 doi:10.20944/preprints202004.0471.v1

Article

Recent advances in geomathematics: Croatian examples from subsurface geological mapping and biostatistics

Tomislav Malvić [1], Marija Bošnjak [4], Josipa Velić [1], Jasenka Sremac [2], Josip Ivšinović [5], Maria Alzira Pimenta Dinis [3], Uroš Barudžija [1]

1 University of Zagreb, Faculty of Mining, Geology and Petroleum Engineering; tomislav.malvic@rgn.unizg.hr , https://orcid.org/0000-0003-2072-9539; josipa.velic@rgn.unizg.hr; uros.barudzija@rgn.unizg.hr , https://orcid.org/0000-0002-1617-9362

2 University of Zagreb, Faculty of Science; jasenka.sremac@geol.pmf.hr , https://orcid.org/0000-0002-4736-7497

3 UFP Energy, Environment and Health Research Unit (FP-ENAS), University Fernando Pessoa (UFP), Praça 9 de Abril 349, 4249-004 Porto, Portugal; madinis@ufp.edu.pt, https://orcid.org/0000-0002-2198-6740

4 Croatian Natural History Museum, Zagreb

5 INA Plc., Zagreb; josip.ivsinovic@ina.hr , https://orcid.org/0000-0002-7451-1677

* Correspondence: marija.bosnjak@hpm.hr

Received: date; Accepted: date; Published: date

Abstract: Geomathematics is extremely important in geosciences, particularly in the geology. The key for any geomathematical analysis is the definition of a typical model to be applied for further prognosis, either through deterministical or stochastical approaches. The selection of the appropriate procedure is presented in this paper. Two different geomathematical subfield datasets were used in subsurface geological mapping and palaeontology and different biostatistics applications, representing important geomathematical subfields in the Croatian geology. The different subsurface interpolation methods, tested, validated and recommended for application, were used to obtain the best possible outcome in reservoir modelling, in the cases with small datasets. Cross-validation may be selected as the main selection criteria, applied to the Croatian part of the Pannonian Basin System (abbr. CPBS). Recent advances in biostatistics applied in palaeontology and case studies from Croatia are also presented, where biometric studies are of significant importance in fossil biota. Data, methods and problems in geosciences is a vast subject, and address a wide spectrum of fundamental science. Because geology includes subsurface and surface geology, and very different datasets regarding variable and number of data, here are chosen two representative case study groups with original samples from Northern Croatia. Subsurface mapping has been presented on limited petrophysical datasets from the Northern Croatian, Miocene, hydrocarbon reservoirs. Biostatistics has been presented on very different samples, allowing to achieve paleoenvironmental reconstructions of size of relevant fossils, as dinosaurs or other species and their paleoenvironments. All examples highlight examples of the valuable application of geomathematical tools in geology. The results, cautiously validated and correlated with other, non-numerical (indicator, categorical) geological knowledge, are of enormous assistance in creating better geological models.

Keywords: geomathematics; geostatistics; subsurface geological mapping; biostatistics; palaeontology; Croatia

1. Introduction

Development of geomathematics during past is a very dynamic and non-linear process. In the early days (until 80's of 20th century) the geomathematics and geostatistics were considered as synonymous, so the pioneer works in geostatistics are considered as breakthroughs in the entire

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 April 2020 doi:10.20944/preprints202004.0471.v1

geomathematics applied in geosciences. The first results of the geostatistical research (different from research in the field of "spatial statistics") had been published by [1,2,3] where the Kriging is described for the first time. The same algorithm had already been applied earlier by [4] for estimation of gold nuggets concentrations in the South African mines. The Matheron's foundation has been based on the least square method and linear Gaussian model, what stayed as base until the present day. Following the linear models, authors as [5] and [6] also developed non-parametric and non-linear geostatistics. In parallel, geostatistics is developed together with the applied statistics by [7]. [8] made an important step toward unification of geostatistics and other data analysis methods in geology, describing three main branches of spatial statistics – geostatistics, spatial variations and spatial point processes. That is what we today call – geomathematics.

Here, it is important to mention that the use of geostatistics (even statistics) is closely related with exploration and production of hydrocarbon reservoirs (e.g., [9,10]). In the late 80's of 20th century, geostatistics offered new algorithms, allowing to obtain much better reservoir characterisation, in particular visualisation. However, from the early days in geostatistics and later geomathematics, the main factor in selection of a method was a number and distribution of data. Those two problems are often intertwined, although distribution of data is considered as a fundament for any later analysis.

As in any data-based analysis, geomathematics is highly dependent on hard data, i.e., measurements, aiming to predict values in non-sampled volumes (Figure 1.1). The problem had been solved differently. As geological variables are mostly presented in deterministic ways, the knowledge about (sub)surface is always partial. In fact, the models are stochastic but too complex so that available mathematical approximations, restricted with limited data, could be presented in such way. The geomathematics offered the approaches designed for object-based models, where objects are datasets analysed and visualised with different spatial methods (e.g., 11,12]). Most of them are deterministical (Kriging) but some approaches could be stochastical (simulations).

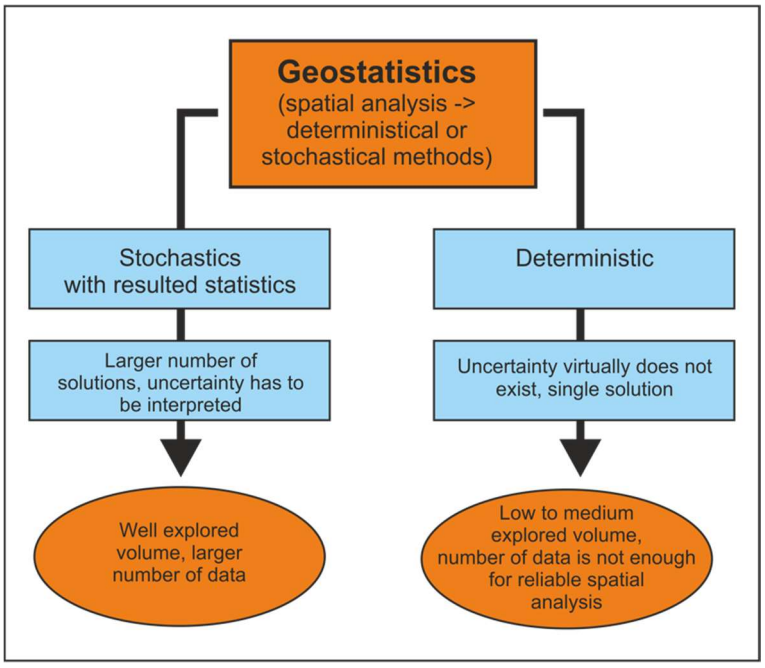


Figure 1.1. Relation between geostatistical approach and number of data.

Any classification of such models can be categorised. Here we decided to divide them into:

1. Full deterministical, models where volume is well known, without uncertainties, possible correlated, and settings are known and established. Such knowledge is rare, but many areas are approximated in such a way.
2. Stochastical volumes, where uncertainties cannot be full described and permanently exists. However, the probability model allows to make predictions and estimations with different

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 April 2020 doi:10.20944/preprints202004.0471.v1

geomathematical algorithms. Such are the most of analysed (sub)surface volumes, but stochastical approach asks for more experiences and, contradictory, more data than deterministical approach.

3. Unpredictable volumes, where analysed variables could not be described by any algorithm or just the number of data is not enough high so that any observation is valid and general.

The key goal for any geomathematical analysis is the definition of a typical model that can be applied for further prognosis in similar or same conditions. Any such prognosis needs to be based on valid choices, grounded on previous case studies where decision trees are made. Such choices are always based on the number of inputs, regarding variables and dataset size. Generally, the explored geological volume is longer researched so that the object of researching could be easier improved with, both, deterministical or stochastical approach. Decision depends on experience, knowledge and readiness to accept uncertainties in future estimations. Any multivariable approach is benefit (like Co-Kriging) but asks for well documented connections among depended variables. Significant inherited (measurements limitations, equipment error) and man-made (biased sampling) uncertainties forced stochastics, but limitation of such approach is a must-have spatial model. However, it is not a condition at all for simpler algorithms like Inverse distance weighting (IDW), Modified Shepard's Method (MSM), Nearest neighbourhood (NN) or similar. The largest limitation is the number of data (Figure 1.2), especially if the primary variable is such to be defined in the entire dataset. The scares or non-representative dataset greatly limit application of statistics. Even statistical representative sets (e.g., n>30) are much easier when analysed with parametric statistics that requires Gaussian distribution. Oppositely, non-parametric statistics is only a choice, which can limit the number of tests and mapping, in particular.

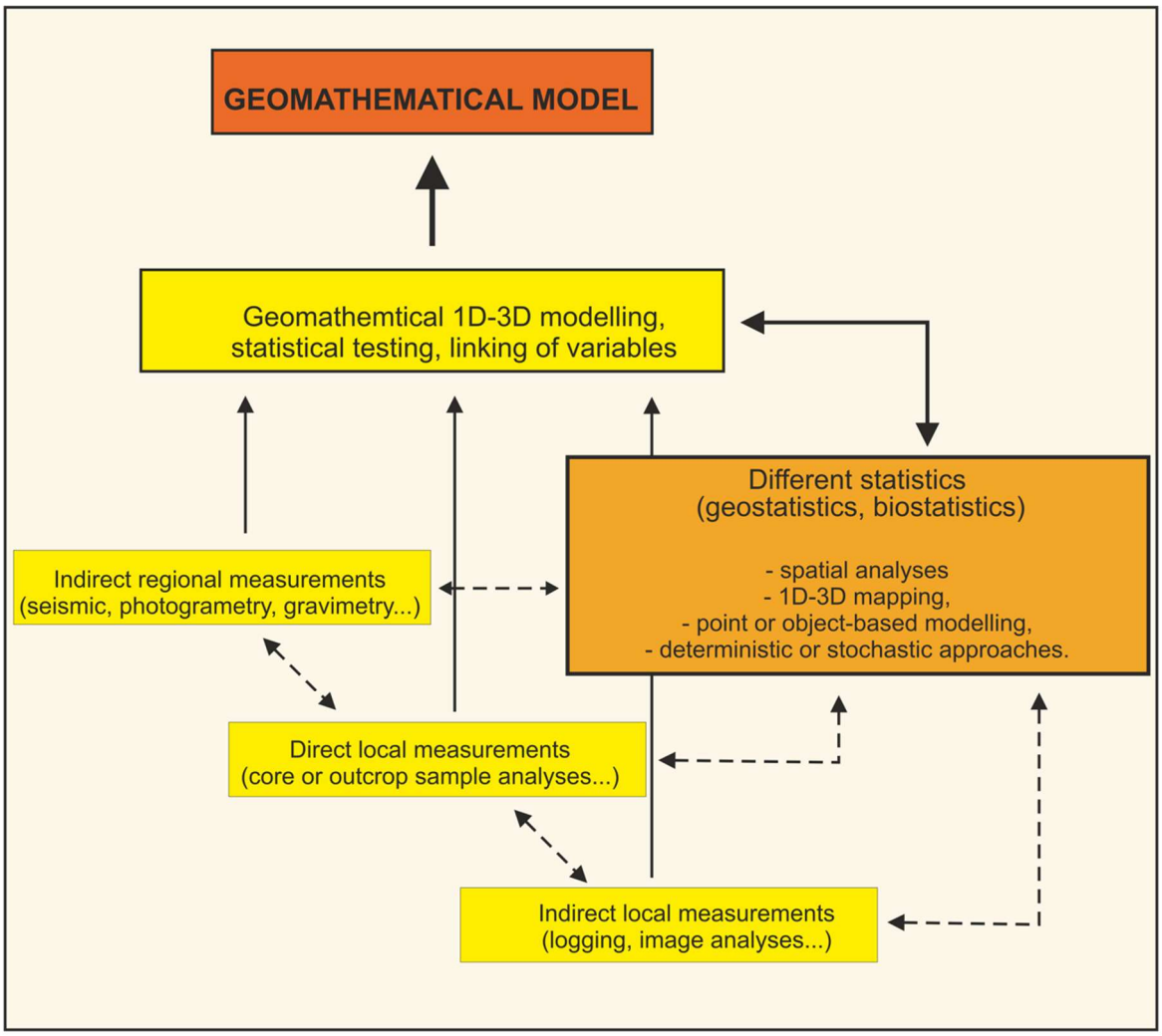


Figure 1.2. Simple decision tree for geomathematical analysis.

However, and despite all limitations, geomathematics has many favourable and robust tools and algorithms for analysis of almost all geological datasets. It is especially valid if geomathematics is considered as a field divided into three sub-fields: geostatistics, statistics applied to geosciences and neural networks applied to geoscientific data.

The main challenge is a selection of an appropriate procedure. How to find such tools, but only in a very tiny spectre of geomathematics and geosciences, has been presented in this work, through the examples of two different geomathematical subfield datasets. The first one refers to the subsurface geological mapping, as described above, and the second one refers to palaeontology and biostatistics.

The starting point in the paleontological research is the assumption presented by day biota and processes which are the key to understand the Earth history. Therefore, the biological research methods and facts, such as biostatistics (biological statistics or biometry), are common in taxonomical and palaeoecological studies in palaeontology due to the assumption that species can be defined by its morphology, including the measurable parameters.

The development of biostatistics dates back to the 19th century, with Francis Galton (1822–1911), "the father of biostatistics and eugenics". His methodology, used in the analysis of biological variation, is considered as the foundation for the application of statistics to biology [13]. The term "biometry" was coined by the zoologist W.F.R. Wilson (1860–1906), who was working with Karl Pearson on the application of statistical methods in biology [13]. The application of biometry in the systematic description of plants and animals was pointed out by [14], where he describes the necessity of specific descriptions of taxon characteristics, in order to precisely describe the specimen. The rising impact of biometry resulted in the establishment of the Biometric Society on September 6, 1947 at Woods Hole, USA, as described by [15]. The first president of the Biometric Society was Sir Ronald Aylmer Fisher (1890–1962). The Society was later renamed to International Biometric Society. The Biometric Section of the American Statistical Association started publishing the Biometrics Bulletin in 1945, which was renamed to Biometrics, in 1947.

Two significantly different datasets and applications in geological subsurface mapping and biostatistics (biometrics) presented in this paper, represent, in the last decade, as well as currently, the most progressive and publicised geomathematical subfields in the Croatian geology.

2. Mathematical basics of algorithms applied in the presented case studies

2.1. Kriging method

The Kriging (as well as the Co-Kriging and stochastical simulations) is a group of statistical estimation methods. The specificity of the Kriging (e.g., [5,16,17]) is the definition as the best linear unbiased estimator (abbr. BLUE), although it is valid only for specific datasets. The strength of Kinging approach is due to the weighting coefficient calculation, the procedure based on the minimisation of Kriging variance. The linear means that estimation has been done by combination of hard data; the unbiased makes sure that the estimation expected value is the real as for the entire possible population. The estimator defines applied methodology. The linear estimation is shown in Equation 2.1:

$$Z_k = \sum_{i=1}^{n} \lambda_i \times Z_i \tag{2.1}$$

Where:

Z_k - *value of the regionalised variable calculated at location "k";*

Z_i - *value of the regionalised variable measured at location "i";*

λ_i - *weighting coefficient calculated by Kriging matrices for location "i".*

The necessary condition for the Kriging estimation is that the measured Zi values are characterised with normal distribution or, at least, that such property is assumed for that variable in the case of a large number of measurements. Compared with simpler estimation algorithms, the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 April 2020 doi:10.20944/preprints202004.0471.v1

Kriging, is more time-consuming interpolation method, but also better tool for handling with highly clustered data. Oppositely, the Kriging results in very weak works with small datasets (n<20), unable to give origin to meaningful spatial models. The spatial (variogram) tool is powerful when applied with enough data and background knowledge. AS mentioned earlier, the main advantage of the Kriging is the weighting coefficient calculation. After the spatial model has been set up, the calculation of coefficient is not dependence on their value, but exclusively on distance between measured points and location where the value is not known. Such value is also called "statistical distance", referring on their derivation from variogram, not from values. The Kriging equations (Equation 2.2) are calculated using matrices. In two of them (W, B), the values are given with variogram values, which depends on distances among observed locations:

$$[W]\times[\lambda]=[B] \tag{2.2}$$

There are numerous Kriging techniques, each of them differenced by some modification in matrices. The most used in Croatian case studies are herein designated by Simple, Ordinary, Indicator and Universal Kriging. The Simple Kriging is the basis for all the other available techniques. The matrix is presented in Equation 2.3:

$$\begin{bmatrix} \gamma(Z_1-Z_1) & \gamma(Z_1-Z_2) & \dots & \gamma(Z_1-Z_{n1}) \\ \gamma(Z_2-Z_1) & \gamma(Z_2-Z_2) & \dots & \gamma(Z_2-Z_n) \\ & & & \\ \gamma(Z_n-Z_1) & \gamma(Z_n-Z_2) & \dots & \gamma(Z_n-Z_n) \end{bmatrix} \times \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \\ \lambda_n \end{bmatrix} = \begin{bmatrix} \gamma(x_1-x) \\ \gamma(x_2-x) \\ \\ \gamma(x_n-x) \end{bmatrix} \tag{2.3}$$

Although the basic technique, it is the only one that do not satisfy the condition of unbiased estimation, because it is the only equation without constraint. Such constraint(s) could be linear or non-linear.

The most often used technique is presented in Equation 2.4 with additional constraint – Lagrange multiplier (μ), aiming to find the local minima and maxima of the function, subjected to equality constraints, i.e., to minimise the Kriging variance.

$$\begin{bmatrix} \gamma(Z_1-Z_1) & \gamma(Z_1-Z_2) & \dots & \gamma(Z_1-Z_n) & 1 \\ \gamma(Z_2-Z_1) & \gamma(Z_2-Z_2) & \dots & \gamma(Z_2-Z_n) & 1 \\ & & & & 1 \\ \gamma(Z_n-Z_1) & \gamma(Z_n-Z_2) & \dots & \gamma(Z_n-Z_n) & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix} \times \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \\ \lambda_n \\ \mu \end{bmatrix} = \begin{bmatrix} \gamma(x_1-x) \\ \gamma(x_2-x) \\ \\ \gamma(x_n-x) \\ 1 \end{bmatrix} \tag{2.4}$$

2.2. Inverse distance weighting (IDW) interpolation method

IDW is a widely used interpolation method, both for small and large datasets. The unknown value is calculated based on all known points and inversely proportional to their distances (Equation 2.5, e.g. [18,19,20]) and is defined as:

$$z_{IU}=\frac{\frac{z_1}{d_1^p}+\frac{z_2}{d_2^p}+\dots\frac{z_n}{d_n^p}}{\frac{1}{d_1^p}+\frac{1}{d_2^p}+\dots\frac{1}{d_n^p}} \tag{2.5}$$

Where:

Z_{IU} - *estimated value,*

$d_1 \dots d_n$ - *distance between estimated value and known value 1…n,*

p - *power (distance) exponent,*

$z_1 \dots z_n$ - *known values at locations 1…n.*

The mapping results are greatly influenced by power exponent, which could stress the influence of more distance points and smooth the map (for p<=2) or force very local estimation (p>2) and even, for large "p", result in zonal estimation, i.e., in map like Voronoi polygons. This method has been

 doi:10.20944/preprints202004.0471.v1

proved for mapping problems in the Croatian part of the Pannonian Basin System (abbr. CPBS) for all datasets where clustering was not largely imposed, and for datasets smaller than 15 points too (e.g., [21,22]).

2.3. Basics of the Nearest neighbourhood (NN) estimation method

NN is the simplest statistical estimation method when unknown point is estimated only from the closest known value. The results are valued polygons, like Voronoi diagram. The distance between the points is Euclidian (Equation 2.6):

$$d(x,T)=\sqrt{\left(X_1-T_1\right)^2+\ldots+\left(X_n-T_n\right)^2} \tag{2.6}$$

Where:

d - *distance,*

n - *n-th pair of points,*

x and T - *unknown and measured points.*

The method is meaningful to apply only for very small datasets, like 5 or less points. The output is not a map, but schematic polygon view.

2.4. Basics of the Natural neighbourhood (NaN) estimation method

NaN is the modification of the NN and results are also shown as Voronoi diagrams (polygons). The unknown point is estimated from the several nearest points (e.g., [23,24,25]) using Equation 2.7:

$$X(x,y)=\sum_{i=1}^{n}(\mathrm{w}_i\, A(\mathrm{X}_i,\mathrm{Y}_i)) \tag{2.7}$$

Where:

X(x,y) - *estimated value in point (x,y),*

A(Xi,Yi) - *known value in point (Xi,Yi),*

wi - *proportion of polygon „i" in total area.*

2.5. Modified Shepard's Method (MSM)

The MSM interpolation is a modification of the IDW method, with the aim of reducing the expressive local values (outliers, extremes) that could cause "bull-eyeing" or "butterfly shape" effects. The method was developed by [26] and it is why is named as Shepard's method. The modification of the method was carried out in the works of, e.g., [27] and [28]. The estimation is done by Equation 2.8:

$$F(x,y)=\frac{\sum_{k=1}^{n}W_k(\mathrm{x},\mathrm{y})\cdot \mathrm{Q}_k(x,y)}{\sum_{i=1}^{n}W_i(x,y)} \tag{2.8}$$

where:

F - *MSM function;*

W - *relative weights;*

Qk - *bivariate quadratic function;*

x, y - *data coordinates;*

n - *number of data.*

MSM used so called relative weights determined (Equation 2.9):

$$W_k(x,y)=\left[\frac{(R_\omega-d_k)_+}{R_\omega\cdot d_k}\right]^2 \tag{2.9}$$

Where:

W - *relative weights;*

d_k - *Euclidean distance between points at locations (x, y) and (x_k, y_k);*
R_ω - *radius of influence around node (x_k, y_k).*

2.6. Cross-validation as numerical estimation of mapping error

The cross-validation is a numerical procedure, which can be applied also as error-based comparison tool for several maps with the same input, but sequentially interpolated with two or more methods. The procedure is repeated as many times as there are measured (hard) values, dropping one known point out and calculating the estimation in the same location from the rest of the hard data (Equation 2.10). The result is often named as Mean Square Error (abbr. MSE, e.g., [21,29, 30,31]). This value is often used as criteria for the most appropriate map selection in the case of small datasets in the CPBS (e.g., [32,33]).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(SV - P)_i^2 \quad (2.10)$$

Where:
MSE - *Mean Square Error value,*
n - *number of known values,*
SV - *measured value of point „i“,*
P - *estimated value of point „i“,*
i - *i-th point.*

2.7. Shannon-Wiener index or Shannon diversity index (H)

In paleoecological analyses, one of the goals is to explore species richness and diversity in the analysed data sets, and to compare biological diversity between the samples with an uneven number of species and individuals. In the examples presented in this paper, authors showed part of the research on the biodiversity of microfossils Foraminifera, where Shannon-Wiener or Shannon diversity index (H) is used as one of the measures of species diversity in one sample, and between samples [34]. The “H” is calculated by the Equation 2.11 [34]:

$$H = -\sum_{i=1}^{R} p_i \times \ln p_i \quad (2.11)$$

Where:
p_i - *a proportion of individuals belonging to the i^{th} species in the sample,*
ln - *a natural logarithm.*

Equation 2.11 shows dependence of H (Shannon index) on pi (proportion of individuals, and if all species in the sample are equally represented, H is at its maximum [34].

3. Recent advances in geomathematical mapping in small datasets and case studies from Croatia

During 2019 and 2020, broad testing of small datasets mapping has been applied [33,35] to the Croatian part of the Pannonian Basin System (abbr. CPBS). A small subsurface sample set is considered to be a set of measurements which includes [33] less than 20 inputs data. Furthermore, such datasets could be subdivided in groups with respect to number of data input: a) 1-5, b) 6-10 & c) 11-19. One example is selected here when the reservoir mapping is done by mathematically and simpler methods (compared with previously widely used Kriging) and results are accepted as the best possible outcomes for further reservoir developing. The permeability maps of the Lower Pontian “K” reservoir (Lower Pontian age, 18 data) of the field "B" are shown in Figure 3.1.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 April 2020 doi:10.20944/preprints202004.0471.v1

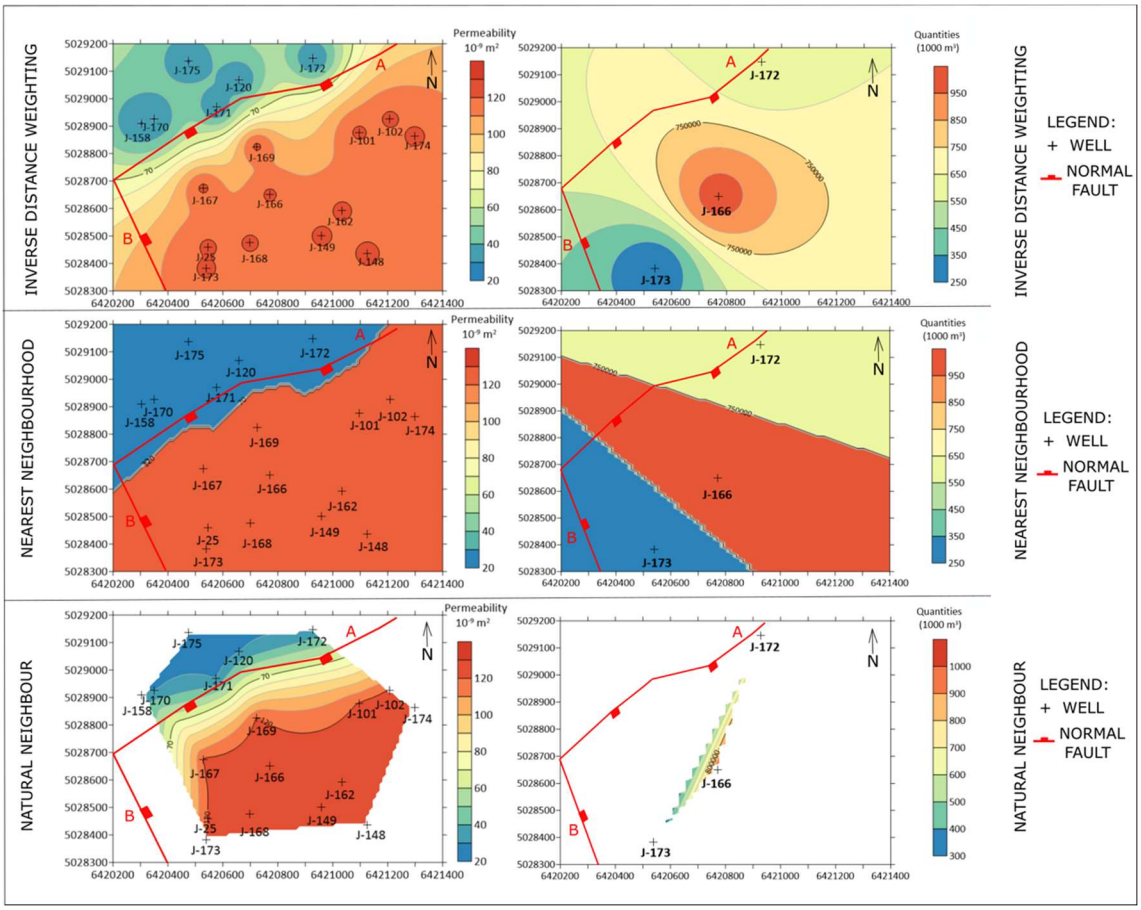


Figure 3.1. Results of IDW, NN and NaN methods (from top to bottom) of the permeability (left) and injected volumes (right) in the "K" reservoir [33].

All maps obtained with different methods (Figure 3.1) are validated with a cross-validation (Table 3.1) and visual assessment (where the larger "bull-eyes" areas mean worse interpolation).

Table 3.1. Summary results of cross-validation for IDW, NN, NaN and MSM methods [33].

Variable	Number of data	Value of Cross-Validation		
		Inverse Distance Weighting	Nearest Neighbourhood	Natural Neighbour
Injected volumes	3	$2.86 \cdot 10^{11}$	$3.96 \cdot 10^{11}$	-
Permeability	18	480.8	1397.4	1044.7

Two interpolation (IDW, NaN) and one zonal (NN) method, gave different mapping results as well as cross-validation errors, as expected. Interestingly, each of them led to at least one useful information about analysed reservoirs, i.e., about connection between permeability and injected water volumes, including the role of some fault zone. The IDW method algorithm remains the main interpolation method of mapping for small reservoir dataset in the Northern Croatia. Other interpolation methods, NN, NaN, may be additional information. But the main advantage was that such datasets could be divided into three classes regarding their mapping, as follows: (a) 1-5, (b) 6-10 and (c) 11-19 inputs. The "class a" could not be analysed with the NaN method because it is often not possible to calculate the cross-validation and the interpolated area is very small regarding unit margins. In the "class b" and "class c", all three methods gave results, and the main selection criteria could be cross-validation.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 April 2020 doi:10.20944/preprints202004.0471.v1

[32] also analysed the possibility of artificially increasing the input data set using the "jack-knifed" method. The presented analysis is the first of such a kind in the Sava Depression (Northern Croatia). It represents the continuation of previous geostatistical analyses conducted in that depression and the entire CPBS. The "jack-knifed" method was applied on porosity of reservoir "K" (19 data) of the "B" field (Figure 3.2).

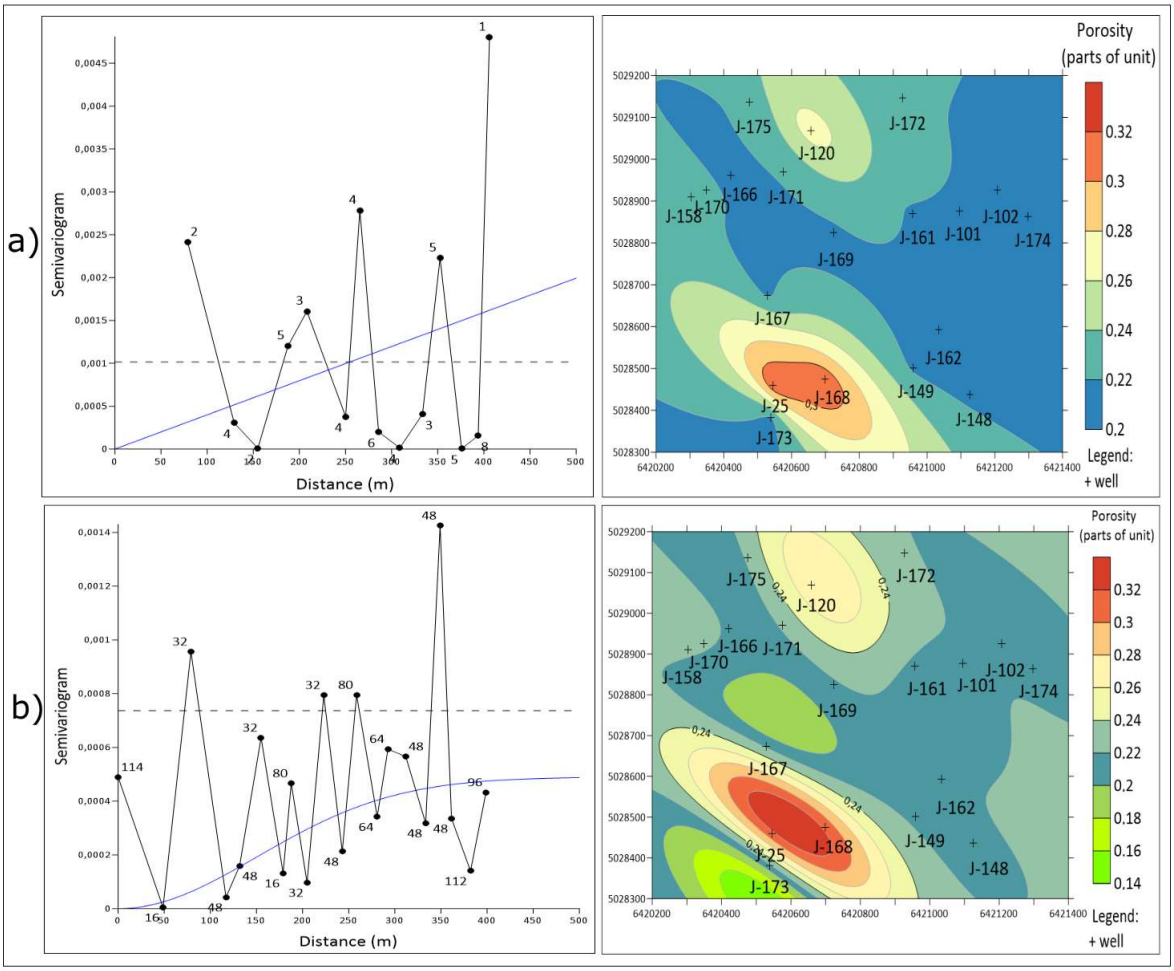


Figure3.2. Experimental semivariograms and porosity maps for the "K" reservoir obtained by the Ordinary Kriging (OK) method: a) without the "jack-knifed" method and b) with the "jack-knifed" method [32].

The obtained porosity maps were analysed by comparing the cross-validation values and expression of the "bull-eyes" effect. The results of the analysis of the "jack-knifing" method are summarized in Table 3.2.

Table 3.2. Comparison of cross-validation values for OK maps based on original and "jack-knifed" semivariograms [32].

Field/ reservoir	OK (original semivariogram)	OK (jack-knifed semivariogram)	Recommendation
"B"/"K"	0.001320 (linear)	0.000970 (Gaussian)	OK with jack-knifed semivariogram

The results in Table 3.2 confirm the possibility of applying the "jack-knifing" method to reservoirs with small data input and should be compared with the maps obtained by the IDW

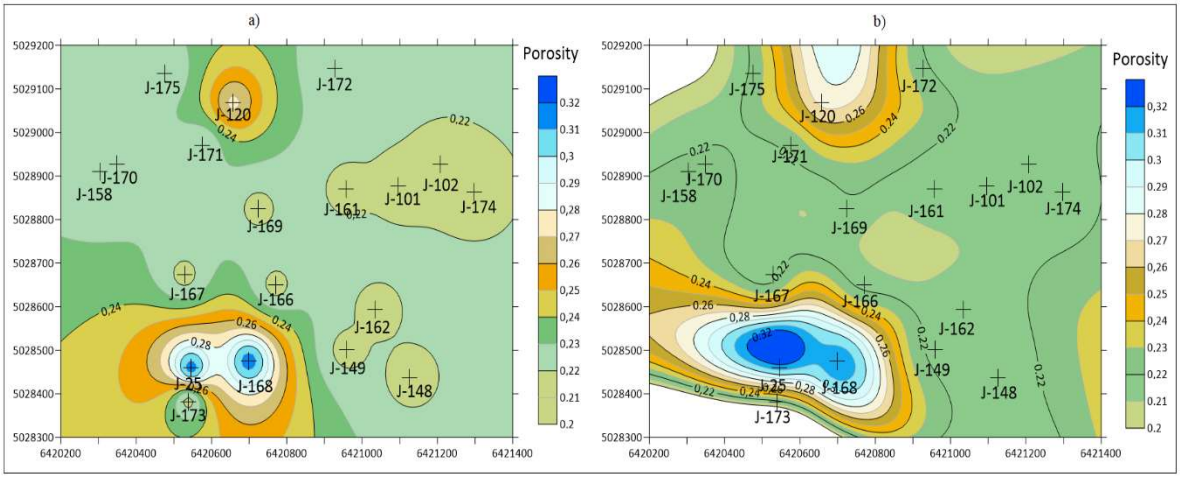
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 April 2020 doi:10.20944/preprints202004.0471.v1

method. Oppositely in another analysed reservoir, the OK was not accepted as interpolation method, but IDW has been accepted. It was the case when jack-knifed did not yield any progress in spatial modelling and the Kriging has been abandoned as an approach.

The permanent problem of small datasets could be oversized with new data. Such data can be obtained with new sampling, but also with the creation of new artificial data, based on the statistical properties of original dataset. The jack-knifing is one of such method, appropriate for datasets of 15–30 points, where the basic, descriptive statistics are more or less representative (variance and mean), and the Gaussian distribution can be assumed. In the presented analysis, the original semivariogram results were highly uncertain, with large oscillations, a small number of data pairs per class and unknown nugget. Consequently, the linear model was the only acceptable theoretical model to use. Due to fact that small dataset could not be statistically representative, the new kriged maps interpolated from "jack-knifed" semivariograms has been tested (a) visually (maps without the "bull-eye" or "butterfly" effects are better) and (b) numerically, using cross-validation and comparing with simpler method of the IDW. Obviously, the results were better in one of the two cases where such validation has been applied.

The next examples are taken from [35] and compare the differences between the results obtained with IDW and MSM (Modified Shepard Method) methods. The IDW does not use weighting coefficient, i.e., each value is "weighted" by a simple (powered) inversely proportional distance from the measured point. The MSM uses relative weights. The porosity, permeability and thickness maps, interpolated with IDW and MSM are given in Figure 3.3. They show the oil reservoir "K" of the Lower Pontian age in the Sava Depression.

The maps obtained by the IDW and SMS methods could be assessed in two ways. One is numerical, using cross-validation. The another is quick-look searching for observable feature of highly expressed local value, i.e., bull-eye or butterfly shape effects. The expected advantage of the MSM is the larger smoothing of the shapes, what is confirmed in that analysis (Figure 3.3). The numerical cross-validation strongly favoured the IDW (Table 3.3).



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 April 2020 doi:10.20944/preprints202004.0471.v1

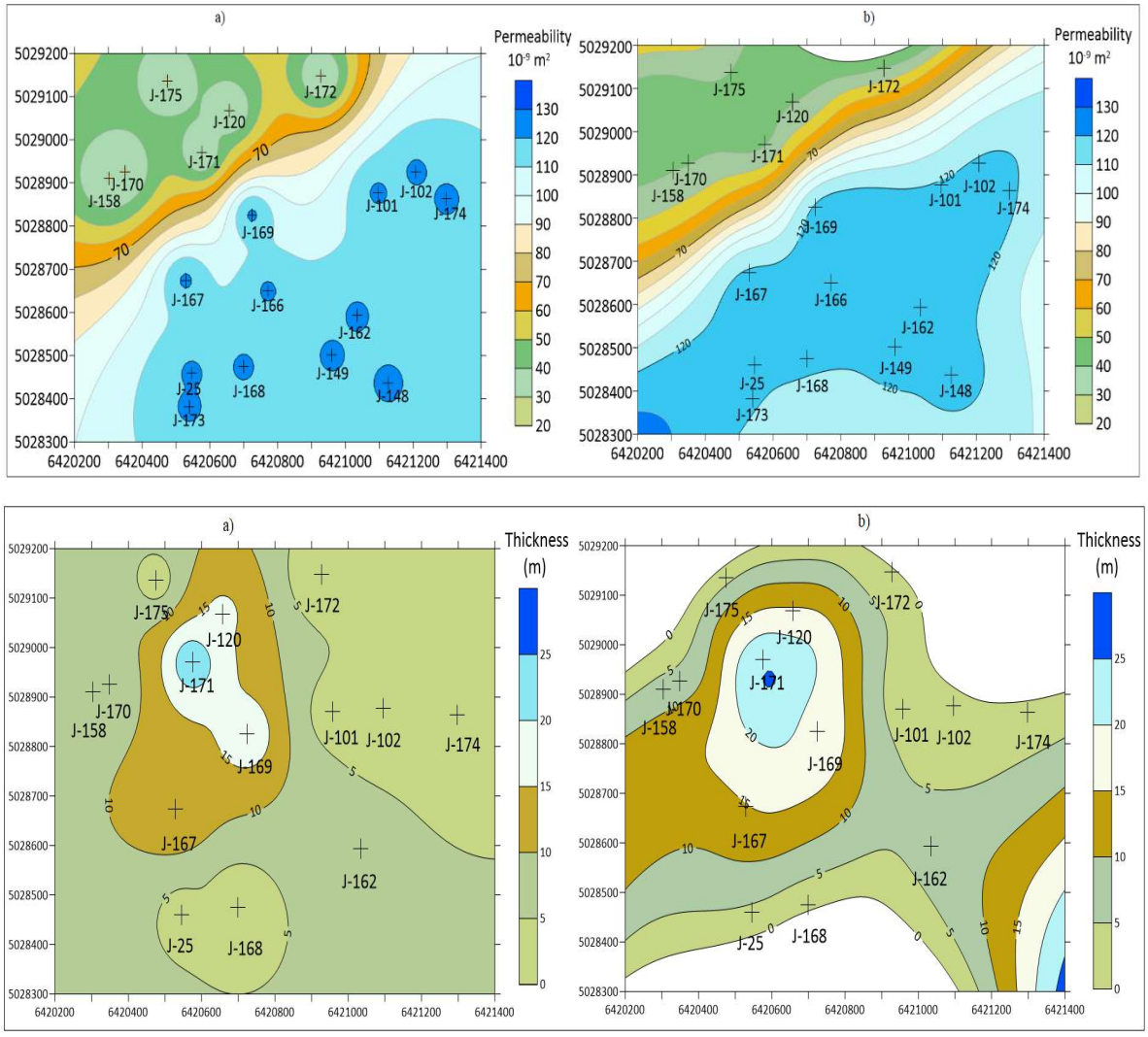


Figure 8. The mapping of the Lower Pontian "K" reservoir, the Sava Depression, Northern Croatia. Left – IDW results, right – MSM results. Top – porosity, middle permeability, down – thickness [35].

The difference resulted from the different mathematical backgrounds of that two methods (Malvić et al. 2020), because the IDW takes into account all measured points or work with general searching radius (or radii for ellipsoid), but the MSM works with local searching by default. It is why cross-validation was higher for MSM - for porosity 289 %, permeability 7 %, and thickness 49 %.

Table 3.3. Cross-validation of the IDW and SMS methods applied in reservoir "K" [35].

Description	No data	Cross-validation	
		Inverse Distance (IDW)	Modified Shepard's Method (MSM)
Porosity	19	0.00119	0.00345
Permeability	18	480.8	516.1
Thickness	14	40.7	60.5

Both methods, obviously led to appropriate quick assessment of the reservoir. However, it was also shown that visual assessment is sometimes the more important criteria than purely numerical cross-validation, what is a crucial conclusion for subsampled reservoirs of the CPBS, and stressed the importance of human and geological expertise, and not purely application of interpolation algorithms. Consequently, [35] recommended the MSM for subsurface geological mapping of

 doi:10.20944/preprints202004.0471.v1

Neogene reservoirs in Northern Croatia in (a) Number of samples smaller than 20 measured values, and/or (b) for early exploration phase or later development phase when the number of measurements of selected property is small, but a quick insight in spatial distribution of such variable is necessary.

4. Recent advances in biostatistics applied in palaeontology and case studies from Croatia and the wider region

Paleontological studies published by numerous authors, including those from Croatia, almost always include basic numerical analyses in recognizing the different taxa. In Croatia, [36] measured the dimensions of the bivalve shells (length, width, length/width ratio of the shell, apical angle) in order to recognize the bivalve subspecies. In her dissertation and several published papers (e.g. [37]), A. Sokač applied biometry in order to present the differences in growth pattern of male and female ostracods. One of the earliest graphically substantiated biometric analysis on the fossil assemblage from Croatia was published by [38], who studied taxonomy and biometry of Eocene corals. The authors distinguished two coral species based on the biometric analyses of the smallest and the largest diameter of the calyx, and the height of the coral calyx plotted in a scatter diagram. Looking at the dispersal of the measured parameters, two areas of dispersal could be recognized, indicating the existence of different species between measured specimens.

During the last decades, a number of global researches were focused on the paleoecology of terrestric, fresh-water or marine biota. In Northern Croatia, Miocene deposits from the Paratethys epicontinental Sea comprise the marine invertebrate fauna, mostly foraminifers, mollusks and ostracods, which were often subject to biostatistics analyses (e.g. [39,40,41,42]). The following data, common in palaeoecological studies, are presented in the referred papers: plankton/benthos ratio, number of species, relative abundance of benthic species within the community, species diversity of benthic foraminifera estimated by the Shannon–Wiener index (H), Dominance (D), Fisher α index (α), Oxygen index and the Infauna/epifauna ratio. Shannon-Wiener index or Shannon diversity index (H) estimates the species diversity in the assemblage, as described in chapter 2.6 of this paper (after [34]). Dominance (D) reflects a distribution of a particular species in the assemblage, and the dominant species are those presented with >10% in the sample [34]. Fisher α index (α) shows the relation of the number of species to the number of the individuals, and to explore the number of species by each individual, a log series distribution is used [34]. This index is used for palaeoecological determinations, because specific values are characteristic for each environment. Depending on the index value range, we can analyse the palaeoecological changes in the environment.

The above mentioned analyses were enhanced by defining and comparing the benthic foraminiferal fauna from different localities conducting the Cluster Analysis and Non-metric Multidimensional Scaling by means of PAST (PAlaeontology STatistic) Program (https://folk.uio.no/ohammer/past/; e.g., [40]; Figure 4.1).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 April 2020 doi:10.20944/preprints202004.0471.v1

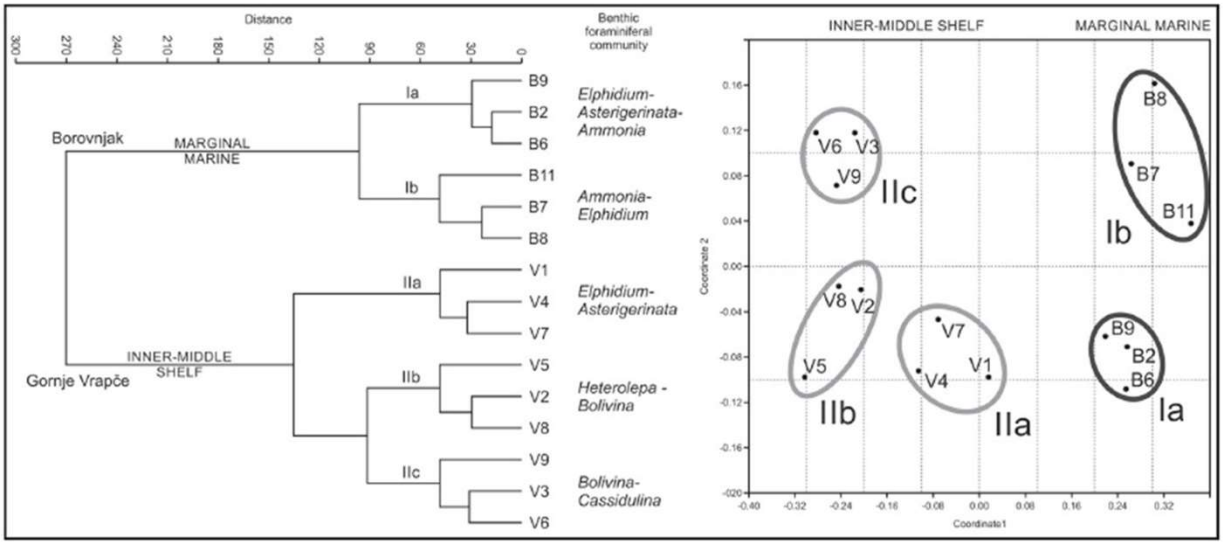


Figure 4.1. Example of statistical comparison of fauna from different localities using Cluster Analysis and Non-metric Multidimensional Scaling analyses (after [40]).

There are a number of other papers dealing with paleoenvironmental reconstructions of fossil communities based on the biometry of benthic foraminifera. Growth characteristics are used as a parameter for the palaeoecological and phylogenetical studies in the wider region (e.g., [43,44]). For example, [45] calibrated test flattening of the foraminifera species Heterostegina depressa as a bathymetric signal (Figure 4.2), using its growth functions and thickness. Similar study can be applied to the Miocene large nummulitids from Northern Croatia.

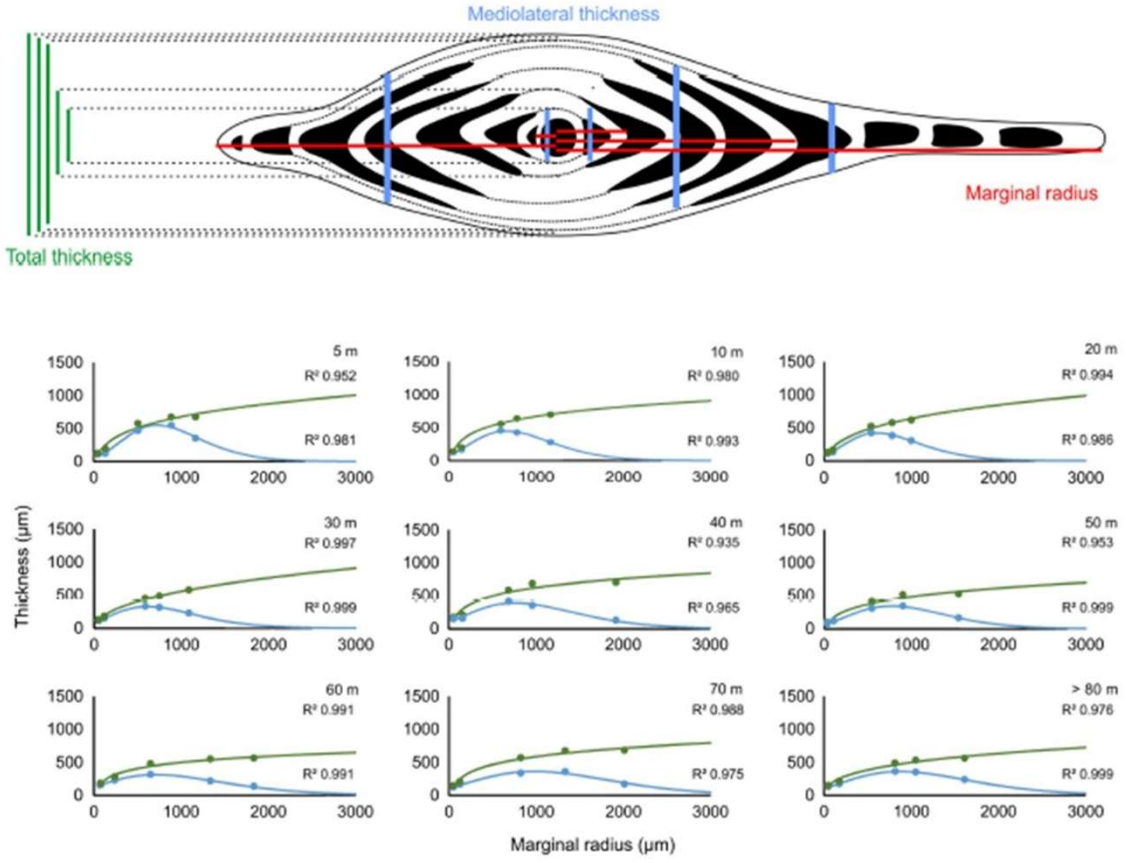


Figure 4.2. Example of using growth characteristics as an indicator on the bathymetry [45].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 April 2020 doi:10.20944/preprints202004.0471.v1

Biometric studies are also commonly applied in taxonomic study of mollusks. For example, a thorough revision based upon this method was made by [46] on gastropod families Conidae and Conorbidae from the Paratethys Sea. The authors measured several shell parameters (shell length, maximum diameter, aperture height, height of maximum diameter, spire angle, apertural length, the angle of the last whorl, length width ratio, relative diameter ratio, position of maximum diameter ratio, relative height of spire ratio, subsutural flexure, mean and standard deviation), analysed by Principal component analysis (PCA). Applying this analysis, authors compared similar species of Conidae and showed the separation of the species and morphospace occupied by genera (Figure 4.3).

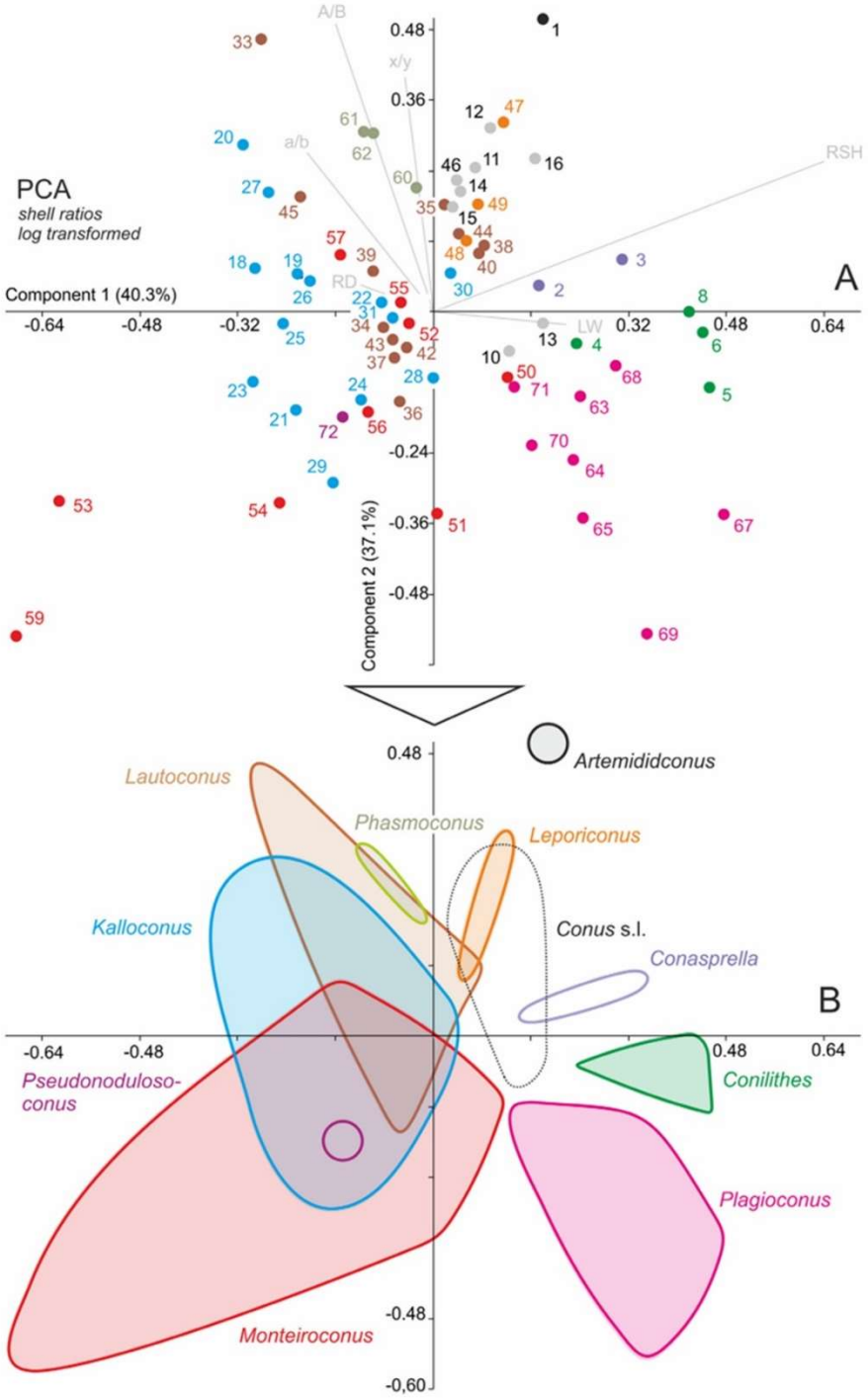


Figure 4.3. Separation of the species and morphospace occupied by genera as shown by the Principal component analysis (after [46]).

 doi:10.20944/preprints202004.0471.v1

Studies from Croatia were mostly focused on gastropods and scaphopods (e.g., [47,48,49]). Several parameters are measured (height and length of the shell, apical angle), defining the basic numerical data useful in species determination, morphometric characteristics of the group, correlation with recent species, comparison with other localities and palaeoecological interpretations.

[47] studied the Miocene planktic gastropods from northern Croatia and, based on the measured shell elements, compared their data with the available published measures of that fossil group found in the Miocene deposits of the neighbouring areas (Figure 4.4).

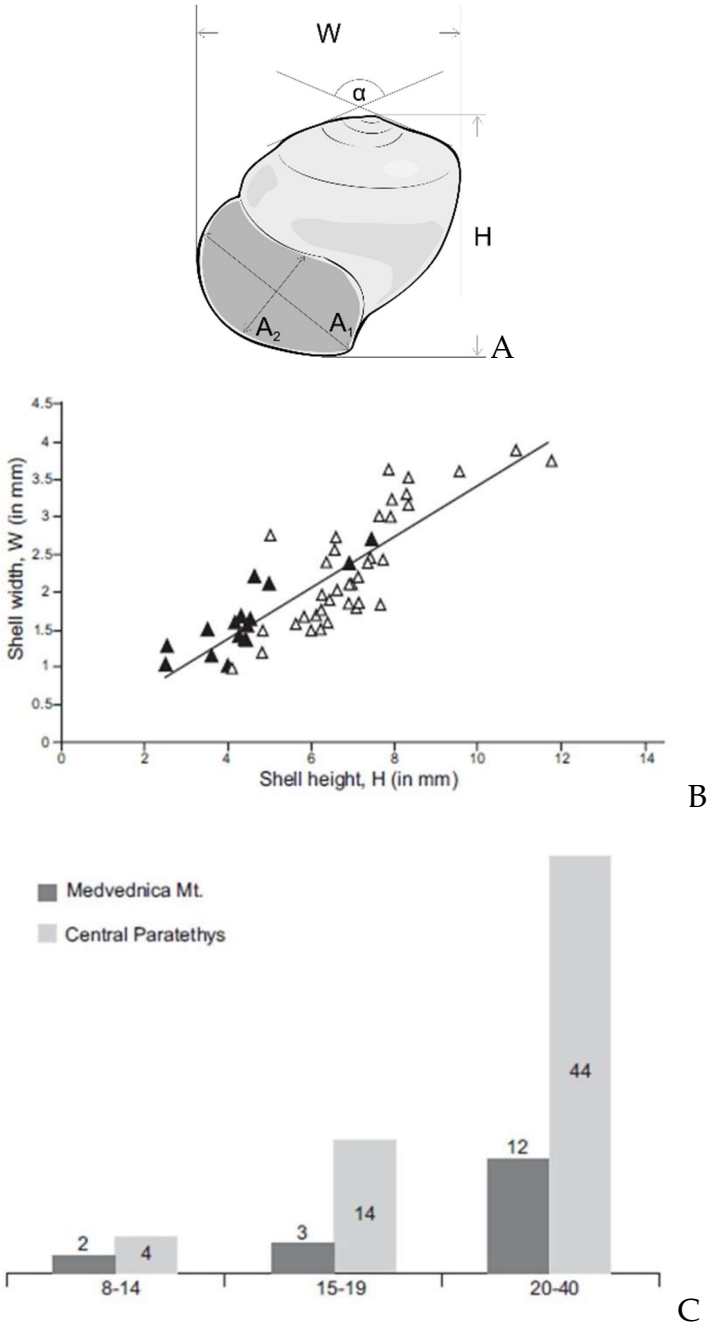


Figure 4.4. Morphometric characteristics and comparison of the planktic gastropods between different localities based on the measured morphometric elements of the shell (after [47]**)**. **A**: Measured parameters on the gastropod shell: H (height of the shell), W (width of the shell), α (apical angle), A_1 and A_2 (aperture diameters). **B.** Comparison of planktic gastropod from different areas (black and white triangles) based on the measured values of the shell height and width. **C.** Comparison of planktic gastropod species from different areas (dark and light grey columns) based on the measured values of the apical angle of the gastropod shells.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 April 2020 doi:10.20944/preprints202004.0471.v1

Bioerosive traces on skeletal remains, in most cases traces of predation, are also rather common topic in biostatistical analyses. Measures and shapes of the drill-holes can indicate the possible predator and help to get better insight in the predator-prey relationship, as described in numerous papers (e.g., [50] and references therein).

One more example of biostatistics analysis is presented in [51]. The authors measured the orientations of oyster attachments on ammonite shells, concluding that the oysters attached themselves while the ammonites were living (Figure 4.5). The results are helpful in palaeoecological studies of fauna in oxygen depleted environments.

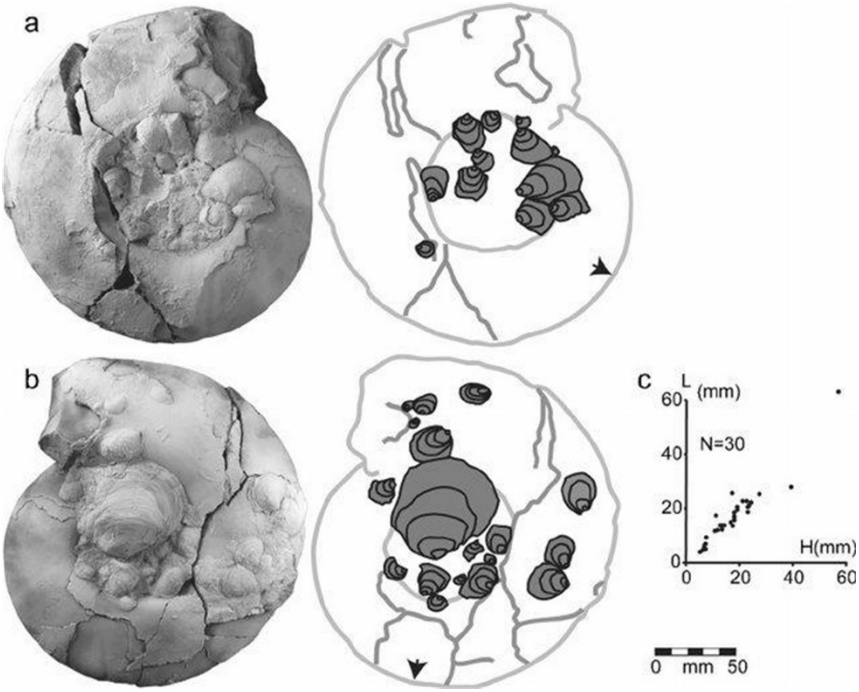


Figure 4.5. Analysis of oyster attachments positions on the ammonite shell [51].

Biostatistical analyses are also common in studies of vertebrates. In Croatia, research included the study of dinosaur footprints. The measured parameters of footprints included width and length of the footprint, and length of the second, third and fourth finger (e.g. [52] and references therein). These studies give insight on the dimensions of the animal (height) and type of their movement (walking) based on the calculations of the movement speed (e.g., [53,54,55] and references therein; Curman, 2017), which gives better insight into the biodynamic of the animal. [56] demonstrated the application of the Geometric Morphometrics as a tool for the shape analysis of the dinosaur footprints and trackways geometric differences (Figure 4.6).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 April 2020 doi:10.20944/preprints202004.0471.v1

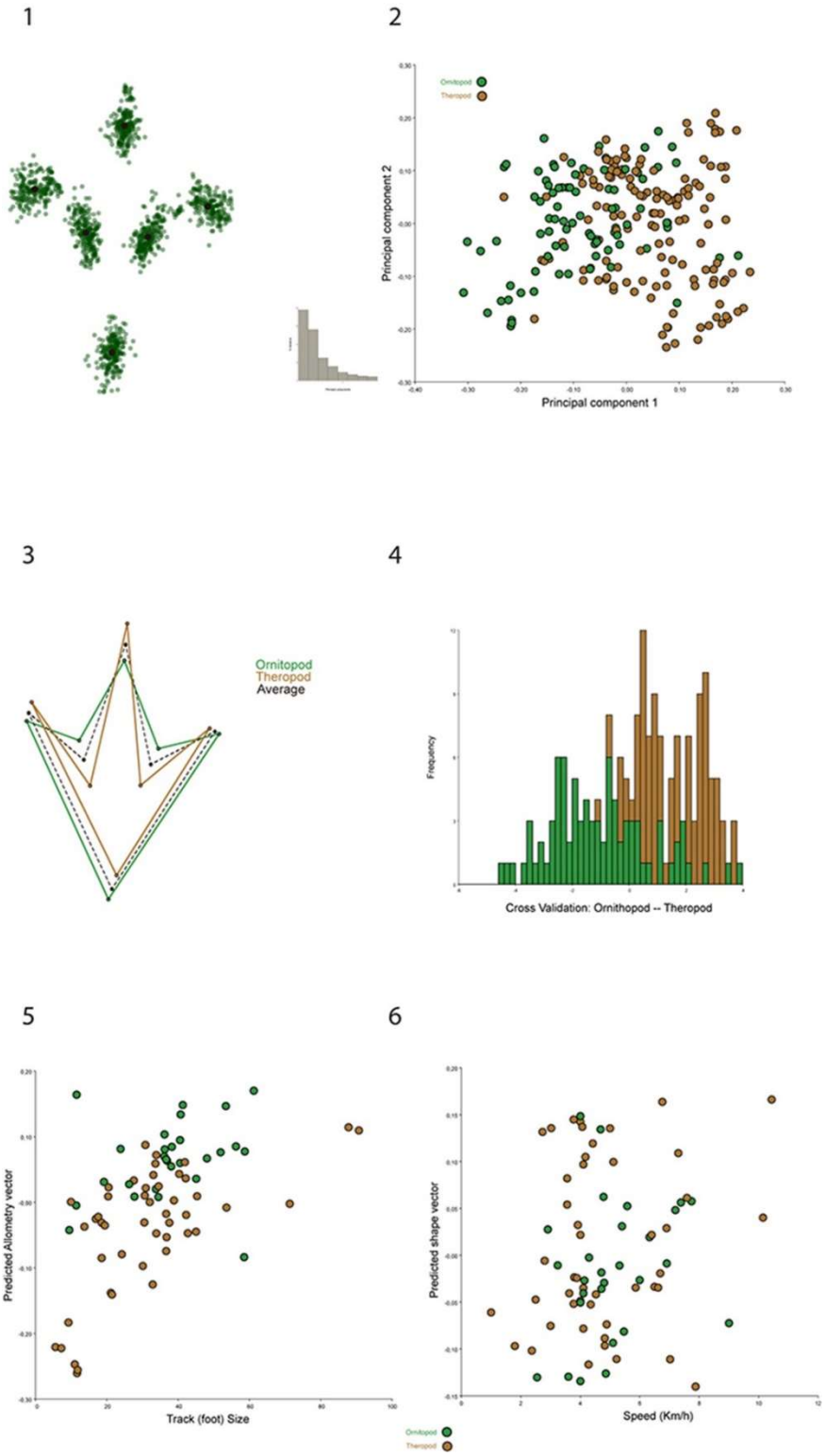


Figure 4.6. Application of the Geometric Morphometrics on the dinosaur footprints analysis [56].

We can conclude that biostatistical analyses generally occurred very early in paleontological studies. In Croatia, their number exhibits the pattern of periodicity. Basic numeric analyses of fossil assemblages were published in the mid-1990s, marking the first peak of biometric studies in Croatia. The second peak was during the first decade of the 2000s, with most research done on microfossils (foraminifers and accompanying ostracods). We can say that the third pulse is happening from 2016 onwards, considering the various groups of fossil biota.

 doi:10.20944/preprints202004.0471.v1

The analyses are mostly made to give more insight in the paleoecology of populations or fossil assemblages, and to help in the species determination. To present the analysed parameters, common statistical tools are used, mostly MS Office Excel and PAST (PALaeontological Statistics) programs.

5. Discussion and conclusion

The topic "Advances in Geosciences" is so broad that any paper publication would hardly cover only the small portion of significant milestones that shaped and led the progress in geosciences in general. The spectre of geosciences includes so many "fundamental" sciences that the ways of progress are very different, regarding data, methods and problems. Geosciences could be found in social (e.g., geography), technical (e.g., geodesy) and natural (e.g., geology) sciences. It is why the authors selected only one science (geology) with only one small segment (subsurface and surface geology) and tiny analytical, numerical methods (small datasets in mapping, larger in biostatistics). Even in such case, the presented cases are given mostly from the researching area where authors worked mostly in the last decade, i.e., original samples taken from the surface and subsurface of the Northern Croatia.

But both examples present the areas where, at least in Croatia, huge progresses are made and referencing methods for later researchers are set up. After more than 15 years of extensive and successful application of the different Kriging techniques in the subsurface mapping of the CPBS, the problem of small dataset where geostatistics cannot be reliable applied has been solved. The several simpler algorithms are tested, validated and recommended for application, namely Inverse Distance Weighting, Nearest Neighbourhood, Natural Neighbourhood and Modified Shepard Method. For such small datasets, the importance of mutual application for cross-validation and visual assessment had been stressed. Additionally, the Kriging was simultaneously tested as alternative or such algorithms, even in cases when variogram model cannot be calculated as reliable value, even as omnidirectional one. The extensive experiments with jack-knifing method have been done on variogram, creating artificial data from original dataset. In some cases, jack-knifed variograms gave competitive the Kriging results, but geostatistics was eliminated as the first choice in mapping analysis of small subsurface datasets.

Application of biostatistics has been presented on very different samples, collected from shallow subsurface or surface outcrops. Here the numerical values characterised not petrophysics, but morphological variables of different fossil groups (foraminifers, molluscs, vertebrates). In the presented examples on molluscs, the parameters like height and length of the shell are measured giving set of numerical values for determination of morphometrics and consequently species which gave more insight on Miocene palaeoecological conditions and environments in the Northern Croatia, especially during the existence of the Paratethys Sea. On larger scale, biostatistical analysis in Croatia helped to reconstruct the size and height of, e.g., dinosaur, using footprints measurements. Two outbursts of the Croatian biostatistical (biometric) analyses, presented with relevant publications, are noted. The first was in the mid-1990s, and the second was during the first decade of the 2000s, with most research done on microfossils (foraminifers and accompanying ostracods). Recently, the Croatian researchers entered in the third fruitful period from 2016 onwards, currently analysing the various marine fossil biota aiming to determine species and their paleoenvironments.

Both examples showed the useful application of geomathematical tools in geology. The first group showed how the small datasets (n<10 data) can be reliably mapped. The second presented how morphometric and surface features could be collected, numerically analysed and applied in paleoenvironmental reconstructions. The uncertainties, of course, remained due to data properties. The most problematics is clustering, which can be hardly handled when datasets are small and/or spatially noisy. In such cases, two crucial statistical properties cannot be reliably checked or established. That are proof of the normal distribution and statistical representativeness of dataset (mean, variance of population). However, the results, carefully validated and correlated with other, non-numerical (indicator, categorical) geological knowledge, are of great help in creating better geological models.

Author Contributions: Conceptualization, T.M., M.B., J.V. and J.S.; Formal analysis, M.B., J.I. Investigation, U.B., M.A.P.D., M.B..; Software, J.I. and M.B.; Supervision, T.M., J.V. and J.S.; Validation, T.M.; Visualization, M.B. and J.I.; Writing - original draft, T.M., M.B., J.I., J.S., U.B., M.A.P.D.; Writing - review & editing, T.M.

Acknowledgments: The authors are grateful for partial support from the project "Mathematical methods in geology IV" (led by T.M.). Funds were given from the University of Zagreb, for the year 2019. Also, the great help had been offered from Mr. Renato Drempetić in drawing the final figures for biostatistical examples.

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

References

1. Matheron, G. (1962): *Traité de géostatistique appliquée.* Tome 1, Editions Technip, 334, Paris.
2. Matheron, G. (1963): *Principles of geostatistics*. Econ. Geol., 58, 1246–1266.
3. Matheron, G. (1965): *Les Variables Régionalisées et leur Estimation*. Masson & Cie, 306, Paris.
4. 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.
5. Journel, A. G. & Huijbregts, C. J. (1978): *Mining Geostatistics*. Academic Press, 600 p., London.
6. Ripley, B. D. (1981): *Spatial Statistics.* Wiley & Sons Ltd., 272, New York.
7. Davis, J.C. & Sampson, R.J. (1973): *Statistics and Data Analysis in Geology.* John Wiley & Sons Inc., 564 p., New York.
8. Cressie, N. (1991): *Statistics for Spatial Data.* Wiley & Sons Ltd., 928 p., New York
9. Isaaks, E. & Srivastava, R. (1989): *An Introduction to Applied Geostatistics*. Oxford University Press Inc., 580 p., New York.
10. Jensen, J. L., Lake, L. W., Corbett, P. W. M. & Goggin, D. J. (2000): *Statistics for Petroleum Engineers and Geoscientists*. Prentice Hall PTR, 390 p., New Jersey.
11. Dubrule, O. (1998): *Geostatistics in Petroleum Geology.* AAPG Education Course Note, Series #38, AAPG and Geological Society Publishing House, 210 p., Tulsa.
12. Kelkar, M. & Perez, G. (2002): *Applied Geostatistics for Reservoir Characterization.* Society of Petroleum Engineers, 264 p., Richardson.
13. Sokal, R. R. & Rohlf, F. J. (2009): Introduction to Biostatistics. Second Edition. Dover Publications Inc., 1-363.
14. Lincoln Edwards, C. (1908): Biometry as a Method in Taxonomy. The American Naturalist, 42, 500, 537-540.
15. Billard, L. (2014): Sir Ronald A. Fisher and The International Biometric Society, Biometrics 70, 259–265, DOI: 10.1111/biom.12153
16. Hohn, M. E. (1988): *Geostatistics and Petroleum Geology*. Van Nostrand Reinhold, 400 p., New York.
17. 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.
18. Balić, D.; Velić, J.; Malvić, T. Selection of the most appropriate interpolation method for sandstone reservoirs in the Kloštar oil and gas field. *Geologia Croatica* **2008**, *61*, 27-35.
19. Medved, I., Pribičević, B., Medak, D. & Kuzmanić, I. Usporedba metoda interpolacije batimetrijskih mjerenja za praćenje promjena volumena jezera (Comparison of Interpolation Methods of Bathymetry Data Used for Monitoring of Lake Volume Change – in Croatian). *Geodetski list* **2010**, *2*, 71–86.
20. Ly, S.; Charles, C.; Degré, A. 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* **2011**, *15*, 2259-2274.
21. Husanović, E.; Malvić, T. Review of deterministic geostatistical mapping methods in Croatian hydrocarbon reservoirs and advantages of such approach. *Nafta* **2014**, *65*, 57-63.
22. Ivšinović, J. Deep mapping of hydrocarbon reservoirs in the case of a small number of data on the example of the Lower Pontian reservoirs of the western part of Sava Depression. Proceedings of the 2nd Croatian congress on geomathematics and geological terminology, 2018, Zagreb, Croatia, 6 October 2018, Malvić, T. (ed.); Velić, J. (ed.); Rajić, R. (ed.), University of Zagreb, Faculty of Mining, Geology and Petroleum Engineering; pp. 59-65.
23. Traversoni, L. Natural neighbour finite elements. *Transactions on Ecology and the Environment* **1994**, *8*, 291-297.

24. Boissonnat, J-D.; Cazals, F. Natural neighbor coordinates of points on a surface. *Computational Geometry* **2001**, *19*, 155–173.
25. Tsidaev, A. Parallel Algorithm for Natural Neighbor Interpolation. Abstract's proceedings of the 2nd Ural Workshop on Parallel, Distributed, and Cloud Computing for Young Scientists, Russia, 2016; Sozykin, A.; Akimova, E., Ustalov, D. Eds.; Pubisher: Ural-PDC, Yekaterinburg, Russia, 2016; 78-83, 83 p.
26. Shepard, D. A two-dimensional interpolation for irregularly spaced data function. Proceedings of the 1968 ACM national conference, New York, USA, 27-29 August 1968, Blue, R. B. (ed.); Rosenberg, A. M. (ed.), Association for Computing Machinery; pp. 517–523.
27. Franke, R.; Nielson, G. Smooth Interpolation of Large Sets of Scattered Data. *International Journal for Numerical Methods in Engineering* 1980, 15, 1691-1704.
28. Renka, R. J. Multivariate Interpolation of Large Sets of Scattered Data. ACM Transaction on Mathematical Software 1988, 14, 139-148.
29. Davis, B. Uses and Abuses of Cross Validation in Geostatistics. *Mathematical Geology* 1987, 19, 3, 241–248, Dordrecht.
30. Rodrıguez, J. D.; Perez, A.; Lozano, J. A. Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation. *IEEE transactions on pattern analysis and machine intelligence* **2010**, *3*, 32, 569-575.
31. Arlot, S.; Lerasle, M. Choice of V for V -Fold Cross-Validation in Least-Squares Density Estimation. *Journal of Machine Learning Research* **2016**, *17*, 1-50.
32. 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* **2019a,** *36*, 9, 1-24.
33. Malvić T.; Ivšinović J.; Velić J.; Rajić, R. Interpolation of Small Datasets in the Sandstone Hydrocarbon Reservoirs, Case Study of the Sava Depression, Croatia. *Geosciences* **2019b**, *9*, 5, 201.
34. Murray, J.W. (2006): Ecology and Applications of Benthic Foraminifera. Cambridge – University Press, Cambridge, 438 p. doi: 10.1017/CBO9780511535529
35. Malvić, T., Ivšinović, J., Velić, J., Sremac, J. & Barudžija, U. Application of the Modified Shepard's Method (MSM): A Case Study with the Interpolation of Neogene Reservoir Variables in Northern Croatia. *Stats* 2020, 3, 68-83.
36. Kochansky-Devidé, V. O fauni marinskog miocena i o tortonskom šliru Medvednice (Zagrebačke gore) [Ueber die Fauna des marinen Miozäns und über den Tortonischen "Schlier" von Medvednica (Zagreber Gebirge)]. *Geološki vjesnik* **1957**, X, 39-50.
37. Sokač, A. Panonska fauna ostrakoda Donjeg Selišta jugozapadno od Gline (Pannonische Ostrakodenfauna von Donje Selište südwestlich von Glina). *Geološki vjesnik* **1963**, 15, 2, 391-401. http://31.147.204.208/clanci/1963_Sokac_189.pdf
38. Prlj Šimić, N., Sremac, J. & Ćosović, V. Taxonomy and Biometry (Applied to the Eocene Corals from the Island of Krk – Croatia). 1st Croatian geological congress, 1995 Abstracts book, 495-498. http://geol.pmf.hr/~jsremac/radovi/znanstveni/1995_opat_koralji.pdf
39. Pezelj, Đ., Sremac, J. & Sokač, A. Palaeoecology of the Late Badenian foraminifera and ostracoda from the SW Central Paratethys (Medvednica Mt., Croatia). *Geologia Croatica* **2007**, 60, 2, 139-150.
40. Pezelj, Đ., Sremac, J. & Bermanec, V. Shallow-water benthic foraminiferal assemblages and their response to the palaeoenvironmental changes — example from the Middle Miocene of Medvednica Mt. (Croatia, Central Paratethys). *Geologica Carpathica* **2016**, 67, 4, 329-345.
41. Pezelj, Đ. & Sremac, J. Badenian Marginal Marine Environment in the Medvednica Mt. (Croatia). *Joannea Geol. Paläont.* **2007**, 9, 83-84.
42. Pezelj, Đ. & Drobnjak, L. Foraminifera-based estimation of water depth in epicontinental seas: Badenian deposits from Glavnica Gornja (Medvednica Mt., Croatia), Central Paratethys. *Geologia Croatica* **2019**, 72, 2, 93-100. doi: 10.4154/gc.2019.08
43. Hohenegger, J. Growth-invariant Meristic Characters Tools to Reveal Phylogenetic Relationships in Nummulitidae (Foraminifera). *Turkish Journal of Earth Sciences* **2011**, 20, 655-681. doi:10.3906/yer-0910-43
44. Hohenegger, J. & Torres-Silva, A.I. Growth-invariant and growth-independent characters in equatorial sections of *Heterostegina* shells relieve phylogenetic and paleobiogeographic interpretations. *Palaios* **2017**, 32, 30-43.
45. Eder, W., Hohenegger, J. & Antonino Briguglio (2018): Test flattening in the larger foraminifer *Heterostegina depressa*: predicting bathymetry from axial sections. Paleobiology, 44, 1, 76–88. DOI: 10.1017/pab.2017.24

46. Harzhauser, M. & Landau, B. A revision of the Neogene Conidae and Conorbidae (Gastropoda) of the Paratethys Sea. *Zootaxa* **2016**, 4210, 1, 001–178. http://doi.org/10.11646/zootaxa.4210.1.1
47. Bošnjak, M., Sremac, J., Vrsaljko, D., Aščić, Š. & Bosak, L. The Miocene "Pteropod event" in the SW part of the Central Paratethys (Medvednica Mt., northern Croatia). *Geologica Carpathica* **2017**, 68, 4, 329-349, with Erratum included. doi: 10.1515/geoca-2017-0023
48. Derežić, I., Bošnjak, M. & Sremac, J. Biostatistic analyses of newly found pteropods (Mollusca, Gastropoda) in the Middle Miocene (Badenian) deposits from the southeastern Medvednica Mt. (Northern Croatia). In: Malvić, T; Velić, J. & Rajić, R. (eds.): Mathematical methods and terminology in geology 2018, Zagreb, Faculty of Mining, Geology and Petroleum Engineering, University of Zagreb, 95–102.
49. Šeparović, A. (2019): Miocene deposits with scaphopods south from Veternica cave (Medvednica). Master Thesis. University of Zagreb, Faculty of Science, 66 p.
50. Kowalewski, M. (2002): The Fossil Record of Predation: An Overview of Analytical Methods.– In: Kowalewski, M. & Kelley, P.H. (eds.): The fossil Record of predation. Paleontological Special Papers, 8, Yale University, New Haven, 3–42.
51. Zell, P., Beckmann, S. & Stinnesbeck, W. (2014): *Liostrea roemeri* (Ostreida, Bivalvia) attached to Upper Jurassic ammonites of northeastern Mexico. Palaeobio Palaeoenv, 94, 439-451. DOI 10.1007/s12549-014-0154-z
52. Curman, D. (2017): Digital model analysis of theropod footprints from Solaris tracksite (Istria). Graduation Thesis. University of Zagreb, Faculty of Science, 1-72.
53. Dalla Vecchia, F.M. Remains of Sauropoda (Reptilia, Saurischia) in the Lower Cretaceous (Upper Hauterivian/Lower Barremian) limestones of SW Istria (Croatia). *Geologia Croatica* **1998**, 51, 2, 105–134.
54. Mezga, A., Cvetko Tešović, B., Bajraktarević, Z. & Bucković, D. (2007): A new dinosaur tracksite in the late Albian of Istria, Croatia. Rivista Italiana di Paleontologia e Stratigrafia, 113, 1, 139-148.
55. Mezga, A., Cvetko Tešović, B., Pretković, V., Jovanović, N. & Bajraktarević, Z. Dinosaur footprints in the Lower Hauterivian deposits of Palud Cove in Istria, Croatia. *Geologia Croatica* **2015**, 68, 2, 113-122. doi: 10.4154/gc.2015.08
56. Costa-Pérez, M., Joaquín Moratalla, J. & Marugán-Lobón, J. (2019): Studying bipedal dinosaur trackways using geometric morphometrics. Palaeontologia Electronica 22.3.pvc-3 1-13. https://doi.org/10.26879/980

Internet sources: https://folk.uio.no/ohammer/past/