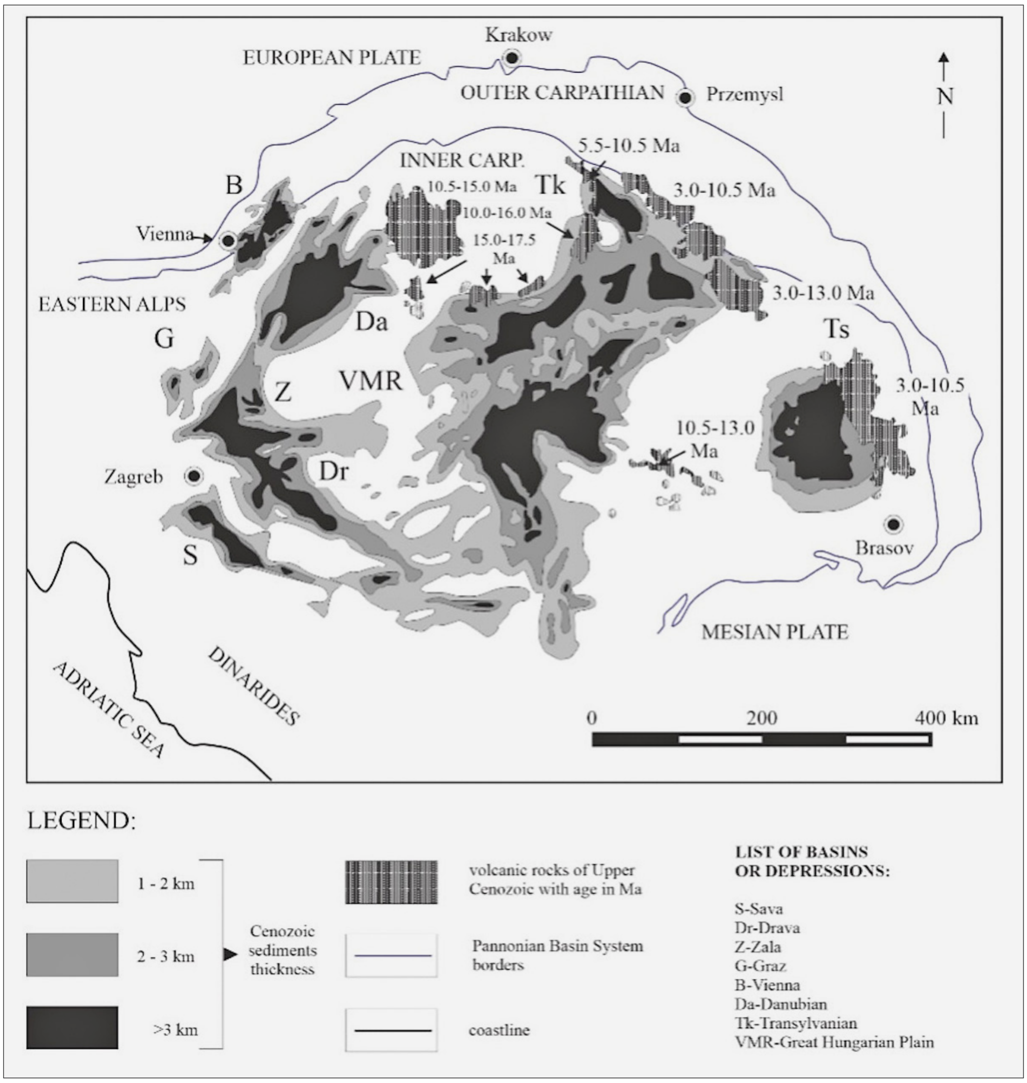


Figure 1. Regional schematic thickness map with largest volcanic outcrop areas and subdivision on depressions and basins of the Pannonian Basin System as geological unit of the highest order (**Malvić et al., 2020**).

engineering (**Golnaraghi et al., 2020**), petroleum industry (**Said, 2023; Al-Mudhafar et al., 2017; Kalla & White, 2007; Yao et al., 2023; Aminian et al, 1991; Wang et al, 2024**) and others.

Furthermore, the Inverse Distance Weighting method has found applications in geography (**Moeletsi et al, 2016**), meteorology (**Chen & Liu, 2012**), geology (**Bokati et al, 2021**), and electrical engineering (**Barlak & Ozkazanc, 2011**). The success of Ordinary Kriging is also not limited to a specific field, as it has been applied in geography (**Rohma, 2022**), geostatistics (**Di et al., 2023**), and environmental science (**Qiao et al, 2018**).

2. Geological settings and location

The PBS, macro unit of the largest order, encompasses areas in eight countries, generally in Central Europe. That system is surrounded by the Carpathians, Alps and Dinarides as the regional mountain chains (orogens). The PBS started to be created during the Ottnangian due to convergence between the Euroasian and African Plates, i.e. subduction of the Apulian Plate below the Dinarides. So, the PBS is a typical back-arc basin system, covered with numerous marine and lacustric environments, among which the largest was (part of) the Paratethys (**Malvić & Velić, 2008**).

The CPBS (see **Figure 2**), at the south-west margin, includes four geological macro units of the 2nd order, namely Mura, Drava, Sava and Slavonia-Srijem Depressions. About 40 hydrocarbon fields were discovered with numerous reservoirs, mostly in the Upper Miocene sandstones.

In the CPBS, two main lithological macro sequences dominate. Their differentiation has been established based on lithology and geological background (**Velić et al., 2012**). The 1st, younger, includes mostly clastic sediments, with minor biochemical (like *Lithotamnium* limestones) or magmatic (like basalts) appearances, of

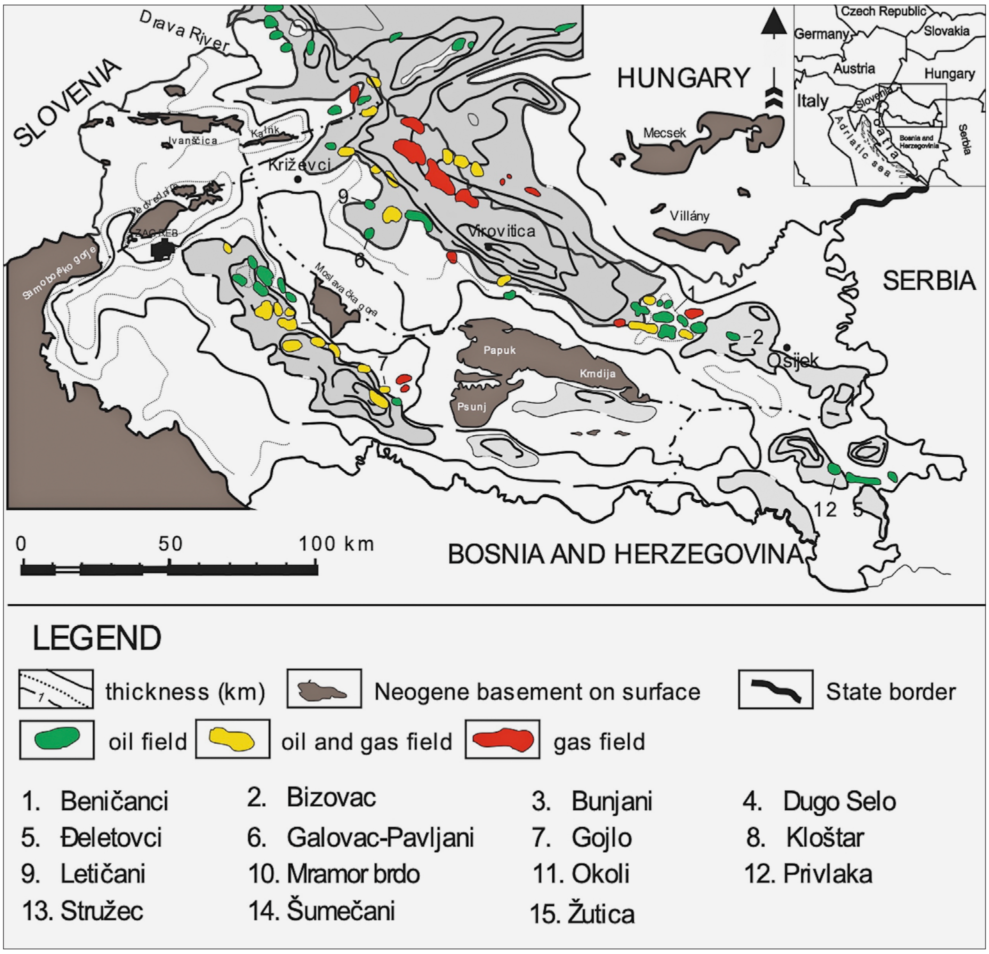
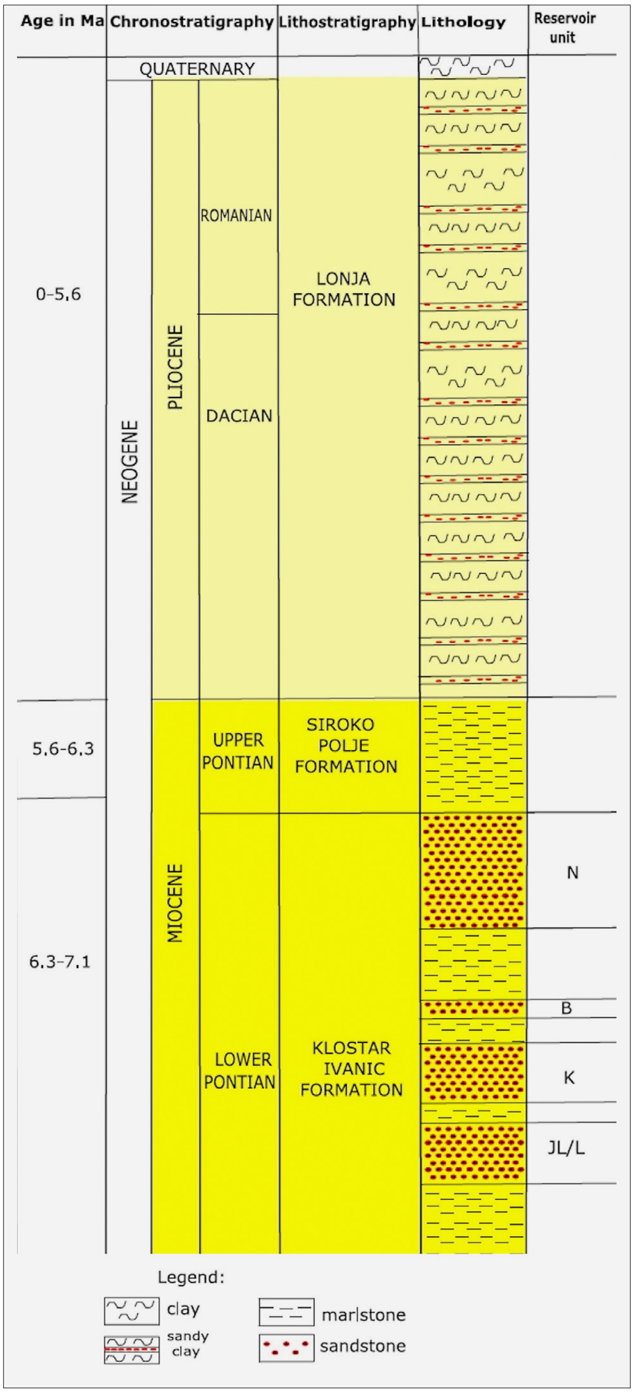


Figure 2. Macro units in the CPBS (modified from **Velić et al., 2012**).

the Neogene and Quaternary periods. The 2nd, older, in the bottom is significantly older, from Mesozoic or Palaeozoic eras, encompassing magmatic or sedimentary rocks, often metamorphosised. The 1st sequence in the Sava Depression is divided into six lithostratigraphic formations (**Velić, 2007**) where 3 of them, important for this research, are presented in **Figure 3**.

During the entire Neogene-Quaternary period there are two transtensional (Badenian and Pannonian-Early Pontian) and two transpressional (Sarmatian and Late Pontian-Quaternary) regional tectonic phases described for the CPBS (**Malvić & Velić, 2011**). Neogene and Quaternary cyclic sediments were divided in megacycles with lithostratigraphic formations and members. Late Miocene (Pannonian and Pontian) cycle includes sedimentary association of the Sava Group (the Ivanić-Grad, Kloštar-Ivanić and Široko Polje Formations) in the Sava and western Drava Depressions. Deposition lasted app. 5.9 million years, and it is represented by the sequences of grey coloured sandstones, siltites and marls. The maximum total thicknesses are deposited in the central part of depressions, pinching out toward the margins. These sandstones, siltites and marls are deposited in deep lake environment basin sedimentation interrupted with turbidites, with the main sources of the material from the Eastern Alps. The main depositional mechanisms during Late Pannonian and Early Pontian were deep water turbidites and sequences of hemipelagic marls were described as well, both in the marginal and the central parts of the basins. Several facies associations have been described: turbidite channel fill facies association - thick-bedded sandstones and thin-bedded sandstones; turbidite overbank-levee facies association - laminated sandstones, siltstones and marls passing into sandstones; distal turbidite facies association - alternating thin sandstones with siltstones and marls; massive marls facies association - marls with rare intercalations of thin siltstone or sandstone laminae.

The Sava Depression has been filled with sediments since the Early Neogene. However, the analysed reservoir rocks ("K" and "L" in Figure 3) are of the Late Neogene (Lower Pontian) age and they are a part of the Kloštar-Ivanić Formation. Arenitic sandstones prevail at the base of the formation, and more fine-grained sandstones intercalated with marl appear more frequently towards the top, as well as in the overlying the Široko Polje Formation. Sandstone units (20-150 m thick) are the main reservoir rocks, and intercalated grey to brown marls (from 30-150 m thick) present the main isolator rocks. These Lower Pontian ("Abichi") beds extend



◄ **Figure 3.** Typical lithological units in the 3 younger lithostratigraphic formations of the Sava Depression (**Ivšinović & Malvić, 2020**).

across the Sava Depression, are comprised of sandstones and marls. Sandstones were deposited from turbidite currents in the largest thicknesses in the central part of the depression, while marls were deposited as turbidites and settled from suspension. Generally, several unique lithofacies are developed which resulted from the Bouma sequence, namely (1) interlaminated marlstone and sandy siltite/shale, (2) laminated marlstone interchanged with clayey-calcitic laminas and laminas with mica flakes and fine quartz grains, (3) fine-grained silty lithic greywacke sandstones to mudstones with quartz grains lithic fragments, bounded in matrix. (4) silty marlstones with kerogen; (5) lithic arenite sandstones with quartz grains and lithic fragments; (6) fine-grained fossiliferous lithic arenite sandstones; (7) coarse-grained lithic arenite sandstone to petromictic breccia/conglomerate; (8) clast-supported petromictic conglomerate with sandy matrix. Some of them can be recognised locally, depending on the age and palaeoenvironment. The analysed reservoirs belong to lithofacies 5 and 6.

The analysed reservoirs "L" and "K" are located at the margin of the western Sava Depression (see **Figure 4**). They largely participate in the hydrocarbon potential of analysed zone. That part includes about 20 hydrocarbon fields still in production, but several others are depleted. The largest ones are the Stružec Field (16x106 m^3 Original Oil In Place, abbr. OOIP) and the Ivanić Field (7x106 m^3 OOIP) (**Malvić & Velić, 2011**). The entire western part covers area of about 8.000 km^2, where surface projection of the field's areas takes about 930 km^2 (**Ivšinović, 2019**). Generally, a little deeper and more to the west, sediments described as the source rocks are located, with fine-grained sediments mainly enriched in kerogen. They correspond to previously described lithofacies 1, 2 and sporadically 3. All are deposited in the deeper lacustrine environment from basin sedimentation or at the distal parts of turbidites.

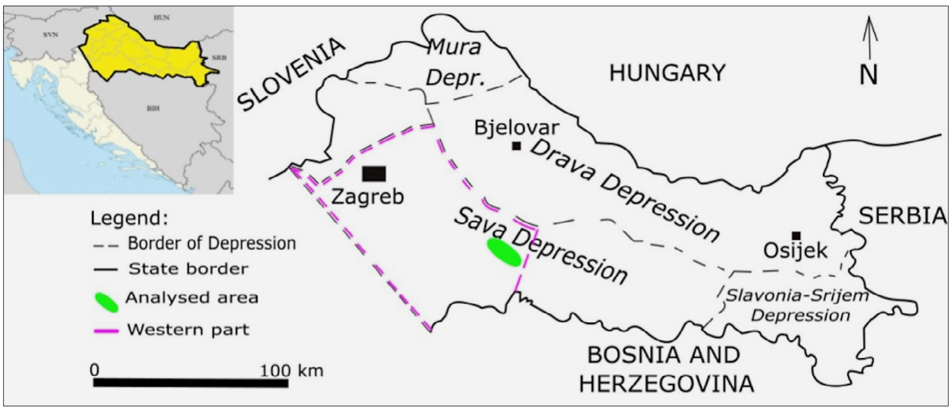


Figure 4. Regional map of the northern Croatia (upper right) enlarged with position of the western part of the Sava Depression. Two analysed fields are located inside marked green (modified from **Ivšinović et al., 2020**).

3. Mathematical basics of applied interpolation methods

Here the analytical methods applied in this research are described. The PR is used as a zonal estimator, and the IDW and OK as a linear interpolator. The Mean Square Error is calculated as an estimator of mapping numerical error.

The PR, as an estimation (here also zonal) method, was in the focus of this paper as a tool that could be a useful companion in the application of other linear interpolators like the IDW or OK. It could be useful when interpolation methods do not provide reliable solutions to the problem of constructing representative spatial distributions, for example in cases like low number of data points, clustering or numerous outliers. Moreover, polygonal methods are considered in geostatistics (beside cell-declustering, e.g. **Journel, 1983; Deutsch, 1989**) as one of the declustering methods, where polygon of influence (known as Thiessen or Voronoi polygon, e.g. **Chow, 1964; Boots, 1987**) includes all data points that are closer to the sample compared to any other measurements. As a result, spatially separated data points will have larger polygons than clustered (grouped) ones. Such polygonal declustering, characteristics for all zonal methods have been tested on two datasets presented in this research. The very unique feature of the PR in the software used in this analysis (Golden software Surfer) is a parameter called Maximum Total Order (MTO). This determines the maximum powers for the X and Y components in the polynomial equation, and all combinations for X and Y are included in the calculation until the sum of powers does not exceed MTO.

3.1. The Polynomial Regression

The polynomial regression is a kind of improvement of simple linear regression, in the case that the relationship between the data is non-linear. In such a case, the polynomial regression can better fit data, adding some polynomial terms into a linear regression equation and a modelling relationship between the dependent (Y) and independent (X) variables with the nth degree polynomial function. The model is called a quadratic if such a polynomial has a degree of 2, a cubic for a degree of 3. The degree of function order can be set on any value (sometimes this order is considered as a hyperparameter as well), but selection must be done carefully because each polynomial function can be easily underfitted or overfitted (even if the least square method is used for minimizing the error). There is the "rule of the thumb" that in the subsurface mapping the fitting function could not be of the 5th or higher order. So, the right polynomial degree would need to lead to reasonable mapping structures (where the main ones are confirmed with other mapping method), but also with smaller cross-validation results (when several models are run).

The Polynomial Regression (PR) method describes the relation between independent variable x and dependent y, using a polynomial of the nth order suitable for the approximation of their relation (**Wang et al., 2024**). The PR can be calculated using **Equation 1**, (e.g. **Yao et al., 2023**), and is used for non-linear relations, which can help in looking for a connection for more complex datasets.

$$y_i = \beta_0 + \beta_1 \cdot x_i^1 + \beta_2 \cdot x_i^2 + \ldots + \beta_m \cdot x_i^m + \varepsilon_i \qquad (1)$$

where:

y_i – ith dependent variable;
x_i – ith independent variable;
m – the mth order of regression;
ε_i – unobserved random error for ith data with mean zero conditioned on a variable x;
β_{0-m} – values adjusted during the calculation.

Equation 1 represents a general polynomial regression term, which always includes model parameters and hyperparameter at the end. Here the "error" is marked as the correction parameter used for fitting the model during iterations. Such a parameter is used to reach the optimal model parameters. In fact, any parameter that models the function shape is called hyperparameter. It could be the ratio between training and validation datasets, etc. Here, such a parameter is called the "error", however, in the applied algorithms, the more important hyperparameter is the order of the fitting function, previous mentioned as the MTO. Here it is tested with several values between 1 to 10, optimising the regression model through an iterative model. The goal was to find the best value of hyperparameter, leading to the optimal map.

3.2. The Inverse Distance Weighting

The Inverse Distance Weighting (IDW) is a relatively simple linear interpolation where an unknown value is calculated from known data inside the search radius and the assumption that closer data is stronger participates in the interpolation of the unknown value. The interpolation of the unknown value is defined by **Equation 2** (e.g. **Setianto & Triandini, 2013**):

$$Z_{IU} = \frac{\sum_{i=1}^{n} \frac{z_i}{d_i^p}}{\sum_{i=1}^{n} \frac{1}{d_i^p}} \qquad (2)$$

where:

ZIU – interpolated value;
d_i – distance to the ith location;
z_i – known value at the ith location;
p – distance exponent.

It is obvious that interpolation is highly dependent on the distance exponent (p) and its selection is led by obtaining the logical visual shapes on the map, as well as minimal numerical error. In mapping, it is very often set on value 2 (**Ly et al., 2011**), especially for subsurface geological mapping in Northern Croatia (**Malvić et al.,**

2019; Ivšinović & Malvić, 2020; Ivšinović & Malvić, 2022; Andrić et al, 2021).

3.3. The Ordinary Kriging

A general linear equation is given in **Equation 3** (e.g. **Novak, 2015**), where the higher value of the weighting coefficient points on a known value closer to the location where the value is estimated.

$$z_k = \sum_{i=1}^{n} \lambda_i \cdot z_i \tag{3}$$

where:

zk – value estimated at location „k“;

λi – weighting coefficient at location „i“;

zi – measured value at location „i“.

The Ordinary Kriging (OK) equation is an upgraded Simple Kriging (SK) with the addition of the Lagrange factor (μ), aiming to minimizing the Kriging variance. Surely, both techniques minimize variance, but the OK is more successful if it is applied properly. Also, all other Kriging techniques (like Universal, Disjunctive, Indicator...) have some additional “factors” with the same purpose, designed for the specific analyses. The SK is the only algorithm where the sum of the weightings is not standardised at 1. Also, in the SK the mean is known, not global, and in the OK, the local one is used. Eventually, in all Kriging techniques (except for the Indicator) the normal distribution of input dataset is highly recommended for more reliable results. It is closely connected with the requirements of the stationarity of the 2nd order where it is implied that the expectation is independent from the number and location of data, and covariance is dependent only on distances among the data (variogram). Only the Indicator Kriging assumed the stationarity of the 3rd order that implied independence of expectation and variogram existence (intrinsic hypothesis). The Lagrange function is represented with **Equation 4**:

$$L(x,\mu) = f(x) + \mu \cdot h(x) \tag{4}$$

where:

L (x, μ) – Lagrange function;

f(x), h(x) – functions for the “x” values;

μ – Lagrange factor.

The Ordinary Kriging matrix equation can be presented with **Equation 5** (e.g. **Mesić Kiš, 2016**):

$$\begin{bmatrix} \gamma(Z_1-Z_1) & \gamma(Z_1-Z_2) & \dots & \gamma(Z_1-Z_n) & 1 \\ \gamma(Z_2-Z_1) & \gamma(Z_2-Z_2) & \dots & \gamma(Z_1-Z_n) & 1 \\ & & & & 1 \\ \gamma(Z_n-Z_1) & \gamma(Z_n-Z_2) & \dots & \gamma(Z_n-Z_n) & 1 \\ 1 & 1 & \dots & 0 & 0 \end{bmatrix} \times \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \\ \lambda_n \\ \mu \end{bmatrix} = \begin{bmatrix} \gamma(Z_1-Z) \\ \gamma(Z_2-Z) \\ \\ \gamma(Z_n-Z) \\ 1 \end{bmatrix} \tag{5}$$

where:

Γ – variogram values;

Z1....Zn – measured values at locations “1...n”;

Z – locations where new value is estimated;

μ – Lagrange factor.

The mentioned property of the Ordinary Kriging is that the technique works with local mean (not global), which can be a pretty good feature for mapping geological subsurface data, at least in the CPBS (**Mesić Kiš & Malvić, 2016**). It is a purely deterministic method, as well as Inverse Distance and Polynomial Regression, for example. All are “numerical and subjective”, and the stationarity (on some order) can be criteria for choosing the right Kriging technique if a researcher knows how to observe stationarity.

3.4. The cross-validation

The cross-validation (CV) is a set of techniques applied for estimation error calculation. Here the mean square error (MSE) technique is selected as one of the most applied in comparison to several interpolations for the same dataset. As a measure of residuals variance, it usually means that a lower MSE would lead to a conclusion of better interpolation. There are also some other measures, like mean absolute error (MAE, the average of the absolute difference between measured and estimated values, i.e. residuals average), root mean square error (RMSE, keeping the same unit as input dataset, allowing easier interpretation), etc.

The MSE is an iterative procedure where one, randomly selected, measured value is ignored and at the same location its value is again estimated (with selected interpolation method) from the rest of the measurements. The difference between those two values is the error in that point. The procedure is repeated for all other measured values and eventually the MSE is calculated using **Equation 6** (e.g. **Hodson, 2022**):

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (Y\text{measured} - Y\text{estimated})^2 \tag{6}$$

where:

MSE – mean square error (of the estimation);

n – number of data;

Ymeasured,i – measured value for location “i”;

Yestimated,i – estimated value for location “i”.

However, the interpretation is not so straightforward, because the maps also can include different isolines shapes, where some forms (structures) could be almost impossible and could eliminate the interpolation method although the accompanied MSE is lower than in other methods.

4. Discussion and results overview

This research has been focused on zonal estimation as a useful addon to classical interpolation methods in the hydrocarbon sandstone reservoirs. The goal has been set up for the CPBS, where numerous interpolations of the mentioned and similar reservoirs had been published. The PR is selected as zonal estimation. The IDW and OK are the used interpolations. The datasets are divided

Table 1. Porosity data for reservoir „K"

Name	Surface X	Surface Y	Porosity
J-101	6421096.00	5028877.00	0.217
J-120	6420658.00	5029068.00	0.272
J-161	6420957.00	5028870.00	0.217
J-162	6421034.00	5028593.00	0.217
J-167	6420529.00	5028674.00	0.217
J-168	6420699.00	5028475.00	0.315
J-169	6420349.00	5028825.00	0.217
J-170	6420349.00	5028926.00	0.223
J-174	6421298.00	5028863.00	0.217
J-175	6420475.00	5029136.00	0.223
J-158	6420303.00	5028910.00	0.223
J-171	6420576.00	5028970.00	0.223
J-172	6420928.00	5029147.00	0.223
J-102	6421208.00	5028926.00	0.217
J-148	6421126.00	5028437.00	0.217
J-149	6420959.00	5028501.00	0.217
J-166	6420771.00	5028650.00	0.217
J-25	6420546.00	5028460.00	0.315
J-173	6420539.00	5028382.00	0.217

into small (19 data, **Table 1**) and large (25 data, **Table 2**) although both are of similar size.

Both sets are limited in their representation, but also solely publicly available for the analysed zones. The given values are mostly derived indirectly from e-logs, and this is why they are the same in numerous wells. They are very similar because they are logged in the same lithofacies. The values represent mean porosity along the entire reservoir in the corresponding well, which is often used in numerous mappings. However, some practitioners used porosities and pondered with other variables, for example, thickness. Such an option is not correct to apply if the researcher does not know (and describe) the depositional model, i.e. how clastics volume and sizes depend on (well) position in such an environment and subsequent compaction. As we did not have such detailed information, the pure mean porosity has been mapped.

Listed data had been interpolated or estimated using IDW, OK and PR. **Table 3** presents the cross-validation values obtained by changing the value of the maximum total order (MTO). The selected MTO (as one hyperparameter related on the order of a fitting function) values are 1, 2, 3, 4, 5, and 10. They determine the maximum sum of the exponents of the independent (X) and (dependent) Y variables, i.e. the degree of polynomial regression, ensuring that the sum of the exponents does not exceed the MTO value (**Draper & Smith, 1981**). For example, for "n" data, the maximum degree of polynomial function and MTO is "n-1". Like it was mentioned for the order of the polynomial function, choosing

Table 2. Porosity data for reservoir „L"

Name	Surface X	Surface Y	Porosity
L-111a	6417747.87	5027750.49	0.239
L-131a	6416846.88	5028084.13	0.156
L-136a	6416153.34	5028514.94	0.145
L-140	6415085.08	5028332.44	0.192
L-142	6415018.82	5028518.52	0.186
L-153	6416755.00	5028207.72	0.239
L-155	6416966.63	5028205.04	0.156
L-156	6415912.39	5028017.76	0.206
L-160	6416409.59	5028202.77	0.197
L-161	6416945.81	5028414.75	0.156
L-27	6416655.05	5028085.51	0.197
L-32	6417390.44	5027719.99	0.239
L-33a	6415763.30	5028687.46	0.214
L-33b	6415763.30	5028687.46	0.214
L-37	6415833.62	5028477.16	0.214
L-4a	6415435.16	5028753.52	0.214
L-5	6417199.92	5027939.22	0.239
L-57	6415945.52	5028103.82	0.206
L-62	6416090.56	5028354.65	0.206
L-65a	6415235.15	5028589.8	0.214
L-66	6415579.42	5028511.51	0.214
L-68	6415314.5	5028205.63	0.214
L-73	6414912.05	5028679.32	0.192
L-79	6414821.26	5028401.83	0.195
L-87a	6416346.64	5028297.46	0.197

an excessively high MTO value may lead to over-fitting, which causes situations where the data or predictions have significant errors or variations, which makes them unreliable (**Wan, 2019**). In contrast, a low value of MTO may result in under-fitting of the model, leading to failure in representing the relationship between the variables (**Araújo, 2018**). For all cases, the MSE is calculated and maps are generated. Regarding the PR, there were several options to estimate data changing the option named as "Max Total Order (MTO)". In this analysis, such an option is varied with values 1, 2, 3, 4, 5 and 10, looking at the lowest MSE (see **Table 3**).

When MTO reached 2, the MSE stayed constant (and overall, the variations are too small). So, this parameter can be considered of low importance, when the threshold of 2 is reached, for the used datasets regarding their abundance and locations. The next step was calculation of the same error using the IDW and OK interpolations in the same reservoirs (see **Table 4**).

In both reservoirs, the MSE values are considerably lower for interpolations than for the estimator (see **Tables 3** and **4**). In the reservoir with more data, those values are lower, which indicated statistically representative datasets, i.e. the fact that an increase in values did not include new "outlier" values and confirmed that both

Table 3. The MSE values for the PR porosity estimation with variable MTO values, for the reservoirs „K“ and „L“

POROSITY OF RESERVOIR "K"			POROSITY OF RESERVOIR "L"		
DATA	MTO	MSE	Data	MTO	MSE
19	1	0.054007	25	1	0.041100
19	2	0.054010	25	2	0.040117
19	3	0.054010	25	3	0.040117
19	4	0.054010	25	4	0.040117
19	5	0.054010	25	5	0.040117
19	10	0.054010	25	10	0.040117

Table 4. The MSE values for the IDW and OK compared with PR, for the reservoirs "K" and "L"

MSE				
Reservoirs	Data	IDW	OK	PR
"K"	19	0.000119	/	0.054010
"L"	25	/	0.000676	0.040117

datasets are representative. Also, the lowest MSE for OK proved that 25 data points were enough for the calculation of a reliable spatial model (variogram) based (probably, because it is not tested) on the normal distribution of porosity (as theoretical assumption for sandstone reservoirs).

The maps made for both datasets using the PR zonal estimator are shown on **Figure 5 left** (reservoir "K") and **Figure 5 right** (reservoir "L"). The main difference is, as in the MSE, in the maps associated with MTO=1 and the maps with MTO>1. As zonal, the maps show sharp borders between different porosity zones and all follow the NW-SE/WNW-ESE strike, which corresponds with the strike of a sandstone depositional environment. For reservoir "K", the zone values (colours) are almost the same for the MTO=2, 3 and 10, which follows the stabilisation of the MSE on the MTO=2 and higher. For reservoir "L", almost the same statement is valid, also with additional features reflected in the "parabolic" borders among zones. Also, the PR is sensitive on higher differ-

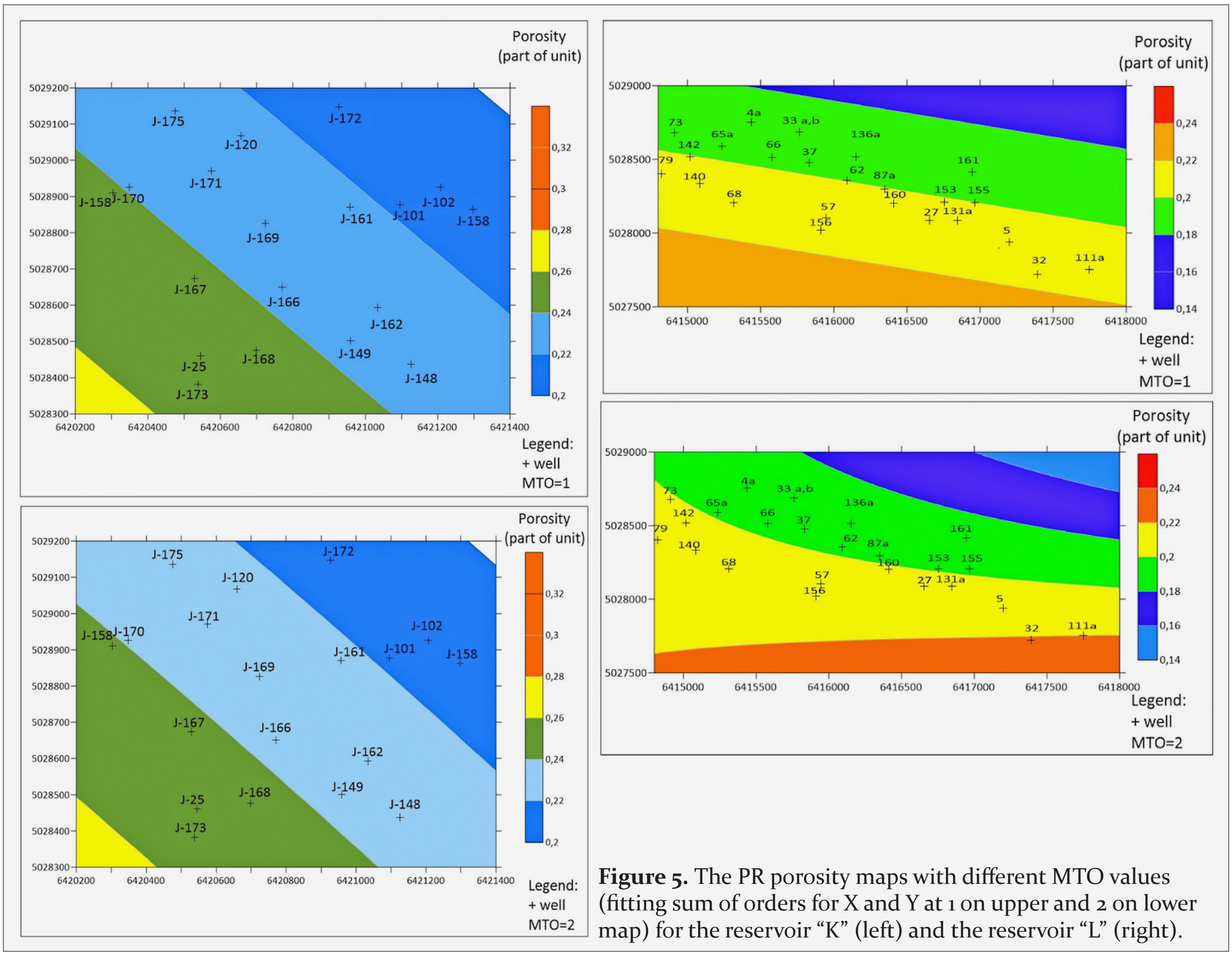


Figure 5. The PR porosity maps with different MTO values (fitting sum of orders for X and Y at 1 on upper and 2 on lower map) for the reservoir "K" (left) and the reservoir "L" (right).

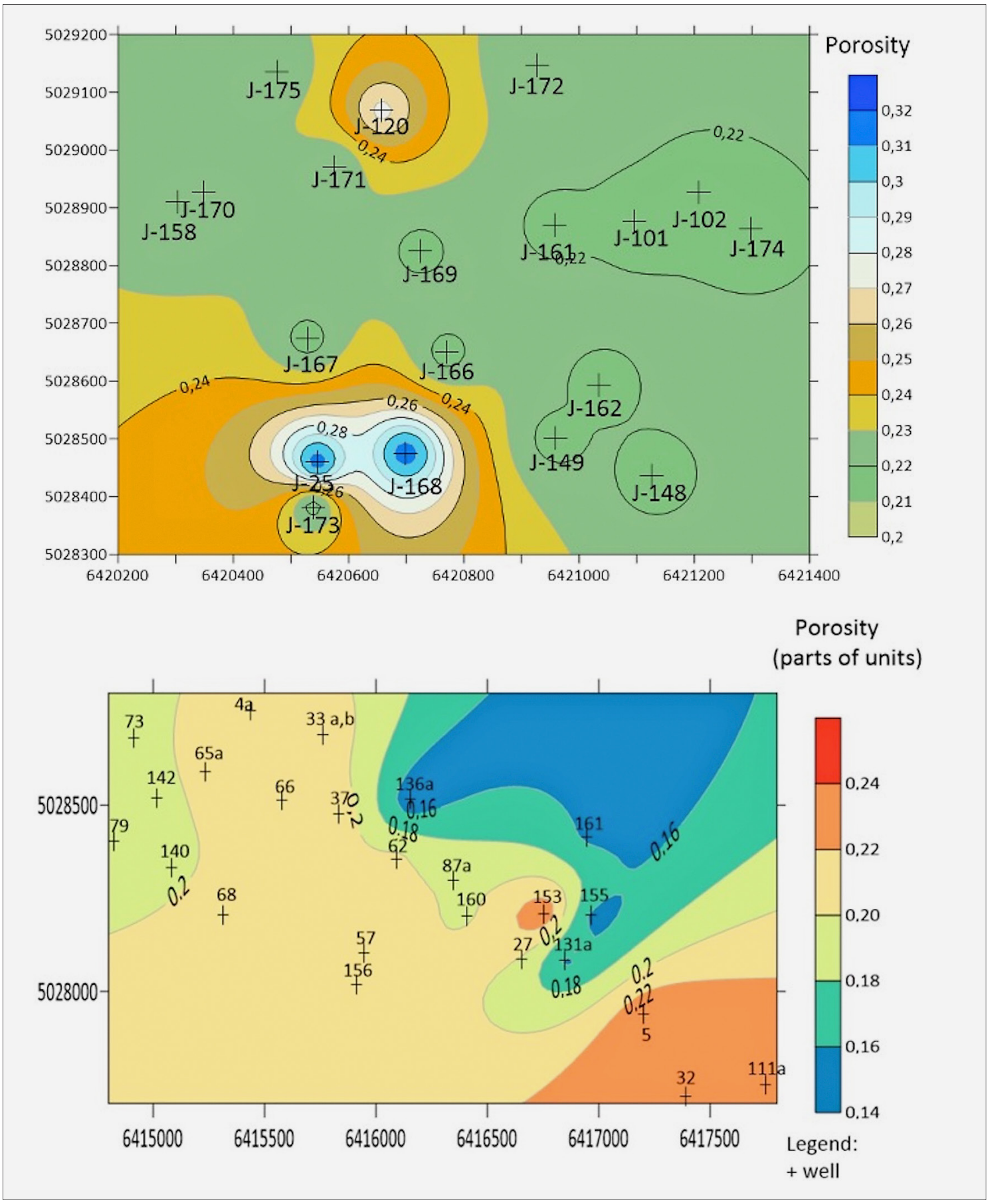


Figure 6. The IDW porosity map for reservoir "K" (top) and the OK porosity map for reservoir "L" (bottom)

ences between close data, which is visible in well pairs J-168 and J-25, i.e. L-27 and L-131a. In any case, when the data are classified mostly in several groups of the same values (see **Tables 1** and **2**) the zonal estimator will group them clear and simple.

The interpolated maps are given on **Figure 6 top** (reservoir "K", the IDW) and **Figure 6 bottom** (reservoir "L", the OK). As expected, the interpolators are based on transitional zones between isoporosity lines on the selected equidistance. So, the spatial interpretation and prediction is much easier than in the zonal representation, however, it can be tricky when numerous points are of the same values, like in the presented reservoirs (see **Tables 1** and **2**). The main difference is the existence of the maximums and minimums, which on the zonal maps were amalgamated into a single zone. Back to MSE, it would be a more appropriate spatial solution than zones.

Application of the OK asked for the definition of a spatial model, which was not an easy task due to the relatively low amount of data. This is why the omnidirectional variogram model had been used (for directional there would not be enough data per sectors of the

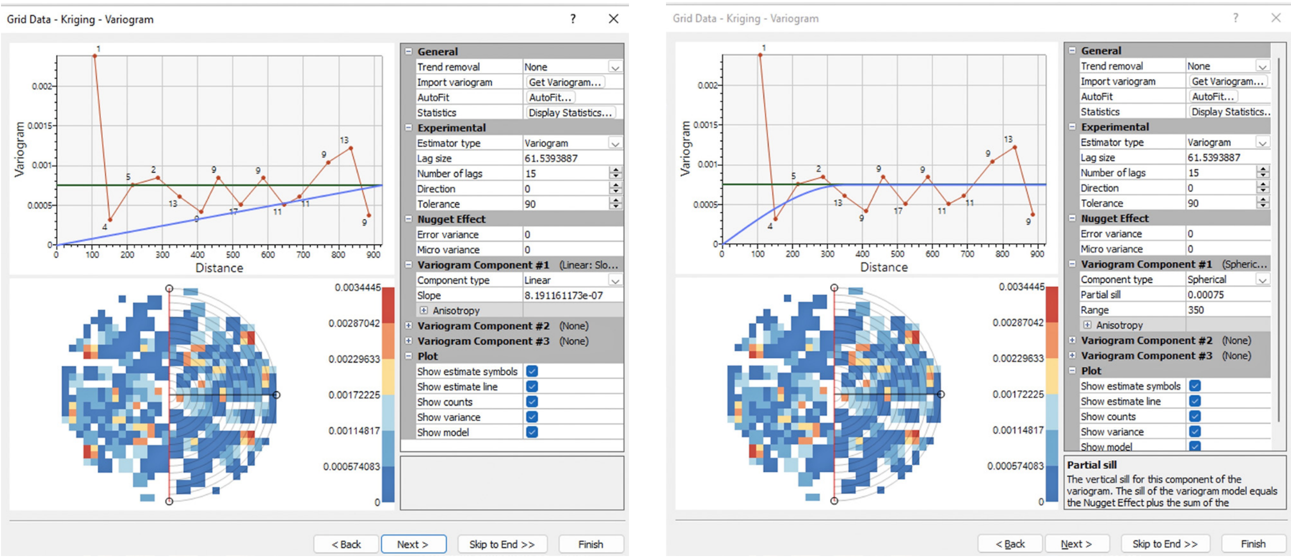


Figure 7. Examples of some tested theoretical variogram models (linear - left, spherical - right)

searching ellipsoid). Some theoretical variogram models (see **Figure 7**) had been tested like linear (range about 900 m) and spherical (range about 350 m) with relatively small changes in visual results and MSE. So, we prefer to use models with lower ranges. Also, the Kriged map used larger equidistance (0.02) than the IDW (0.01) just to be sure that a different spatial model will not significantly change the shapes of the main "structures". It was not influenced by the MSE calculation.

5. Conclusions

The main achievements can be summarised as follows:

1. Zonal estimators and interpolators are two different approaches for subsurface sandstone mapping, both with their own properties and advantages.

2. Larger datasets will favour interpolators, smaller datasets favour estimators.

3. Here it is proven that both datasets can be interpolated with the IDW (19 data, "K", 0.000119) and even using the OK (25 data, "L", 0.000676), and in both cases the MSE will be significantly lower than in the estimation with the PR (0.054010 in "K" and 0.040117 in "L").

4. However, both datasets are very specific, meaning that many values are the same, which practically decreased the number of different data from 19 or 25 to only several classes. In such a case pure interpolation will not be so useful, because almost all grid cells are not confirmed with the measured value, i.e. hard data, but are artificially calculated from an interpolation algorithm.

5. Such a disadvantage is the result of the way that data are calculated, where they are indirectly averaged from a mixture of well logs and cores in the sandstone reservoir intervals across the field.

6. Alternatively, their interpretation could be improved with simultaneously running a zonal estimator algorithm, where the mentioned value classes could be easily recognised inside zones, revealing their area and strike (see **Figure 5a, b**).

7. Zonal estimation enhances reservoir development when depositional models are well-defined, like in the presented sandstone reservoirs or their counterparts in other regions of the CPBS.

Acknowledgement

This research was partly carried out within the framework of the projects "Mathematical Research in Geology IX/2024 and X/2025" (led by T. Malvić) and for the bachelor theses of Ivana Brajnović and Maria Rudec (co-mentoring by J. Ivšinović & T. Malvić).

6. References

Al-Mudhafar, W. J.; Dalton, C. A.; & Al Musabeh, M. I. (2017). Metamodeling via Hybridized Particle Swarm with Polynomial and Splines Regression for Optimization of CO2-EOR in Unconventional Oil Reservoirs. SPE Reservoir Characterisation and Simulation Conference and Exhibition, Abu Dhabi, UAE.

Aminian, K.; Ameri, S.; Abbitt, W. E. & Cunningham, L. E. (1991). Polynomial Approximations for Gas Pseudopressure and Pseudotime (SPE-23439-MS), SPE Eastern Regional Meeing.

Andrić, K.; Malvić, T.; Velić, J.; Barudžija, U. and Rajić, R. (2021). Poboljšana analiza i kartiranje geokemijskih i geoloških varijabli u prostoru Bjelovarske subdepresije (Sjeverna Hrvatska), primjenom metode inverzne udaljenosti. (in Croatian)

Araújo, A. (2018). Polynomial regression with reduced over-fitting-The PALS technique. *Measurement,* 124, 515-521.

Bal, S. R. R. (2022). Prediction Of Heat Transfer Performance Using Polynomial Regression. Second International Conference on Artificial Intelligence and Smart Energy (ICAIS), Coimbatore, India, 1735-1740.

Barlak, C. and Ozkazanc, Y. (2011). Battery capacity estimation with inverse distance weighting. *Communications Faculty of Sciences University of Ankara Series A2-A3 Physical Sciences and Engineering*, 52(1), 1-16.

Bokati, L.; Velasco, A. & Kreinovich, V. (2022). Scale-Invariance and Fuzzy Techniques Explain the Empirical Success of Inverse Distance Weighting and of Dual Inverse Distance Weighting in Geosciences. North American Fuzzy Information Processing Society Annual Conference, Halifax, NS, Canada.

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

Chen, F.-W. and Liu, C.-W. (2012). Estimation of the spatial rainfall distribution using inverse distance weighting (IDW) in the middle of Taiwan. *Paddy and Water Environment*, 10(3), 209-222.

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.

Di, J.; Xu, J.; Fu, J.; Li, B. & Wu, M. (2023). Study on the regional ASF prediction method based on the ordinary kriging interpolation. *Physica Scripta*, 99.

Draper, N.R. and Smith, H. (1981). Applied Regression Analysis, New York. A Whiley-Interscience Publication.

Golnaraghi, S.; Moselhi, O.; Alkass, S. & Zangenehmadar, Z. (2020). Modelling construction labour productivity using evolutionary polynomial regression. *International Journal of Productivity and Quality Managemen*t, 31(2), 207-226.

Hodson, T. O. (2022). Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. *Geoscientific Model Development,* 15(14), 5481-5487.

Ivšinović, J. (2019). Odabir i geomatematička obradba varijabli za skupove manje od 50 podataka pri kreiranju poboljšanoga dubinskogeološkoga modela na primjeru iz zapadnoga dijela Savske depresije (*Selection and geomathematical calculation of variables for sets with less than 50 data regarding the creation of an improved subsurface model, case study from the western part of the Sava Depression*), dissertation. Faculty of Mining, Geology and Petroleum Engineering, University of Zagreb, Zagreb. (in Croatian)

Ivšinović, J. and Malvić, T. (2020). Application of the Radial Basis Function interpolation method in selected reservoirs of the Croatian part of the Pannonian Basin System. *Mining of Mineral Deposits*, 14. 37-42.

Ivšinović, J.; Malvić, T.; Velić, J. & Sremac, J. (2020). Geological Probability of Success (POS), case study in the Late Miocene structures of the western part of the Sava Depression, Croatia. *Arab J Geosci*, 13, 714.

Ivšinović, J. and Malvić, T. (2022). Comparison of Mapping Efficiency for Small Datasets using Inverse Distance Weighting vs. Moving Average, Northern Croatia Miocene Hydrocarbon Reservoir. *Geologija*, 65, 1, 47-57.

Jiménez-López, D.; Corcho-Caballero, P.; Zamora, S. & Ascasibar, Y. (2021). Polynomial expansion of the star formation history in galaxies. *Astronomy & Astrophysics*.

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

Kalla, S. and White, C. D. (2007). Efficient Design of Reservoir Simulation Studies for Development and Optimization. *SPE Res Eval & Eng*, 10, 629–637.

Ly, S.; Charles, C. & Degre, A. (2011). Geostatistical interpolation of daily rainfall at catchment scale: The use of several variogram models in the Ourthe and Ambleve catchments, Belgium. *Hydrology and Earth System Sciences*, 15(7), 2259–2274.

Malvić, T. and Velić, J. (2008). Geologija ležišta fluida (Geology of Reservoir Fluids). Faculty of Mining, Geology and Petroleum Engineering, University of Zagreb, Zagreb. (in Croatian)

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. Schattner, U.), InTech, Rijeka, 215-238.

Malvić, T.; Ivšinović, J.; Velić, J. & Rajić, R. (2019). Interpolation of Small Datasets in the Sandstone Hydrocarbon Reservoirs, Case Study of the Sava Depression, Croatia. *Geosciences*, ***9*,** 201.

Malvić, T.; Pimenta Dinis, M.A.; Velić, J.; Sremac, J.; Ivšinović, J.; Bošnjak, M.; Barudžija, U.; Veinović, Ž. & Sousa, H.F.P.e. (2020). Geological Risk Calculation through Probability of Success (PoS), Applied to Radioactive Waste Disposal in Deep Wells: A Conceptual Study in the Pre-Neogene Basement in the Northern Croatia. *Processes*, 8, 755.

Mesić Kiš, I. (2016). Comparison of ordinary and universal kriging interpolation techniques on a depth variable (a case of linear spatial trend), case study of the Šandrovac field. *Rudarsko-geološko-naftni zbornik*, 31 (2), 41-58.

Mesić Kiš, I. and Malvić, T. (2016). Variables selection and linear equations calculation for description of geological surface regional dip. *Proceedings of the Inova Exhibition,* Zagreb.

Moeletsi, M.; Shabalala, Z.; Nysschen, G. & Walker, S. (2016). Evaluation of an inverse distance weighting method for patching daily and dekadal rainfall over the Free State Province, South Africa. *Water SA*, 42 (3), 466-474.

Najafzadeh, M. and Kargar, A. R. (2019). Gene-Expression Programming, Evolutionary Polynomial Regression, and Model Tree to Evaluate Local Scour Depth at Culvert Outlets. *Journal of Pipeline Systems Engineering and Practice*, 10(3).

Novak, K. (2015). Modeliranje površinskoga transporta i geološki aspekti skladištenja ugljikova dioksida u neogenska pješčenjačka ležišta Sjeverne Hrvatske na primjeru polja Ivanić (*Surface transportation modelling and geological aspects of carbon-dioxide storage into Northern Croatian* *neogene sandstone reservoirs, case study Ivanić field*), dissertation. Faculty of Mining, Geology and Petroleum Engineering, University of Zagreb, Zagreb. (in Croatian)

Qiao, P.; Lei, M. & Yang, S. (2018). Comparing ordinary kriging and inverse distance weighting for soil as pollution in Beijing. *Environ. Sci. Pollut. Res.* 25, 15597–15608.

Rohma, N. (2022). Estimation of Ordinary Kriging Method with Jackknife Technique on Rainfall Data in Malang Raya. *International Journal on Information and Communication Technology (IJoICT),* 8, 22-39.

Said, L. (2023). Forecasting Reservoir Pressure for Vertical Oil Well Using Supervised Machine Learning Model. SPE/IADC Middle East Drilling Technology Conference and Exhibition, Abu Dhabi, UAE.

Setianto, A. and Triandini, T. (2013). Comparison of kriging and inverse distance weighted (IDW) interpolation methods in lineament extraction and analysis. *Journal of Applied Geology*, 5(1), 21-29.

Trebuňa, F; Ostertagová, E.; Frankovský, P. & Ostertag, O. (2016). Application of Polynomial Regression Models in Prediction of Residual Stresses of a Transversal Beam. *American Journal of Mechanical Engineering*, 4(7), 247-251.

Velić, J. (2007). Geology of Oil and Gas Reservoir. Faculty of Mining, Geology and Petroleum Engineering, University of Zagreb, Zagreb. (in Croatian)

Velić, J.; Malvić, T.; Cvetković, M. & Vrbanac, B. (2012). Reservoir Geology, Hydrocarbon Reserves and Production in the Croatian part of the Pannonian Basin System. *Geologia Croatica*, 65(1), 91-101.

Wan, X. (2019). The Influence of Polynomial Order in Logistic Regression on Decision Boundary. IOP Conference Series Earth and Environmental Science, Baubau, Indonesia.

Wang, J.; Li, C.; Cheng, P.; Yu, J.; Cheng, C.; Ozbayoglu, E. & Baldino, S. (2024). Data Integration Enabling Advanced Machine Learning ROP Predictions and its Applications (OTC-35395-MS). Offshore Technology Conference.

Weslati, O.; Bouaziz, S. & Serbaji, M. M. (2022). The Efficiency of Polynomial Regression Algorithms and Pearson Correlation (r) in Visualizing and Forecasting Weather Change Scenarios. In; Recent Advances in Polynomials. pp. 20.

Wu, S. and Zhang, T. (2021). Active Learning Model on Wind Turbine Power Generation Based on Polynomial Regression. In: Proceedings of the 2021 5th International Conference on Deep Learning Technologies (ICDLT ‘21). Association for Computing Machinery, New York, NY, USA, 111-115.

Yao, C.; Ren, X.; Valiveti, D.; Ryu, S.; Chaney, C. & Zeng, Y. (2023). Machine Learning Based FPSO Topsides Weight Estimation for a Project on an Early Stage (OTC-32304-MS). Offshore Technology Conference.

SAŽETAK

Usporedba rezultata interpolacije polinomnom regresijom te onih inverznom udaljenošću i običnim krigiranjem; primjer neogenskih pješčenjaka u Savskoj depresiji, Hrvatska

Analizirana su dva skupa podataka šupljikavosti. Oni su prostorno prikazani polinomnom regresijom, skr. PR (zonalna procjena), inverznom udaljenošću, skr. IU, i običnim krigiranjem, skr. OK (interpolacije). Podatci su prikupljeni iz pješčenjačkih ležišta „K" (19 točaka) i „L" (25 točaka), donjopontske starosti, smještenih u Savskoj depresiji u Hrvatskoj. Karte su uspoređene rezultatima krosvalidacije (srednjom kvadratnom pogrješkom, skr. SKP) te vizualnom usporedbom glavnih oblika linija izoporoznosti. Obrađeni skupovi podataka razmatrani su kao mali (ležište „K") i veliki (ležište „L"). Za ležište „K" SKP iznosi 0,000119 (IU) nasuprot 0,05401 (PR). U ležištu „L" SKP je iznosio 0,000676 (OK) nasuprot 0,040117 (PR). Zonalna procjena nije se pokazala primarnim izborom za kartiranje takvih skupova od 20-ak podataka. Nasuprot njoj, linearni interpolatori (IU i OK) predstavljaju mnogo razumniji odabir, posebno ako se može modelirati pouzdan prostorni model temeljen na normalnoj razdiobi analizirane varijable (šupljikavosti). U takvim slučajevima preferira se tehnika OK. Zonalna procjena može se smatrati uporabljivim dodatkom interpretaciji ležišta, posebno u onim dijelovima gdje je rješenje manje pouzdano, npr. u prijelaznim zonama.

Ključne riječi:
pješčenjačka ležišta, šupljikavost, zonalna procjena, interpolacija, Savska depresija, Hrvatska

Author's contribution

Tomislav Malvić (Full tenured Professor): conceptualization, methodology, investigation, validation, writing and editing. **Josip Ivšinović** (PhD): conceptualization, methodology investigation, software, writing and editing. **Uroš Barudžija** (Associate Professor): investigation, validation, writing and editing. **Ivana Brajnović** (student): visualization, writing and editing, **Maria Rudec** (student): visualization, writing and editing.
All authors have read and agreed to the published version of the manuscript.