GOSPODARKA SUROWCAMI MINERALNYMI – MINERAL RESOURCES MANAGEMENT

2023 Volume 39 Issue 3 Pages 149–172

DOI: 10.24425/gsm.2023.147555

JUSTYNA SOWIŃSKA-BOTOR[1], WOJCIECH MASTEJ[2], TOMASZ MAĆKOWSKI[3]

# Ranking of the utility of selected geostatistical interpolation methods in conditions of highly skewed seismic data distributions: a case study of the Baltic Basin (Poland)

## Introduction

Geostatistics is now widely applied in various areas of seismic research results (e.g. Chilès and Delfiner 2012; Parra and Emery 2013; Azevedo and Demyanov 2019; Chahooki et al. 2019; Aleardi et al. 2020; Dip et al. 2021). However, geostatistical interpolation of seismic data is often difficult due to large deviations from the normality of variable distributions. Frequently, even extremely positively skewed data distribution with a form similar to the log-normal distribution and outliers are common here. The type of distribution

✉ Corresponding Author: Justyna Sowińska-Botor; e-mail: justyna.sowinska@gmail.com

[1] AGH University of Science and Technology, Faculty of Geology, Geophysics and Environmental Protection, Kraków, Poland; e-mail: justyna.sowinska@gmail.com

[2] AGH University of Science and Technology, Faculty of Geology, Geophysics and Environmental Protection, Kraków, Poland; ORCID iD: 0000-0001-9147-8146; e-mail: wmastej@agh.edu.pl

[3] AGH University of Science and Technology, Faculty of Geology, Geophysics and Environmental Protection, Kraków, Poland; ORCID iD: 0000-0002-8366-0332; e-mail: mackowsk@agh.edu.pl

does not affect the form of kriging equations, but their non-normality causes an increase in interpolation errors, including the incorrect determination of local confidence intervals for the estimated values (e.g. Isaaks and Srivastava 1989). Since there is no clear solution to this problem, various attempts have been made to overcome it. It is generally known, as confirmed by Vann and Guibal (Vann and Guibal 1998), that histogram deskewing and a decrease in variability can be achieved by increasing the support, for example, by using block kriging instead of point kriging. However, this will not be the subject of this study.

Kerry and Oliver (Kerry and Oliver 2007a) has suggested checking for outliers. If their appearance is not the result of errors (the erroneous should of course be rejected), then it should be checked whether their presence and strong skewness of distribution is not caused by mixing different data populations (e.g., Kokesz 2006; Kerry and Oliver 2007a; Nieć and Mucha 2007). If the populations can be separated in terms of area, then a good, though labor-intensive and with no guarantee of success is the division of the area into quasi-homogeneous sub-areas within which the distributions are already more similar to normal. Another, easier approach to addressing these problems (Kerry and Oliver 2007a) is to use variogram estimators that are more robust with regard to deviation from normality than the classic Matheron estimator. Kerry and Oliver (Kerry and Oliver 2007b) give an overview of such estimators. However, Lark (Lark 2008) stated that these estimators are useful only when the skewness of data is caused by the presence of outliers. This does not mean, however, that you can skip editing data and sometimes their transformation before calculating the variogram, which is also necessary when the skewness has a general cause, i.e. it is caused not only by outliers.

The Matheron estimator is also not robust to non-ergodicity (preferentially space-distributed data). It is widely believed that in cases of such data, which are too strongly skewed, non-ergodic estimators, as proposed by Isaaks and Srivastava (Isaaks and Srivastava 1988) and Srivastava and Parker (Srivastava and Parker 1989), should be used.

There have been signals that such estimators do not have any advantage over traditional estimators, and in the case of isotropic data, they are equivalent (Curriero et al. 2002), but a number of works confirm their usefulness (e.g. Rossi et al. 1992; Nieć and Mucha 2007; Kokesz 2010). It has also long been obvious (e.g. Kerry and Oliver 2007a) that Matheron's estimator is not robust with regard to strongly skewed data, regardless of the cause of this skewness. It is not obvious, however, whether it will be robust for data that is admittedly standardized but not entirely ergodic. In this study, the somewhat forgotten Inverted Covariance Variogram estimator, robust with regard to non-ergodicity of data, called InvCov (Englund and Sparks 1999) was tested. Its utility postulates, for example, Nieć and Mucha (Nieć and Mucha 2007).

If the skewness of the distribution is not caused by outliers, then "standard best practice" (Kerry and Oliver 2007a) is the transformation of primary data into a form close to a normal distribution. Usually, this is logarithmization (Journel 1980; Saito and Goovaerts 2000) or normalization (Goovaerts 1997; Deutsch and Journel 1997). In the case of logarithmization,

however, there are concerns about maintaining non-biasness. It may be lost by even a slight deviation from the logonormality of the primary data of Vann and Guibal (Vann and Guibal 1998) or during the inverse transformation of kriging estimates (Roth 1998; Clark 1999; Yamamoto 2007). Another transformation type that is often used is indicator transformation. Such a transformation, despite causing the loss of some information, allows the data to be used without any distribution requirements. Disjunctive or indicator kriging is usually used to interpolate such data (Journel and Deutsch 1997). The use of such transformations causes the interpolators to become non-linear. Vann and Guibal (Vann and Guibal 1998) assume that in the case of strongly skewed distributions, no linear interpolators should be used.

The purpose of this work was to checking the usefulness of some popular and low-labor methods using data transformation: logarithmization, normalization in the variant – anamorphosis and indicator transformation in combination with a simplified version of the median indicator kriging. The previously mentioned InvCov spatial variability estimator was also tested.

## 1. Geological setting

The study area is located on the SW slope of the East European Craton (EEC), where the Baltic Basin can be distinguished in the northern part of the Polish sector (Figure 1). The EEC margin is limited by the Teisseyre-Tornquist Zone (TTZ) (e.g. Guterch et al. 2010; Mazur et al. 2018a, b). In NE Poland, which is the study area of this work, Neoproterozoic, Palaeozoic, Mesozoic and Cainozoic strata occur above the Precambrian crystalline basement (e.g., Guterch et al. 2010; Mazur et al. 2018a, b). In the Baltic Basin, middle Cambrian sandstones represent major reservoirs of conventional oil and gas deposits, which are exploited in the central offshore part of the Polish Baltic Basin (as summarized by Pletsch et al. 2010). Therefore, in the study area, geophysical methods, including seismic approaches, have recently been extensively applied (e.g. Kasperska et al. 2019; Domagała et al. 2021; Kwietniak et al. 2021) for both conventional and unconventional petroleum exploration (e.g. Botor et al. 2019a, b).

The study area is located on the Quaternary glacial plateau, 30 km to the SW of Gdańsk, near the village of Wysin (Figures 1 and 2). Cenozoic strata, forming the uppermost part of the sedimentary section, are composed of varied, poorly consolidated deposits of thicknesses of usually only a few hundred meters. In the Quaternary section, fluvioglacial sediments and boulder clays dominate in the study area. There are practically only outcrops of sediments of the upper glaciation of the Vistula (Baltic) on the surface. The first surface layer covers the soil (about 0.5 m in thickness) and the upper part of the order of boulder clay or sands and glacial gravels, and in rare occasions, both of these sediment variability.

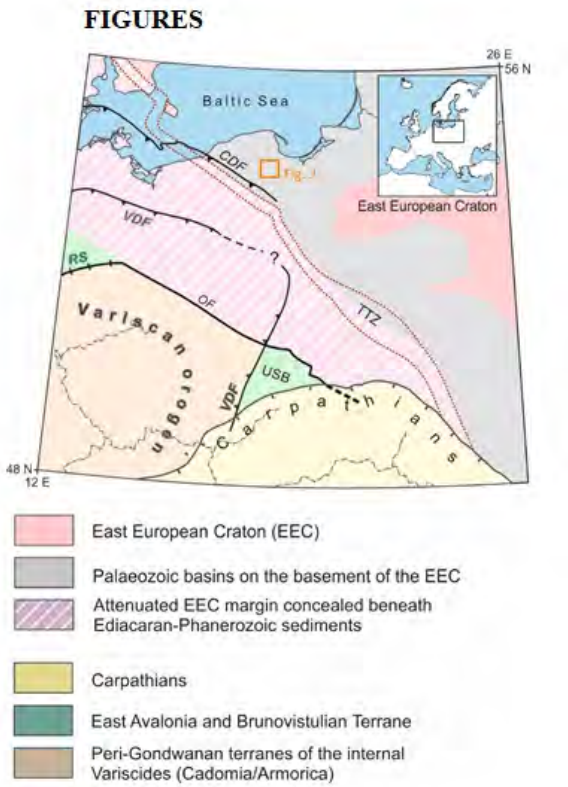


Fig. 1. Simplified tectonic map of Central Europe (modified after Mazur et al. 2018b).
The boundary of the Variscan orogen, main structural elements and possible terrane boundaries are also shown.
CDF – Caledonian Deformation Front; OF – Odra Fault; RS – Rheic Suture;
VDF – Variscan Deformation Front; TTZ – Teisseyre-Tornquist Zone; USB – Upper Silesia Block

Rys. 1. Uproszczona mapa tektoniczna Europy Środkowej.
Pokazano również granicę orogenu waryscyjskiego, główne elementy strukturalne i możliwe granice tektoniczne
CDF – Kaledoński Front Deformacji; OF – uskok Odry; RS – szew Rheic;
VDF – waryscyjski przód deformacji; TTZ – strefa Teisseyre-Tornquista; USB – Blok Górnośląski

## 2. Data

The study material is data on the thickness of the close-to-surface layer (variable H1_80) and the velocity of seismic waves inside it (variable V1_80). The data was obtained from eighty measurement points, located within an approximately regular square grid with a spacing of 2–5 km over an area of around 20 × 20 km (Figure 2). Data was collected by micro-profiling upholes (sixty points), using a seismic wave source on the surface and receivers (uphole probes) placed in shallow (40–60 m) holes and using the method of shallow refraction measurement (twenty points). The datasets H1_80 and V_80 were divided into

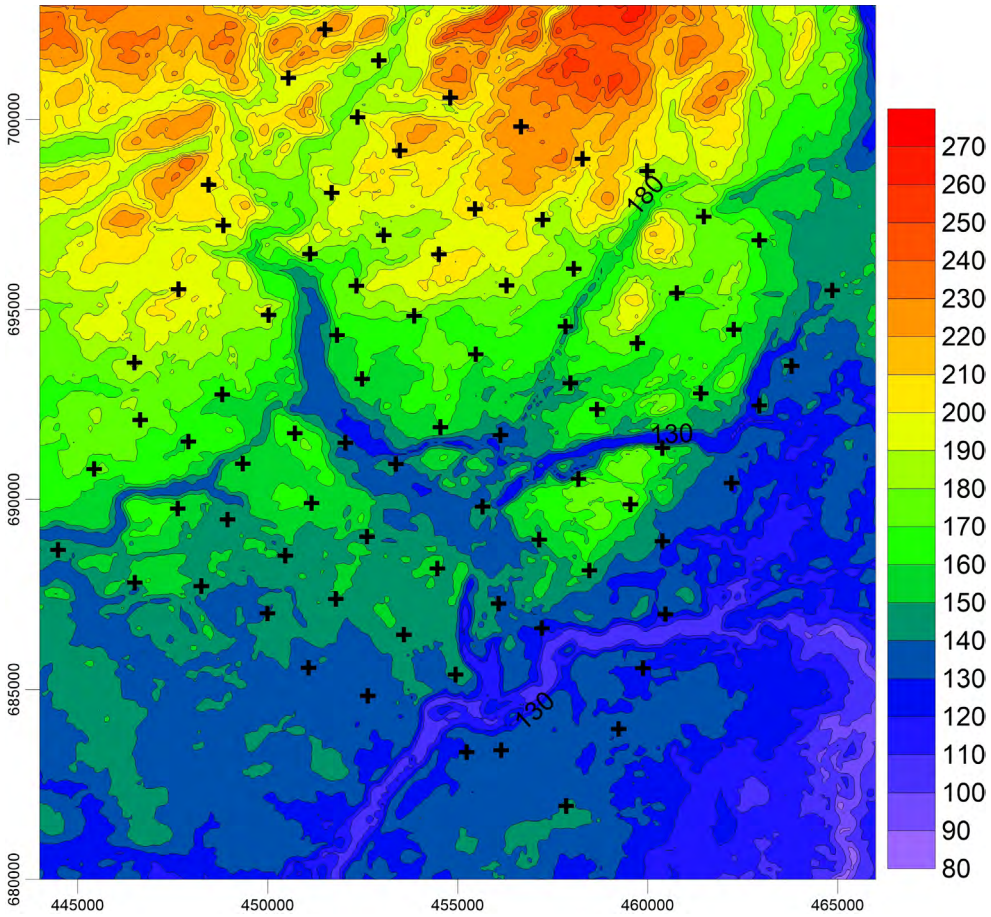


Fig. 2. Measuring points (uphole) on the background of topographic map of the study area. The points where the shallowest surface layer covers soil and boulder clays are marked in black the soil and fluvioglacial sediments are magenta

Rys. 2. Punkty pomiarowe (typu *uphole*) na tle ukształtowania morfologii obszaru badań. Na czarno zaznaczono punkty, w których najpłytsza warstwa przypowierzchniowa obejmuje glebę i gliny zwałowe, na różowo – glebę i utwory fluwioglacjalne

two parts corresponding to the two different methods of their measurement: H1_60_uph and V1_60_uph (micro-profile upholes), and H1_20_ref and V1_20_ref (refraction method). The next variable (H1_76) was created by removing four outliers from the dataset H1_80, precisely from its part that was obtained through refraction measurement. This enabled the definition of another dataset obtained through the refraction method: H1_16_ref. The criterion for identifying outliers was subjective. Layer thickness data (variables H1_80, H1_67 and its parts: H1_60_uph, H1_20_ref and H1_16_ref) derived from low velocity zone micro-profiling in upholes is characterized by lower variability compared to data from shallow refraction

measurements (Figure 3a). Furthermore, the H1_76 data has lower outlier values. In turn, data on the velocity of seismic waves (variable V1_80) from these two sources have similar variability but slightly different average values (Figure 3b). Therefore, taking into account refraction data increases the variability, asymmetry and kurtosis of distributions (Figure 4), which is unfavorable during interpolation by kriging methods. However, this publication combines this data to test the usefulness of different variants of interpolation methodology in difficult conditions. For this reason, the H1_80, H1_76 and the V1_80 datasets were used

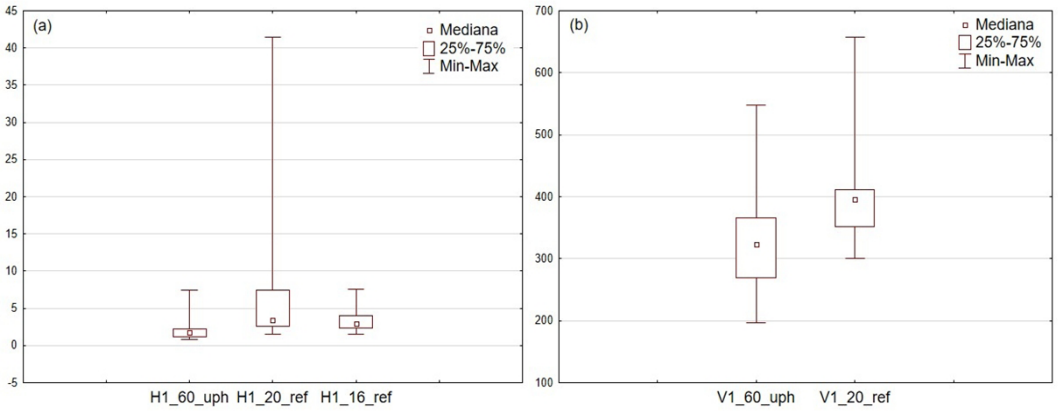


Fig. 3. Comparison of data from seismic micro-profiling (uph) and shallow refraction measurements (ref) for variables: a) H1_80 and H1_76, b) V1_80

Rys. 3. Porównanie danych z mikroprofilowań sejsmicznych (uph) i z pomiarów płytkiej refrakcji (ref) dla zmiennych (a) H1_80 i H1_76 oraz dla zmiennej (b) V1_80

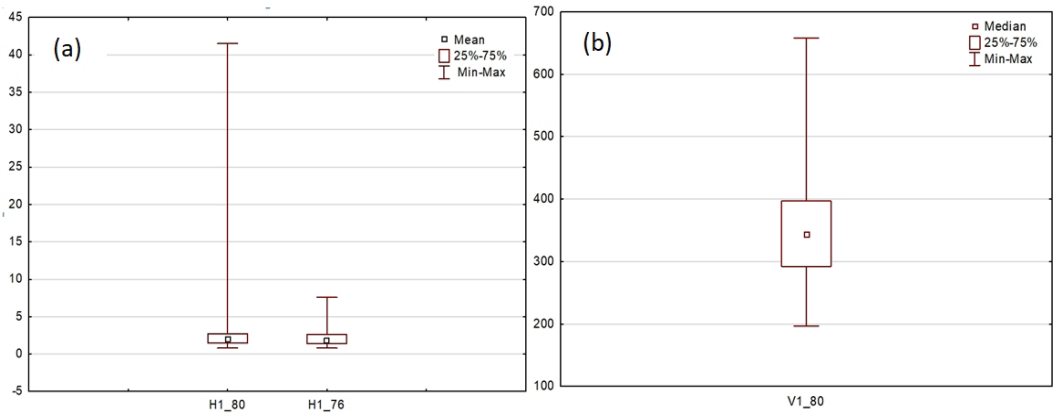


Fig. 4. Box and whisker plots of variables: (a) H1_80 and H1_76, (b) V1_80

Rys. 4. Wykresy typu pudełko-wąsy zmiennych: (a) H1_80 and H1_76, (b) V1_80

to calculate the static correction, which included the effect of terrain topography and low-velocity surface zone in the seismic survey. Ultimately, this survey was used to search for the shale gas deposit in the Silurian-Ordovician Baltic basin (Stara Kiszewa concession).

Significant lithological variability, especially among the boulder clays, had a significant impact on the variability, skewness and tailedness of the distributions of the studied variables (high values of the coefficient of variation V, asymmetry A and kurtosis K; see Table 1, Figure 4). These statistical parameters are clearly correlated with each other. Their highest values appear in the case of the H1_80 variable, and after removing the four outliers with the highest values (over 8 m), they decrease significantly (the H1_76 variable), although the distribution still has asymmetry and three outliers in the range of 7 to 8 m. One can notice that the distribution for H1_76 would still remain skewed even if these three outliers were removed. Both variables (H1_80 and H1_76) have distributions similar to lognormal distributions, while the variable V1-80 is characterized by almost normal distribution, with one outlier not too distant from the other values. In the conducted tests, the latter data was considered as "ideal" data, which were expected to have a high kriging estimation quality, regardless of the estimation procedure. The variable H1_80 has the highest values of coefficients V, A and K, the H1_76 – intermediate, and the V1_80 – the lowest. This directly impacts into the degree of difficulty in estimation – the greatest difficulties can be expected in the case of the variable H1_80, intermediate in the case of H1_76 and the lowest for V1_80. A facilitation in the research was the negligibly small trend and anisotropy observed in these data sets.

Table 1. Basic statistical parameters of variables

Tabela 1. Podstawowe parametry statystyczne zmiennych

| Variable | $n$ | $\bar{z}$ | Minium | Maxium | $s$ | $V$ (%) | $A$ | $K$ | No. of outliers |
|---|---|---|---|---|---|---|---|---|---|
| H1_80 | 80 | 3.09 | 0.80 | 41.50 | 4.86 | 158 | 6.57 | 50.40 | 7 |
| H1_76 | 76 | 2.27 | 0.80 | 7.60 | 1.44 | 63 | 2.16 | 5.32 | 3 |
| V1_80 | 80 | 349 | 196 | 658 | 83 | 24 | 0.93 | 1.62 | 1 |

Parameters for variables H1_80 and V1_80 were calculated for 60 upholes of seismic micro-profiling and 20 measurements of shallow refraction, but for variable H1_76 – for 60 upholes and 16 refraction points.

$n$ – sample size, $z_i$ – ith value of the variable, $\bar{z}$ – average value of the variable, $s$ – standard deviation, $V$ – coefficient of variation, $A$ – asymmetry coefficient, $K$ – kurtosis. The asymmetry coefficient $A$ and kurtosis $K$ were calculated from the formulas:

$$A = \frac{n\sum_{i=1}^{n}(z_i - \bar{z})^3}{(n-1)(n-2)s^3}; \qquad K = \frac{n(n+1)}{(n-1)(n-2)(n-3)}\sum_{i=1}^{n}\left(\frac{z_i - \bar{z}}{s}\right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)}$$

# 3. Methods

## 3.1. Method of kriging estimation validation

The leave-one-out-cross-validation (LOOCV) method was used to validate the quality of kriging estimation, as it is known (Jain et al. 2000) that it guarantees unbiased validation. For this reason, it is widely used in geostatistics to check the correct selection of the spatial variability model and parameters of the mobile neighborhood used in kriging procedures. If we have sampling points, the LOOCV procedure consists in the sequential calculation of kriging estimates at each of these points based on information from the remaining points, and then comparing the calculated kriging estimations with the observed, "real" values obtained during sampling. Estimation quality measures in the LOOCV tests can be found, for example, in Isaaks and Srivastava (Isaaks and Srivastava 1989), Johnnston et al. (Johnnston et al. 2001), Atkinson and Lloyd (Atkinson and Lloyd 2010), Li and Zhao (Li and Zhao 2009), Chilès and Delfiner (Chilès and Delfiner 2012). In the case of one computational procedure (median indicator kriging – see Section 3.2), due to limitations of the Isatis software used, the hold-out method (HO) was employed instead of the LOOCV. This involved calculating kriging estimates and kriging standard deviations for an additional twenty-nine points within the study area, where the "true" values of variables H1 and V1 were known. These calculations were based solely on the previously described datasets with eighty and seventy-six elements.

In the paper, several parameters listed in Tables 2, 3 and 4 (left, outermost column) were used for validating the quality of estimation. The first of these was standardized kriging error; at point its value will be:

$$\varepsilon_{st}(x_i) = \frac{\left[\hat{z}(x_i) - z(x_i)\right]}{\hat{\sigma}_K(x_i)} \tag{1}$$

- $\hat{z}(x_i)$ and $z(x_i)$ – estimate and observed value at this point $x_i$;
- $\hat{\sigma}_K(x_i)$ – kriging standard deviation at this point.

Probability distribution of this error should be close to standardized normal distribution. The best measures of estimation quality are mean value of this error – MSE (mean standardized error), showing bias of the kriging estimator and standard deviation the error – RMSSE (root mean square standardized error). A RMSSE value greater than 1 indicates an underestimation and a lower value indicates an overestimation. These measures are given by formulas (e.g. Johnston et al. 2001):

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\varepsilon_{st}(x_i) \tag{2}$$

$$RMSSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{st}\left(x_i\right)^2} \tag{3}$$

It is assumed that the MSE and the RMSSE should not deviate by more than 0.05 from zero and one, respectively.

Another kind of the error measure used here was the average absolute relative error:

$$\overline{\varepsilon}_{Arel} = \frac{1}{N}\sum_{i=1}^{N}\left|\varepsilon_{rel}(x_i)\right| \tag{4}$$

where:

$$\varepsilon_{rel}(x_i) = \frac{\left[\hat{z}(x_i) - z(x_i)\right]}{z(x_i)} \cdot 100\% \tag{5}$$

This measure can only be calculated in locations where observed values occur. To estimate errors at the nodes of the interpolation grid, the predicted average absolute error is commonly used. In order to compare it with the error $\overline{\varepsilon}_{Arel}$, it was calculated in this publication at the locations where the data are located:

$$\overline{\varepsilon}_{progArel} = \frac{1}{N}\sum_{i=1}^{N}\left|\varepsilon_{progrel}(x_i)\right| \tag{6}$$

where:

$$\varepsilon_{progrel}(x_i) = \frac{\hat{\sigma}_K(x_i)}{\hat{z}(x_i)} \cdot 100\% \tag{7}$$

In addition to estimation error measures, the correlation coefficients $R_1$ (see Tables 2, 3 and 4) between the estimates $\hat{z}(x_i)$ and the observed values $z(x_i)$ were calculated. Such a correlation is very desirable (Isaaks and Srivastava 1989), where it should be a linear correlation, and the regression line should have a slope of 45° and pass as close as possible to the beginning of the coordinate system. However, a correlation between estimates and errors is undesirable. To assess the strength of this correlation, the correlation coefficient $R_2$ was calculated for variables $\hat{z}$ vs. $\varepsilon_{st}$, i.e. $\hat{z}$ vs. $\left(\hat{z}-z\right)/\hat{\sigma}_K$, as well as the $R_3$ coefficient for variables $\hat{z}$ vs. $(\hat{z}-z)$.

### 3.2. Geostatistical procedures

It has been known for many years (e.g. Clark 1999) that geostatistical methods do not manage well with data that deviates strongly from normality. In order to test the resistance of selected methods to such data character, this paper compares several geostatistical procedures, which are combinations of different variants of data transformation, variogram estimation and kriging techniques. These procedures were divided into four groups. The first group, A, consists of two procedures using ordinary kriging (OK), without data transformation. It was anticipated that these procedures would not perform well in the interpolation of data of large deviations from the normality of variable distributions. Therefore, it was initially assumed that they would only serve as a reference point for more advanced procedures. They differ in the type of spatial variability estimator. The Matheron's estimator was used in the procedure A1 (Chilès and Delfiner 2012):

$$\hat{\gamma}_M(h) = \frac{1}{2n_h}\sum_{i=1}^{n_h}\left[z(x_i+h) - z(x_i)\right]^2 \tag{8}$$

- $z(x_i)$ and $z(x_i + h)$ are the values of the regionalized variable at all measurement points separated by the shift vector $h$; $n_h$ is the number of pairs of measurement points distant by the shift vector $h$. Due to the irregularity in the data distribution, the distance $h$ is not constant but takes values from a certain range; the calculated variogram value is assigned to the average value of $h$ in this range.

In the procedure A2, the InCov estimator was used (Isaaks and Srivastava 1988; Srivastava and Parker 1988; Englund and Sparks 1999), given by the formula:

$$\hat{\gamma}_{IC}(h) = s^2 - C(h) = s^2 - \frac{1}{n_h}\sum_{i=1}^{n_h}\left[z(x_i)z(x_i+h) - \overline{z}_i\overline{z}_{i+h}\right] \tag{9}$$

- $s^2$ – sample variance;
  $C(h)$ – autocovariance;
  $z(x_i)$ $z(x_i)$, $z(x_i + h)$ – variable values at points distant $h$;
  $\overline{z}_i$, $\overline{z}_{i+h}$ – mean values of the variable in the variogram points: $i$ and: $i + h$.

The second group of estimation procedures (B) is characterized by the use of the Matheron estimator for spatial variability and lognormal kriging (LGK) in the variant with ordinary kriging (OK, procedure B1) or with simple kriging (SK, procedure B2), but with previous logarithmization of data. This was reasoned by the fact that the variables H1_80 and H1_76 had distributions close to lognormal, so after logarithmization, the distributions should be

almost standardized normal in character. This standardized form of distributions enables use of not only ordinary but also simple kriging. During the application of ordinary kriging, it is assumed that the global mean is not known. However, in the case of using simple kriging (procedures from groups B and C), a global mean is imposed. Since simple kriging was used for data with standardized normal distributions, a global mean of zero was assumed. Logarithmic transformation took the form (Bleinès et al. 2014):

$$y = \ln\left(z+1\right) \tag{10}$$

Inverse transformation of kriging estimates was performed in variant B2, after applying simple kriging:

$$\hat{z} = \exp\left(\hat{y} + \frac{1}{2}\hat{\sigma}^2_{Ky}\right) - 1 \tag{11}$$

↳ $\hat{\sigma}^2_{Ky}$ – variance of log-data kriging.

The third group of procedures (C) differed from the B procedures in the type of data transformation. Instead of logarithmization, normalization using Gaussian anamorphosis was used (Bleinès et al. 2014). This anamorphosis is a function transforming a variable with normal distribution into a variable with any distribution, using Hermite polynomials. Ordinary kriging (OK) was used in procedure C1 and simple kriging (SK) was used in C2. In both of these procedures, the Matheron spatial variability estimator was used, while in procedure C3, the InvCov estimator was combined with simple kriging.

The fourth group (D) is only one indicator kriging procedure in a simplified version, i.e. median indicator kriging (MIK) (Isaack and Srivastava 1989; Deutsch and Journel 1997; Badel et al. 2011). Spatial variability was assessed using the Matheron estimator. Indicator Kriging is particularly recommended for estimating parameters characterized by high or extremely high variability, strongly asymmetrical probability distributions and by the occurrence of anomalous values. In the paper, the value of the parameter measured at the *i*-th pointis subject to a non-linear, binary transformation according to the following principle:

$$\begin{aligned} I_l\left(c_l\right) &= 1 \quad \text{when} \quad z_i \geq c_l \\ I_l\left(c_l\right) &= 0 \quad \text{when} \quad z_i < c_l \end{aligned} \tag{12}$$

↳ $c_l$ is an $l$ – cutoff value.

The obtained values for the variable H1_80 required correction of order relation deviations (Deutsch and Journel 1997). These authors recommend the MIK procedure in a situation where it can be assumed that if the full indicator kriging procedure (IK) was used, the variograms for *K* cutoff values would be similar. This would, however, increase labor inten-

sity to a level similar to that in the IK procedure, which would remove the sense of using the MIK procedure. Thus, without any initial assumptions, the MIK procedure calculates the variogram for only one cut-off value – the median. In general, eight cutoffs were used for the H1_80 and H1_76 datasets, while nine cutoffs were used for the V1_80 dataset. Then, using ordinary kriging applied to the transformed data, probabilities are calculated for the values of the variable of interest to fall within the intervals determined by successive cutoffs. This, in turn, enables interpolation, i.e. estimating the mean at a specific point (block).

## 4. Results and discussion

Each variant of the estimation procedure from the group A began with the calculation of an experimental variogram to which the spatial variation model was fitted. A movable window (neighborhood) in the shape of a circle with a radius of 4000 m, divided into four sectors was used in the all procedures. In the remaining groups (B, C, D), this stage was preceded by data transformation. Example models are shown in Figure 5, and model parameters are presented in Tables 2, 3 and 4. All the models used were spherical. No trends or anisotropy were observed. The lack of trends is evidenced by bounded variograms (Figure 5), and the lack of anisotropy by slight differences between directional variograms (the ratio of variability in the direction of maximum and minimum variability does not exceed 2; not reported here). Spherical isotropic models were fitted for the all variables (H1_80, H1_76, V1_80).

The estimation procedures were evaluated in terms of their suitability for use for highly asymmetric data distributions with outliers. As the evaluation criteria were preliminarily recognized estimation errors measures (Tables 2, 3 and 4): MSE, RMSSE, $\overline{\varepsilon}_{Arel}$ and $\overline{\varepsilon}_{progArel}$, and correlation coefficients $\hat{z}$ vs. $z(R_1)$, $\hat{z}$ vs. $\varepsilon_{st}$, i.e. $\hat{z}$ vs. $(\hat{z}-z)/\hat{\sigma}_K(R_2)$ and $\hat{z}$ vs. $(z-\hat{z})\,(R_3)$. Due to strong deviations from the normal distribution of the variables H1_80 and H1_76, credible correlation coefficients are not Pearson's linear correlation measures but Spearman's rank correlation measures (Tables 2 and 3). Pearson's coefficients were calculated for the variable V1_80 with a distribution close to normal (Table 4). Additionally, for this variable, Spearman's correlation coefficients were also calculated, enabling their comparison with the values of such coefficients in the H1-80 and H1-76 datasets (Table 7).

From the above set of criteria, it was necessary to remove those that did not sufficiently differentiate the estimation results. Therefore, MSE and RMSSE errors were removed, because deviations of from zero and one, respectively, were acceptable in each of the variants. Although these errors for the procedure D are significantly higher (Tables 2, 3 and 4), this may be due to the use of the HO method instead of the LOOCV for validation. The $\overline{\varepsilon}_{progArel}$ measure of error was also recognized to be of little use, as it can be considered rather a rough estimate of estimation error (e.g. Wasilewska and Mucha 2005); moreover, it cannot be calculated for variants A2 and C1, C2, C3.

All correlation coefficients also proved to be of little use. The desired, weak correlations of estimates with errors: $\hat{z}$ vs. $\varepsilon_{st}$ and $\hat{z}$ vs. $(z-\hat{z})$ were similar in each of the variants.

Table 2. Variable H1_80, parameters of variogram models and measures of estimation quality for different variants of the estimation method

Tabela 2. Parametry modelu wariogramu oraz miary jakości estymacji dla różnych metod estymacji zmiennej H1_80

| Parameters of variogram models and measures of estimation quality | | Variant of computing procedure | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | (A1) Matheron + OK | (A2) InvCov +OK | (B1) Matheron + + LNK (OK) | (B2) Matheron + + LNK (SK) | (C1) Matheron + + Norm. + OK | (C2) Matheron + + Norm. + SK | (C3) InvCov + + Norm. + SK | (D) Matheron + + MIK |
| Parameters of variogram models | $C_0$ | 5.0 | 20.0 | 0.090 | 0.090 | 0.400 | 0.400 | 0.700 | 0.203 |
| | $C$ | 17.0 | 4.0 | 0.150 | 0.150 | 0.600 | 0.600 | 0.300 | 0.033 |
| | $L$ | 0.23 | 0.83 | 0.38 | 0.38 | 0.40 | 0.40 | 0.70 | 0.86 |
| | $a$ (m) | 1750 | 4000 | 2000 | 2000 | 4000 | 4000 | 7000 | 7986 |
| Errors of kriging estimation | *MSE* | -0.03 | -0.03 | -0.01 | (0.00) | (0.00) | (-0.02) | (-0.01) | 0.14 |
| | *RMSSE* | 1.00 | 1.00 | 1.00 | (1.04) | (1.06) | (1.04) | (1.02) | 0.93 |
| | $\overline{\varepsilon}_{Arel}$ (%) | 79 | 78 | (31) | (31) 69 | 50 | 47 | 45 | 82 |
| | $\overline{\varepsilon}_{progArel}$ (%) | 193 | 201 | (44) | (41) 19 | – | – | – | 85 |
| Correlations | $R_{1S}$ | 0.11 | 0.13 | (0.19) | (0.19) 0.16 | (0.19) 0.19 | (0.21) 0.21 | (0.21) 0.21 | 0.20 |
| | $R_{2S}$ | 0.23 | 0.10 | (0.14) | (0.08) 0.05 | (0.13) | (0.04) | (0.21) | 0.20 |
| | $R_{3S}$ | 0.45 | 0.44 | (0.26) | (-0.11) -0.04 | (0.32) 0.24 | (0.16) 0.07 | (0.11) 0.02 | 0.52 |

Parameters of variogram models: $C_0$ – nugget effect; $C$ – sill; $L = C_0/(C_0 + C)$; $a$ – range of variogram (range of autocorrelation); errors of kriging estimation and coefficients of correlations – explained in text; Matheron, InvCov – types of space variability estimator; OK – ordinary kriging, SK – simple kriging, MIK – median indicator kriging (8 cutoffs), LNK – lognormal kriging; Norm. – normalization of the data. In some cases computing were impossible due to unbelievable values of kriging standard deviation (variants C1, C2, C3). In brackets – parameters before back-transformation of data. Statistically significant values ($\alpha = 0.05$) of the correlation coefficients are marked in red font. Further explanations see in the text.

Table 3. Variable H1_76, parameters of variogram models and measures of estimation quality for different variants of the estimation method

Tabela 3. Parametry modelu wariogramu oraz miary jakości estymacji dla różnych metod estymacji zmiennej H1_76

| Parameters of variogram models and measures of estimation quality | | Variant of computing procedure | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | (A1) Matheron + OK | (A2) InvCov +OK | (B1) Matheron + + LNK (OK) | (B2) Matheron + + LNK (SK) | (C1) Matheron + + Norm. + OK | (C2) Matheron + + Norm. + SK | (C3) InvCov + + Norm. + SK | (D) Matheron + + MIK |
| Parameters of variogram models | $C_0$ | 0.900 | 1.200 | 0.070 | 0.070 | 0.626 | 0.626 | 0.650 | 0.210 |
| | $C$ | 1.250 | 1.100 | 0.065 | 0.065 | 0.387 | 0.387 | 0.350 | 0.032 |
| | $L$ | 0.42 | 0.52 | 0.52 | 0.52 | 0.62 | 0.62 | 0.65 | 0.87 |
| | $a$ (m) | 5,000 | 10,000 | 6,669 | 6,670 | 6,631 | 6,631 | 9,000 | 7474 |
| Errors of kriging estimation | *MSE* | 0.03 | 0.03 | (0.03) | (0.01) | (0.03) | (0.01) | (0.01) | –0.16 |
| | *RMSSE* | 1.02 | 1.02 | (1.00) | (1.00) | (1.02) | (1.02) | (1.02) | 1.35 |
| | $\overline{\varepsilon}_{Arel}$ (%) | 50 | 48 | (26) | (26) 47 | 41 | 40 | 40 | 54 |
| | $\overline{\varepsilon}_{progArel}$ (%) | 63 | (61) | (30) | (30) 16 | – | – | – | 63 |
| Correlations | $R_{1S}$ | 0.24 | 0.27 | (0.27) | (0.26) -0.26 | (0.26) 0.26 | (0.25) 0.25 | (0.28) 0.28 | 0.30 |
| | $R_{2S}$ | 0.03 | 0.12 | (0.17) | (0.14) 0.22 | (0.25) | (0.08) | (0.09) | 0.26 |
| | $R_{3S}$ | 0.39 | 0.33 | (0.24) | (0.17) 0.22 | (0.25) 0.17 | (0.66) 0.07 | (0.65) 0.04 | 0.39 |

Explanation as in Table 2.

Table 4. Variable V1_80; parameters of variogram models and measures of estimation quality for different variants of the estimation method

Tabela 4. Parametry modelu wariogramu oraz miary jakości estymacji dla różnych metod estymacji zmiennej V1_80

| Parameters of variogram models and measures of estimation quality | | Variant of computing procedure | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | (A1) Matheron + OK | (A2) InvCov +OK | (B1) Matheron + + LNK (OK) | (B2) Matheron + + LNK (SK) | (C1) Matheron + + Norm. + OK | (C2) Matheron + + Norm. + SK | (C3) InvCov + + Norm. + SK | (D) Matheron + + MIK |
| Parameters of variogram models | $C_0$ | 5,100.0 | 5,700.0 | 0.042 | 0.042 | 0.79 | 0.79 | 0.770 | 0.160 |
| | $C$ | 1,920.0 | 1,500.0 | 0.014 | 0.014 | 0.31 | 0.31 | 0.230 | 0.100 |
| | $L$ | 0.73 | 0.79 | 0.75 | 0.75 | 0.72 | 0.72 | 0.77 | 0.62 |
| | $a$ (m) | 5,900 | 10,000 | 8,551 | 8,552 | 10,000 | 10,000 | 7,000 | 6,210 |
| Errors of kriging estimation | *MSE* | 0.05 | 0.01 | (0.01) | (–0.02) | (0.01) | (–0.01) | (–0.02) | 0.07 |
| | *RMSSE* | 1.00 | 1.00 | (1.00) | (1.00) | (1.00) | (0.99) | (1.00) | 0.97 |
| | $\overline{\varepsilon}_{Arel}$ (%) | 19 | 19 | (3) | (3) 19 | 18 | 18 | 18 | 22 |
| | $\overline{\varepsilon}_{progArel}$ (%) | 24 | 24 | (4) | (4) 0 | – | – | – | 28 |
| Correlations | $R_{1P}$ | 0.18 | 0.19 | (0.19) | (0.20) | (0.19) | (0.21) | (0.21) | 0.32 |
| | $R_{1S}$ | 0.18 | 0.19 | (0.20) | (0.21) | (0.20) | (0.21) | (0.20) | 0.20 |
| | $R_{2P}$ | 0.18 | 0.17 | (0.18) | (0.08) | (0.18) | (0.10) | (0.05) | 0.26 |
| | $R_{2P}$ | 0.21 | 0.17 | (0.18) | (0.08) 0.08 | (0.17) 0.16 | (0.10) 0.09 | (0.05) 0.05 | 0.22 |

Explanation as in Table 2.

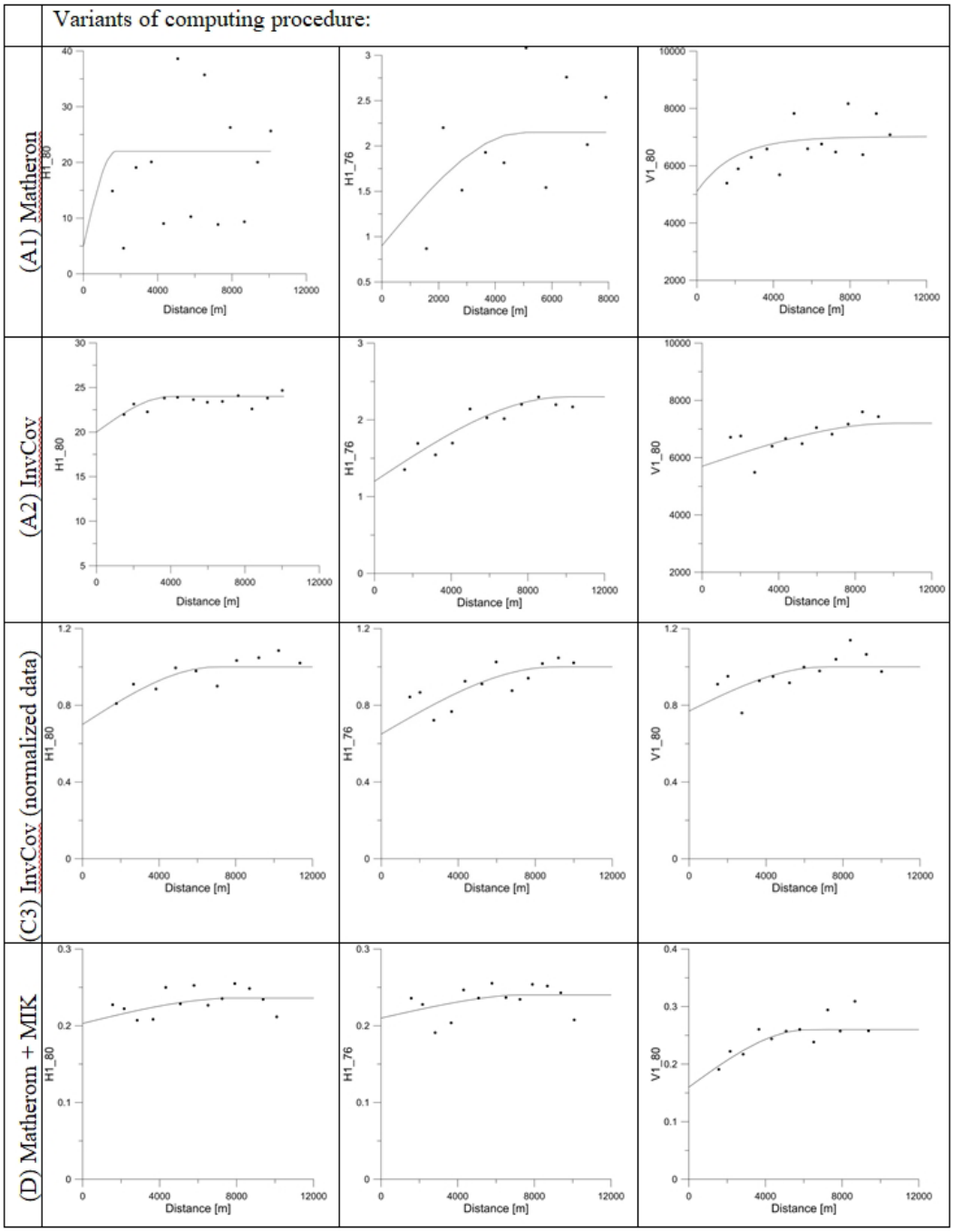


Fig. 5. Examples of spatial variability models. See and compare in Table 2, 3, 4.
Further explanations in the text

Rys. 5. Przykłady modeli zmienności przestrzennej

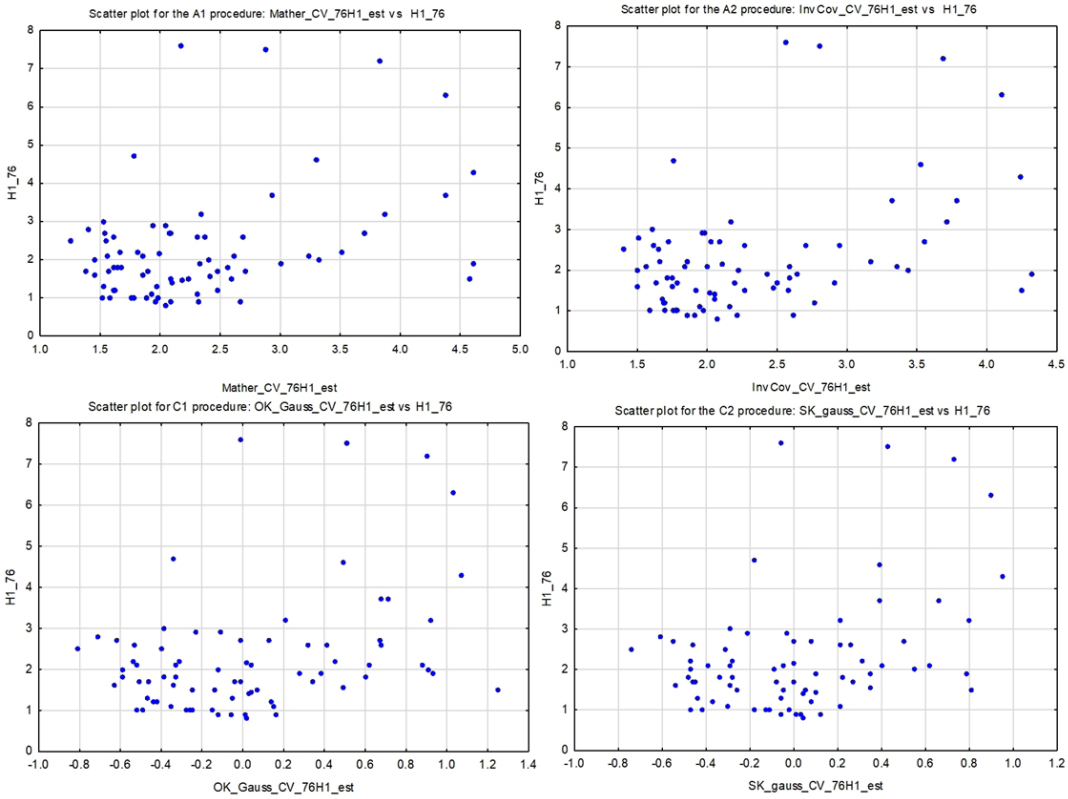


Fig. 6. Comparison of scatter diagrams between the estimates $\hat{z}(x_i)$ (abscissa axes) and the observed values $z(x_i)$ (ordinate axes) in the procedures A1, A2, C2, and C3, for cases where Spearman correlation coefficients values were statistically significant, i.e., for the dataset H1_76

Rys. 6. Porównanie diagramów rozrzutu między estymatami $\hat{z}(x_i)$ (osie odciętych) i wartościami obserwowanymi $z(x_i)$ (osie rzędnych) w procedurach A1, A2, C2 i C3, dla przypadków, gdzie wartości współczynników korelacji Spearmana były istotne statystycznie, tj. dla zbioru danych H1_76

These weak correlations often proved to be statistically significant in the H1_80 and H1_76 datasets (Tables 2 and 3), which should not really be the case. However, the undesirable, weak and statistically insignificant correlations of estimates with observed values: $\hat{z}$ vs. $z$, revealed in the H1_80 and V1_80 datasets (Tables 2 and 4), poorly predict the quality of future interpolation. Only for the H1_76 data (Table 3, see also scatterplots – Figure 6), the $R_{1S}$ coefficient values are statistically significant across all procedures. However, they indicate very weak correlation relationships, and although the systematically higher values of $R_{1S}$ in procedures where the InvCov estimator was used (A2 and C3) are intriguing, the remaining values are only slightly smaller. The values of the $R_{1S}$ coefficient, therefore, are not a good criterion for evaluating the applied procedures; they can rather serve to validate the calculations on the H1_76 dataset.

Finally, the error measure $\overline{\varepsilon}_{Arel}$ was recognized as the only useful criterion for the evaluation of estimation procedures (Table 5). One can see a decrease in this error measure in all variants after cutting off four outliers, i.e. when moving from the most difficult (H1_80) to intermediate (H1_76) case for estimation. The lowest value of the error measure appears for the easiest case (V1_80). It is strange in this case that the values of the $R_{1P}$ coefficient are statistically insignificant. This regularity also works for data before logarithmic back transformation (variants B1, B2), although one can easily notice that this transformation introduces a large, additional error. In further considerations, in relation to the variants in which backward transformation was used (B1, B2, C1, C2, C3), we will limit ourselves only to errors after this transformation.

For variables H1_80 and H1_76, which are difficult during estimation, with highly skewed distributions, high variability and kurtosis, and also having outliers, it is easy to indicate which variants give the lowest errors. These are type C variants with data normalization. In the particularly challenging case (H1_80), the most effective approach is the C3 variant, using the InvCov estimator, and simple kriging (SK). The error rate is, however, too high, even in the C variants (for example in the C3 variant: H1_80 – 45%, H1_76 – 40%) to allow one to use point kriging (OK, SK). Low efficiency of logonormal kriging (B1, B2) as well as median indicator kriging (D) is puzzling. In the easiest case (V1_80), all estimation procedures work well, giving a low enough error (maximum of 19%; Table 4), so if it were not for the lack of correlation vs. $z$, point interpolation would be possible.

The phenomenon of lack of correlation is probably caused by the high values of the relative nugget effect $L$ (Tables 2, 3 and 4), i.e. high share of the random component of variance – the nugget effect ($C_0$) in the overall variance, which causes strong averaging in the kriging procedure – underestimation of maxima and overestimation of minima (Hohn 1988; Deutsch and Journel 1997). However, systematically the smallest $L$ values for the variable H1_80, intermediate for H1_76, and the largest for V1_80 (Tables 6 and 7), received by mean of the Matheron's estimator of variogram, do not have an expected effect on the strength of the said correlation. It would be expected that the $R_{1S}$ coefficients would have the highest values for V1_80, intermediate for H1_76 and the lowest for H1_80, but this did not happen. This expected correct relationship between $L$ and $R_{1S}$ could happen if the $L$ value for the variable H1_80 were higher. Such relationship only occurred when InvCov (variants A3 and C3) was used as the estimator, and not the Matheron estimator. Therefore, it seems legitimate to conclude that with this estimator, it was possible to better assess the $L$ value in cases of highly positively skewed data, which was represented by H1_80. Better assessment of $L$ gave a better assessment of the $C_0$ and $C$ parameters of the spatial variability model.

It is widely accepted that non-ergodic variogram estimators, such as the InvCov, are also robust for skewed data. This was commonly postulated (e.g., Rossi et al. 1992; Nieć and Mucha 2007; Kokesz 2010), although this has recently been called into question (Lark 2008; Curriero et al. 2002). Our work, although based on limited data, suggests that the first authors are probably right in saying that the InvCov estimator is more robust to data skewness

than the Matheron estimator. Moreover, since the use of a non-ergodic estimator for ergodic data does not affect the quality of the model, it is safer to use the InvCov estimator if the nature of the data is not known.

Table 5. Comparison of the values

Tabela 5. Porównanie wartości

| Variable | (A1) Matheron + +OK | (A2) InvCov + +OK | (B1) Matheron + + LNK (OK) | (B2) Matheron + + LNK (SK) | (C1) Matheron + + Norm. + + OK | (C2) Matheron + + Norm. + + SK | (C3) InvCov + + Norm. + + SK | (D) Matheron + + MIK |
|---|---|---|---|---|---|---|---|---|
| H1_80 | 79 | 78 | (31) | (31) 69 | 50 | 47 | 45 | 82 |
| H1_76 | 50 | 48 | (26) | (26) 47 | 41 | 40 | 40 | 54 |
| V1_80 | 19 | 19 | (3) | (3) 19 | 18 | 18 | 18 | 22 |

Explanation as in Table 2. Data come from Tables 2, 3 and 4.

Table 6. Comparison of the $L$ values and of differences of $L$ values among A2 and A1, as well as C3 and C2 procedures

Tabela 6. Porównanie wartości $L$ i różnic wartości $L$ między procedurami A2 i A1 oraz C3 i C2

| Variable | (A1) Matheron + + OK | (A2) InvCov + + OK | Delta $L$ (A2–A1) | (C2) Matheron + + Norm. + SK | (C3) InvCov + + Norm. + SK | Delta $L$ (C2–C2) |
|---|---|---|---|---|---|---|
| H1_80 | 0.23 | 0.83 | 0.60 | 0.40 | 0.70 | 0.30 |
| H1_76 | 0.42 | 0.52 | 0.10 | 0.62 | 0.65 | 0.03 |
| V1_80 | 0.73 | 0.79 | 0.06 | 0.72 | 0.77 | 0.05 |

Explanation as in Table 2. Data come from Tables 2, 3 and 4.

Table 7. Comparison of the $R_{1S}$ values among A2 and A1, as well as C3 and C2 procedures

Tabela 7. Porównanie wartości $R_{1S}$ pomiędzy procedurami A2 i A1 oraz C3 i C2

| Variable | (A1) Matheron + + OK | (A2) InvCov + + OK | Delta $R_{1S}$ (A2–A1) | (C2) Matheron + + Norm. + SK | (C3) InvCov + + Norm. + SK | Delta $R_{1S}$ (C2–C2) |
|---|---|---|---|---|---|---|
| H1_80 | 0.11 | 0.13 | 0.02 | (0.19) | (0.21) | 0.02 |
| H1_76 | 0.24 | 0.27 | 0.03 | (0.25) | (0.28) | 0.03 |
| V1_80 | 0.18 | 0.19 | 0.01 | (0.21) | (0.21) | 0.00 |

In brackets – values for normalized data; other explanations as on the Table 2. Data come from Tables 2, 3 and 4.

## Conclusions

In the paper, based on the three datasets, the usefulness of several low-laborious geostatistical procedures in interpolation highly positively skewed seismic data but without significant influence of trend and anisotropy have been tested. To prepare the usability ranking, several measures of interpolation error were used, as well as two correlation coefficients: kriging estimates vs. observed values and kriging estimates vs. interpolation errors. Only one measure was recognized as a useful criterion that was able to differentiate the quality of interpolation. This was the average absolute relative error $\overline{\varepsilon}_{Arel}$. On the basis of this criterion, one could confirm the fundamental meaning of data normalization, which was expected. The low usability of the lognormal kriging and the simple median indicator kriging methods was surprising here.

Low usefulness of the mentioned coefficients of correlation between kriging estimates and observed values for distinguishing the quality of interpolation probably was caused by high values of share of random component in general variance *L*. Apart from this, it was expected that the correlations would be stronger when lower *L* values occurred. It was easy to notice that such a dependence, albeit weak, took place only when inverted covariance variogram estimator (InvCov) was used. This is a premise that the InvCov estimator works better in difficult cases than the Matheron estimator; moreover, it works with more stability and is easy with regard to modeling.

Ultimately, the question arises of whether interpolation is possible based on any of the three tested datasets. In the H1_76 dataset, there are relatively high errors (the lowest after data normalization), but it should be noted that they were estimated using different variants of point kriging. Most likely, these errors would decrease to an acceptable level in interpolation using block kriging. Additionally, significant albeit low Spearman correlation coefficients between kriging estimates and "true" values would be favorable during interpolation. In the case of the V1_80 dataset, the values of these coefficients are low and insignificant, which is most likely due to high values of the relative nugget effect L. However, the errors $\overline{\varepsilon}_{Arel}$ are low enough that even point kriging could be used for interpolation. Nonetheless, using geostatistical methods in such cases is not very effective, although it likely would not worsen the interpolation. As for the H1_80 dataset, which differs from H1_76 only by the presence of four outliers, it can be concluded that interpolation is not possible. The reasons are high errors $\overline{\varepsilon}_{Arel}$ and the practical lack of correlation between kriging estimates and "true" values.

*This research has been funded by the Polish National Centre for Research and Development (NCRD) grant under the "Blue Gas Programme" – "Badania sejsmiczne i ich zastosowanie dla detekcji stref występowania gazu z łupków. Dobór optymalnych parametrów akwizycji i przetwarzania w celu odwzorowania budowy strukturalnej oraz rozkładu parametrów petrofizycznych i geomechanicznych skał perspektywicznych" („Seismic surveys and their application for the detection of shale gas zones. Selection of optimal acquisition and processing parameters for the imaging of structural*

*architecture and distribution of petrophysical and geomechanical parameters in prospective rock formations"). The Polish Oil and Gas Company in Warsaw (PGNiG) is acknowledged for access to seismic data.*

## REFERENCES

Aleardi et al. 2020 – Aleardi, M., Salusti, A. and Pierini, S. 2020. Trans-dimensional and Hamiltonian Monte Carlo inversions of Rayleigh-wave dispersion curves: A comparison on synthetic datasets. *Near Surface Geophysics* 18(5), pp. 515–543, DOI: 10.1002/nsg.12100.

Atkinson, P.M., and Lloyd, C.D. (eds.) 2010. GeoENV VII – Geostatistics for Environmental Applications: *Proceedings of the Seventh European Conference on Geostatistics for Environmental Applications*. Springer Science and Business Media, pp. 205.

Azevedo, L. and Demyanov, V. 2019. Multiscale uncertainty assessment in geostatistical seismic inversion. *Geophysics* 84(3), pp. 1–67, DOI: 10.1190/geo2018-0329.1.

Badel et al. 2011 – Badel, M., Angorani, S., and Panahi, M.S. 2011. The application of median indicator kriging and neural network in modeling mixed population in an iron ore deposit. *Computers and Geosciences* 37, pp. 530–540, DOI: 10.1016/j.cageo.2010.07.009.

Bleinès et al. 2014 – Bleinès, C., Bourges, M., Deraisme, J., Geffroy, F., Jeannée, N., Lemarchand, O., Perseval, S., Rambert, F., Renard, D., Touffait, Y. and Wagner, L. 2014. *Isatis Technical References Software*. Avon, France.

Botor et al. 2019a – Botor, D., Golonka, J., Anczkiewicz, A. A., Dunkl, I., Papiernik, B., Zając, J. and Guzy, P. 2019a. Burial and thermal history of the Lower Paleozoic petroleum source rocks in the SW margin of the East European Craton (Poland). *Annales Societatis Geologorum Poloniae* 89, pp. 31–62, DOI: 10.14241/asgp.2019.12.

Botor et al. 2019b – Botor, D., Golonka, J., Zając, J., Papiernik, B. and Guzy, P. 2019b. Petroleum generation and expulsion in the Lower Palaeozoic petroleum source rocks at the SW margin of the East European Craton (Poland). *Annales Societatis Geologorum Poloniae* 89, pp. 63–89, DOI: 10.14241/asgp.2019.11.

Chahooki et al. 2019 – Chahooki, M.Z, Javaherian, A. and Saberi, M.R. 2019. Realization ranking of seismic geostatistical inversion based on a Bayesian lithofacies classification – A case study from an offshore field. *Journal of Applied Geophysics* 170, DOI: 10.1016/j.jappgeo.2019.07.008.

Chilès, J.P. and Delfiner, P. 2012. *Geostatistics. Modeling Spatial Uncertainty*. Second Edition, Wiley and Sons.

Clark, I. 1999. Geostatistical estimation applied to highly skewed data. *Proceedings of Joint Statistical Meetings*, Dallas, Texas.

Curriero et al. 2002 – Curriero, F.C., Hohn, M.E., Liebhold, A.M. and Lele, S.R. 2002. A statistical evaluation of non-ergodic variogram estimators. *Environmental and Ecological Statistics* 9(1), pp. 89–110, DOI: 10.1023/A:1013771109591.

Deutsch, C.V. and Journel, A.G. 1997. *GSLIB Geostatistical Software Library and User's Guide*. New York: Oxford University Press, 340 pp.

Dip et al. 2021 – Dip, A.C., Giroux, B. and Gloaguen, E. 2021. Microseismic monitoring of rockbursts with the ensemble Kalman filter. *Near Surface Geophysics* 19, pp. 429–445, DOI: 10.1002/nsg.12158.

Domagała et al. 2021 – Domagała, K., Maćkowski, T., Stefaniuk, M. and Reicher, B. 2021. Prediction of Reservoir Parameters of Cambrian Sandstones Using Petrophysical Modelling – Geothermal Potential Study of Polish Mainland Part of the Baltic Basin. *Energies* 14, DOI: 10.3390/en14133942

Englund, E. and Sparks, A. 1999. GEO – EAS 1.2.1 Geostatistical Environmental Assessment Software. User's Guide. Environmental Monitoring Systems Laboratory Office of Research And Development U.S. Environmental Protection Agency Las Vegas, Nevada 89119.

Goovaerts, P. 1997. Geostatistics for Natural Resources Evaluation. New York: Oxford University Press, 483 pp.

Guterch et al. 2010 – Guterch, A., Wybraniec, S., Grad, M., Chadwick, A., Krawczyk, C.M., Ziegler, P.A., Thybo, H. and De Vos, W. 2010. Chapter 2: *Crustal structure and structural framework*. [In:] Doornebal, H. and Steven-

son, A. (eds.), *Petroleum Geological Atlas of the Southern Permian Basin Area*. EAGE Publications, Houten, the Netherlands, pp. 11–23.

Hohn, E.M. 1988. *Geostatistics and petroleum geology*. Van Nostrand Reinhold, New York, USA. 264 pp.

Isaaks, E.H. and Srivastava, R.M. 1988. Spatial continuity measures for probabilistic and deterministic geostatistics. *Mathematical Geology* 20, pp. 313–341, DOI: 10.1007/BF00892982.

Isaaks, E.H., and Srivastava, R.M. 1989. *Applied Geostatistics*. New York–Oxford: Oxford University Press, pp. 561.

Jain et al. 2000 – Jain, A.K., Duin, R.P.W. and Mao, J. 2000. Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(1), pp. 4–37, DOI: 10.1109/34.824819.

Johnnston et al. 2001 – Johnnston, K., Hoef, J.M., Krivoruchko, K., and Lucas, N. 2001. *Using ArcGIS Geostatistical Analysis*. GIS User Manual by ESRI, New York.

Journel, A.G. 1980. The lognormal approach to predicting local distributions of selective mining unit grades. *Mathematical Geology* 12, pp. 285–303, DOI: 10.1007/BF01029417.

Journel, A.G. and Deutsch, C.V. 1997. *Rank order geostatistics: a proposal for a unique coding and common processing of diverse data*. [In:] Baafi, E.Y., Schofield, N.A. (eds.), Geostatistics Wollongong '96. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 174–187.

Kasperska et al. 2019 – Kasperska, M., Marzec, P., Pietsch, K. and Golonka, J. 2019. Seismo-geological model of the Baltic Basin (Poland). *Annales Societatis Geologorum Poloniae* 89, pp. 195–213, DOI: 10.14241/asgp.2019.02.

Kerry, R. and Oliver, M.A. 2007a. Determining the effect of asymmetric data on the variogram. I. Underlying asymmetry. *Computers and Geosciences* 33, pp. 1212–1232.

Kerry, R. and Oliver, M.A. 2007b. Determining the effect of asymmetric data on the variogram. II. Outliers. *Computers and Geosciences* 33, pp. 1233–1260.

Kokesz, Z. 2006. Difficulties and limitations in geostatistical modelling of mineral deposits variabilities and resources/reserves estimation by kriging (*Trudności i ograniczenia w geostatystycznym modelowaniu zmienności złóż i szacowaniu zasobów metodą krigingu*). *Gospodarka Surowcami Mineralnymi – Mineral Resources Management* 22(3), pp. 5–20 (*in Polish*).

Kokesz, Z. 2010. Constraints on ordinary kriging application for contour maps construction. *Bulletin of Polish Geological Institute* 439, pp. 403–408.

Kwietniak et al. 2021 – Kwietniak, A., Maćkowski, T. and Cichostępski, K. 2021. Seismic signature of transition zone (Wolf ramp) in shale deposits with application of frequency analysis. *Geofluids,* p. 1–16, DOI: 10.1155/2021/6614081.

Lark, R.M. 2008. A comparison of some robust estimators of the variogram for use in soil survey. *European Journal of Soil Science* 51(1), pp. 137–157, DOI: 10.1046/j.1365-2389.2000.00280.x.

Li, D. and Zhao, C. (eds.) 2009. Computer and Computing Technologies in Agriculture II, Volume 1: The Second IFIP International Conference on Computer and Computing Technologies in Agriculture (CCTA2008), October 18–20, 2008, Beijing, China. Springer Science and Business Media, pp. 130.

Mazur et al. 2018a – Mazur, S., Krzywiec, P., Malinowski, M., Lewandowski, M., Aleksandrowski, P. and Mikołajczak, M. 2018a. On the nature of the Teisseyre-Tornquist Zone. *Geology, Geophysics and Environment* 44(1), pp. 17–30, DOI: 10.7494/geol.2018.44.1.17.

Mazur et al. 2018b – Mazur, S., Gągała, Ł., Kufrasa, M. and Krzywiec, P. 2018b. Application of two-dimensional gravity models as input parameters to balanced cross-sections across the margin of the East European Craton in SE Poland. *Journal of Structural Geology* 116, pp. 223–233, DOI: 10.1016/j.jsg.2018.05.013.

Nieć, M. and Mucha, J. 2007. From statistics to geostatistics in geological investigations of Lower Silesia copper ore deposits – 50 years of experience. *Bulletin of Polish Geological Institute* 423, pp. 59–67.

Parra, J. and Emery, X. 2013. Geostatistics applied to cross-well reflection seismic for imaging carbonate aquifers. *Journal of Applied Geophysics* 92, pp. 68–75, DOI: 10.1016/j.jappgeo.2013.02.010.

Pletsch et al. 2010 – Pletsch, T., Appel, J., Botor, D., Clayton, C.J., Duin, E.J.T., Faber, E., Górecki, W., Kombrink, H., Kosakowski, P., Kuper, G., Kus, J., Lutz, R., Mathiesen, A., Ostertag, C., Papiernik, B. and van Bergen, F. 2010. *Petroleum generation and migration*. [In:] Doornenbal, J.C., Stevenson, A. (eds.) *Petroleum Geological Atlas of the Southern Permian, Basin Area*. Houten: EAGE Publications, pp. 225–253.

Rossi et al. 1992 – Rossi, R.E., Mulla, D.J. and Journel, A.J. 1992. Geostatistical tools for modelling and interpreting ecological spatial dependence. *Ecological Monographs* 62, pp. 277–314, DOI: 10.2307/2937096.

Roth, C. 1998. Is lognormal kriging suitable for local estimation? *Mathematical Geology* 30(8), pp. 999–1009, DOI: 10.1023/A:1021733609645.

Saito, H. and Goovaerts, P. 2000. Geostatistical interpolation of positively skewed and censored data in a dioxin-contaminated site. *Environmental Science and Technology* 34(19), pp. 4228–4235, DOI: 10.1021/es991450y.

Srivastava, R.M. and Parker, H.M. 1988. *Robust measures of spatial continuity*. [In:] Armstrong, M. et al. (eds), *Third International Geostatistics Congress*, D. Reidel, Dodrecht, the Netherlands, pp. 295–308.

Vann, J. and Guibal, D. 1998. *Beyond ordinary kriging: An overview of non-linear estimation*. [In:] Vann, J. (ed.), *Beyond Ordinary Kriging Seminar. Perth, Western Australia: Geostatistical Association of Australasia* (Monograph 1).

Wasilewska, M. and Mucha, J. 2005. Kriging as a method of interpolation of parameters describing the quality of hard coal in the Upper Silesian Coal Basin (GZW) (*Kriging jako metoda interpolacji parametrów opisujących jakość węgla kamiennego w pokładach Górnośląskiego Zagłębia Węglowego (GZW)*). *Materiały Sympozjum Warsztaty Górnicze z cyklu „Zagrożenia naturalne w górnictwie", Kazimierz Dolny nad Wisłą*, 20–22 June 2005, pp. 341–354 (*in Polish*).

Yamamoto, J.K. 2007. On unbiased backtransform of lognormal kriging estimates. *Computational Geosciences* 11, pp. 219–234, DOI: 10.1007/s10596-007-9046-x.

## RANKING OF THE UTILITY OF SELECTED GEOSTATISTICAL INTERPOLATION METHODS IN CONDITIONS OF HIGHLY SKEWED SEISMIC DATA DISTRIBUTIONS: A CASE STUDY OF THE BALTIC BASIN (POLAND)

Keywords

seismic, data processing, shallow subsurface, uncertainty, variability

Abstract

The suitability of several low-labor geostatistical procedures in the interpolation of highly positively skewed seismic data distributions was tested in the Baltic Basin. These procedures were a combination of various estimators of the model of spatial variation (theoretical variogram) and kriging techniques, together with the initial data transformation to normal distribution or lack thereof. This transformation consisted of logarithmization or normalization using the anamorphosis technique. Two variations of the theoretical variogram estimator were used: the commonly used classical Matheron estimator and the inverse covariance estimator (InvCov), which is robust with regard to non-ergodic data. It was expected that the latter would also be resistant to strongly skewed data distributions. The kriging techniques used included the commonly used ordinary kriging, simple kriging useful for standardized data and the non-linear median indicator kriging technique. It was confirmed that normalization (anamorphosis) is the most useful and less laborious geostatistical procedure of those suitable for such data, which results in a standardized normal distribution. The second, not obvious statement for highly skewed data distributions suggests that the non-ergodic inverted covariance (InvCov) estimator of variogram has an advantage over the Matheron's estimator. It gives a better assessment of the $C_0$ (nugget effect) and $C$ (sill) parameters of the spatial variability model. Such a conclusion can be drawn from the fact that the higher the estimation of the relative nugget effect

$L = C_0/(C_0 + C)$ using the InvCov estimator, the weaker the correlation between the kriging estimates and the observed values. The values of the coefficient $L$ estimates obtained by using the Matheron's estimator do not meet this expectation.

## RANKING PRZYDATNOŚCI WYBRANYCH METOD INTERPOLACJI GEOSTATYSTYCZNEJ W WARUNKACH SILNIE SKOŚNYCH ROZKŁADÓW DANYCH SEJSMICZNYCH: STUDIUM PRZYPADKU BASENU BAŁTYCKIEGO (POLSKA)

Słowa kluczowe

sejsmika, przetwarzanie danych, strefa przypowierzchniowa, niepewność, zmienność

Streszczenie

W ramach studium przypadku w rejonie basenu bałtyckiego przetestowano przydatność kilku mało pracochłonnych procedur geostatystycznych do interpolacji silnie skośnych rozkładów danych sejsmicznych. Były one kombinacją różnych estymatorów modelu zmienności przestrzennej (wariogramu teoretycznego) i technik krigingu, wraz ze wstępną transformacją danych do rozkładu normalnego lub jej brakiem. Transformacja ta polegała na logarytmowaniu bądź na normalizacji z użyciem techniki anamorfozy. Zastosowano dwie odmiany estymatora wariogramu teoretycznego: powszechnie stosowany klasyczny estymator Matherona oraz estymator odwróconej kowariancji (InvCov) odporny na dane nieergodyczne. Spodziewano się, że ten drugi okaże się również odporny na silnie skośne rozkłady dane. Wśród zastosowanych technik krigingu znalazł się powszechnie stosowany kriging zwyczajny, kriging prosty użyteczny dla danych zestandaryzowanych i nieliniowa technika krigingu wskaźnikowego. Najbardziej użyteczną i mało pracochłonną procedurą geostatystyczną, nadającą się do zastosowania w przypadku takich danych, okazała się normalizacja (anamorfoza), w efekcie której uzyskuje się rozkład normalny standaryzowany. Drugim, nieoczywistym wnioskiem dla silnie skośnych rozkładów danych, jest sugestia, iż estymator InvCov ma przewagę nad estymatorem Matherona, ponieważ pozwala na bardziej realistyczną ocenę parametrów $C_0$ (efektu samorodka) i $C$ (wariancji progowej) modelu zmienności przestrzennej. Taki wniosek można wyciągnąć z faktu, że im wyższa wartość relatywnego efektu samorodków $L = C_0/(C_0 + C)$ obliczona za pomocą estymatora InvCov, tym słabsza korelacja między wartościami obliczonymi a danymi. Wartości współczynnika $L$ obliczone za pomocą estymatora Matherona nie posiadają tej właściwości.