

## Geostatystyka dla niematematyków. Przewodnik praktyczny – *appendix*

Wojciech Naworyta<sup>1</sup>



**Geostatistics for non-mathematicians. Practical guide – appendix.** Prz. Geol., 72: 685–691; doi: 10.7306/2024.52

*Abstract.* Geostatistical tools are useful and even necessary in many fields, not only in geology. The obstacle to their widespread use is the seemingly difficult mathematical foundations with which the teaching of the subject usually begins. Many years of experience and observations allow me to claim that even if geostatistical methods are used, they are often incomplete and imperfect. In a series of three articles, I tried to familiarize potential non-mathematicians with the most important methods and tools from the arsenal of spatial statistics and how to use them properly. I indicated the areas where these methods can be used, explained whether they can always be used, showed what decisions should be made during calculations and how the results obtained should be interpreted. It is not a compendium, but rather a pocket guide facilitating the reader's first contact with geostatistics.

After publishing my triptych under the common title "Geostatistics for non-mathematicians", I decided to add a supplement in which I will discuss several important issues. These include the condition of stationarity, the problem of the presence of a trend in the variability of the analysed phenomenon, and the situation where there are observations that differ significantly from the others, called outliers.

**Keywords:** geostatistics, stationarity, trend, outliers

Od omówienia warunku stacjonarności rozpoczyna się każdy klasyczny kurs geostatystyki. Swój cykl artykułów określiłem jako *praktyczny przewodnik*, więc pozwoliłem sobie na dużą dowolność w doborze zagadnień. Położyłem nacisk na te służące praktycznemu zastosowaniu narzędzi geostatystycznych. Skupiając się na zagadnieniach praktycznych, zagadnienia teoretyczne poruszałem tylko w takim zakresie, w jakim to było konieczne dla zrozumienia tych pierwszych. Aby jednak być uczciwym wobec Czytelnika, nie mogę pominąć kluczowego z punktu widzenia teorii geostatystyki założenia stacjonarności. Dla ułatwienia pojęcie stacjonarność zastąpiłbym mianem homogeniczności. Zjawisko, które modelujemy z wykorzystaniem narzędzi geostatystycznych, powinno być w całym modelowanym obszarze homogeniczne pod względem zmienności.

### **CZY I JAK SPRAWDZIĆ, CZY MODELOWANE WIELKOŚCI WYKAZUJĄ STACJONARNOŚĆ?**

Wracając do nomenklatury geologicznej, zamiast o zmienności zjawiska przestrzennego częściej będę mówił o zmienności wybranego parametru złożowego, np. miąższości pokładu, zawartości składnika użytecznego, wartości opałowej itp. Parametr w obszarze złoża jest zmienny. Jego zmienność w modelowanym złożu powinna być wszędzie podobna, czyli niezależnie od miejsca w złożu różnice pomiędzy parami obserwacji w takich samych odległościach pomiędzy obserwacjami powinny być statystycznie podobne. W procesie modelowania wykorzystujemy model wariogramu obliczony na podstawie wszystkich obserwacji, które później będą wykorzystane w procesie modelowania. Model wariogramu pokazuje, jak statystycznie różnią się (różnice kwadratowe) obserwacje w parach odległych od siebie o pewną odległość  $h$  (Nawory-

ta, 2024b). Wariogram pokazuje zależność statystyczną, czyli średnią. Stosując model wariogramu w procesie modelowania parametru złoża zakładamy, że zależność widoczna w kształcie wariogramu dotyczy całego złoża – zarówno obserwacji, na podstawie których ta zależność została obliczona (zidentyfikowana), jak i całej przestrzeni pomiędzy tymi obserwacjami, np. w węzłach siatki interpolacyjnej.

Gdyby warunek stacjonarności nie był spełniony, albo był spełniony tylko w pewnej części złoża, a gdzie indziej nie, to stosowanie modelu wariogramu w procesie krigingu do modelowania zmienności parametru złożowego w całej przestrzeni złoża byłoby nieuprawnione. Mówiąc kolokwialnie – zależność zobrazowana wariogramem powinna być ważna albo obowiązująca w całej przestrzeni modelowanego zjawiska. Tyle teoria, a jak wygląda rzeczywistość i jak to sprawdzić w praktyce?

Najczęściej modelowane zjawisko, np. parametr złożowy, nie podporządkowuje się ściśle teoretycznej zasadzie stacjonarności. Złoża powstają w różnych okolicznościach, w ich zmienności występują zaburzenia, np. w formie uskoków albo w skutek sfałdowań, niektóre powstają w rowach tektonicznych, więc w obszarach brzegowych ich parametry wykazują zmienność odmienną w stosunku do parametrów w części wewnętrznej złoża. Mógłbym więc zaryzykować stwierdzenie, że zasada stacjonarności jako zasada teoretyczna w praktyce ma ograniczone zastosowanie. Uważam, że nestorzy geostatystyki musieli wprowadzić to założenie, aby dało się wyprowadzić dalsze zależności i stworzyć narzędzia, takie jak wariogram czy kriging. Jednak to, co napisałem, nie uważa nas do ignorowania tego założenia. Jak więc dowiedzieć się, czy analizowane zjawisko jest stacjonarne czy nie. Czy musimy to robić? A co, jeśli okaże się, że zjawisko nie jest stacjonarne? Czy stosowanie metod geostatystycznych będzie wtedy błędem?

<sup>1</sup> Wydział Inżynierii Lądowej i Gospodarki Zasobami, Akademia Górniczo-Hutnicza im. St. Staszica, al. Adama Mickiewicza 30; 30–059 Kraków; naworyta@agh.edu.pl; ORCID ID: 0000-0003-4569-3907

Teoretycznie należałoby losowo wybrać kilka podobszarów w rejonie złoża i dla każdego z tych podobszarów obliczyć osobny wariogram. Jeżeli zmienność zjawiska wykazuje stacjonarność, to wszystkie wariogramy dla wybranych podobszarów powinny być do siebie podobne. W praktyce jednak dysponujemy ograniczoną liczbą obserwacji, często jest ich za mało, aby spośród nich wybrać jeszcze podzbiory w określonych podobszarach złoża i na ich podstawie obliczyć osobne wariogramy. Aby obliczyć wariogram, powinniśmy mieć do dyspozycji co najmniej kilkadziesiąt obserwacji. Według mnie 20 obserwacji to w analizach geostatystycznych absolutne minimum. Wartości na wariogramie empirycznym to wielkości statystyczne, średnie, aby były statystycznie istotne, powinny być obliczone na podstawie wielu par obserwacji – im więcej tym lepiej. Trudno jest udowodnić stacjonarność zjawiska przestrzennego, obliczając wariogramy w podobszarach na podstawie niewielkich podzbiorów obserwacji.

Jak zatem dowiedzieć się, czy zmienność interesującego nas zjawiska wykazuje stacjonarność czy nie? Najlepiej wykonać interpolację, wykorzystując do tego jedną z metod niegeostatystycznych, np. odwrotność odległości do potęgi (*inverse distance to a power*) albo minimum krzywizny (*minimum curvature*) i ocenić wizualnie, czy zmienność zjawiska w całym obszarze ma podobny przebieg, czy też może występują jakieś wyraźnie widoczne odchylenia, które upoważniałyby do twierdzenia, że wariogram nie będzie dobrze reprezentował zmienności zjawiska w całym obszarze. Wymienione metody interpolacyjne nie wymagają od operatora dużego nakładu pracy. Taką wstępną analizę określa się mianem szybka interpolacja (*quick interpolation*). Jest to proces, w którym

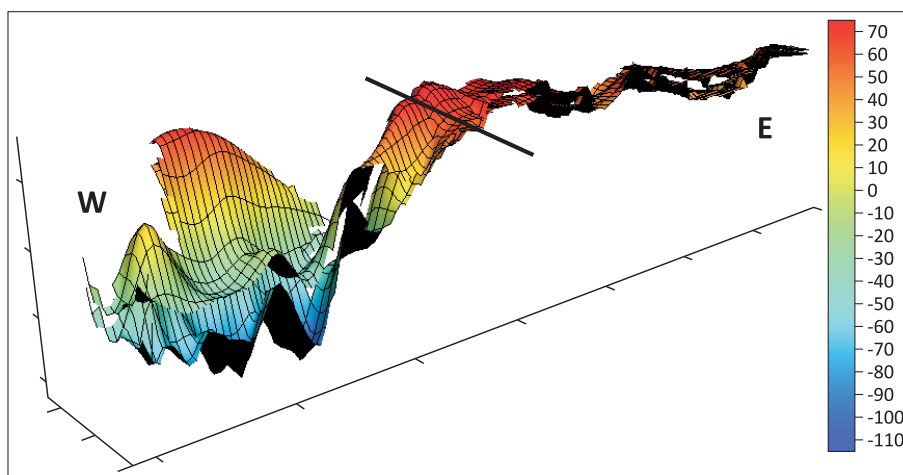
nie chodzi o dokładność modelu, tylko o możliwość szybkiego wglądu w jego podstawowe cechy.

Dla zilustrowania zjawiska braku stacjonarności przedstawiłem model jednego ze złóż węgla brunatnego (ryc. 1). Dwie powierzchnie na modelu trójwymiarowym obrazują strop i spąg pokładu. Już na pierwszy rzut oka widać, że pod względem budowy składa się ono z dwóch wyraźnie różniących się części. Część wschodnia (E) to pokład o prostej budowie, niewielkiej zmienności miąższości – od 2,4 do 18,6 m, ze średnią  $m = 9,62$  m i odchyleniem standardowym  $s = 5,3$  m. Natomiast w części zachodniej (W) złożo ma zupełnie inny kształt. Miąższość pokładu dochodzi tu do 106,7 m, ze średnią  $m = 28,3$  m i odchyleniem standardowym  $s = 25,82$  m. W tabeli 1 zestawiliem podstawowe cechy statystyczne dwóch części analizowanego złoża (Naworyta, Zajączkowski, 2018). Jeżeli policzy się wariogram na podstawie obserwacji z całego złoża, to nie będzie on dobrze reprezentował ani części wschodniej (E), gdzie zmienność jest niewielka, ani części zachodniej (W), gdzie rozpiętość wartości zarówno w stropie, jak i spągu jest znacząco większa. Stosując metody geostatystyczne, należałoby podzielić złożo na dwie części i każdą z nich modelować osobno.

O ile ocena stacjonarności w zmienności parametrów strukturalnych złoża przedstawionego na ryc. 1 jest dość oczywista, to jednak nie zawsze jest tak łatwo i granica pomiędzy stacjonarnością a jej brakiem jest dość płynna. Trudno wyznaczyć uniwersalne kryteria, dla których należałoby przyjąć, że parametr złożowy nie wykazuje stacjonarności i trzeba odstąpić od stosowania metod geostatystycznych. Najlepiej chyba w takich przypadkach polegać na wyczuciu i doświadczeniu, których nabiera się z czasem.

→

**Ryc. 1.** Model strukturalny złoża o skomplikowanej budowie ilustrujący brak stacjonarności w zmienności rzędnych stropu i spągu  
**Fig. 1.** Structural model of a deposit with a complex structure, illustrating the lack of stationarity in the variability of the elevations of the top and base



**Tab. 1.** Podstawowe cechy strukturalne złoża przedstawionego na ryc. 1  
**Table 1.** Basic structural features of the deposit shown in Fig. 1

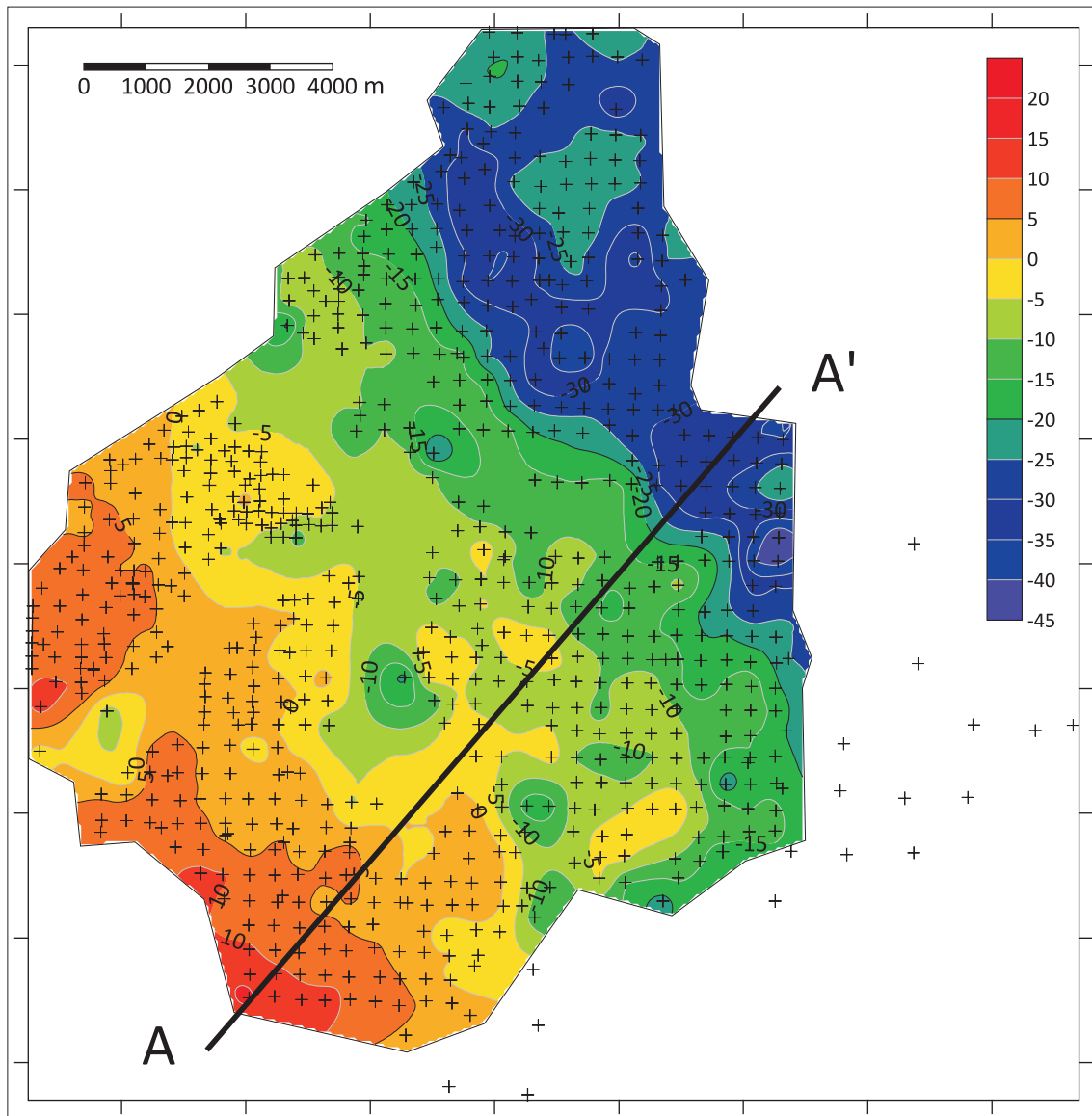
Cecha Feature	Wschodnia część złoża – E East part of the deposit – E	Zachodnia część złoża – W West part of the deposit – W
Rzędne spągu / Elevations of the base	od -130,7 do 58,8 m n.p.m.	od -3,3 do 86,2 m n.p.m.
Rzędne stropu / Elevations of the top	od -90,9 do 72,5 m n.p.m.	od 0,0 do 88,8 m n.p.m.
Miąższość pokładu / Thickness of the seam	od 1,0 do 106,7 m	od 2,4 do 18,6 m n.p.m.
Średnia miąższość pokładu / Average seam thickness	28,3 m	9,62 m
Odchylenie standardowe miąższości pokładu / Standard deviation of the seam thickness	25,82 m	5,3 m

## WYSTĘPOWANIE TRENDU W ZMIENNOŚCI MODELOWANEGO ZJAWISKA I JEGO WPŁYW NA PROCES MODELOWANIA

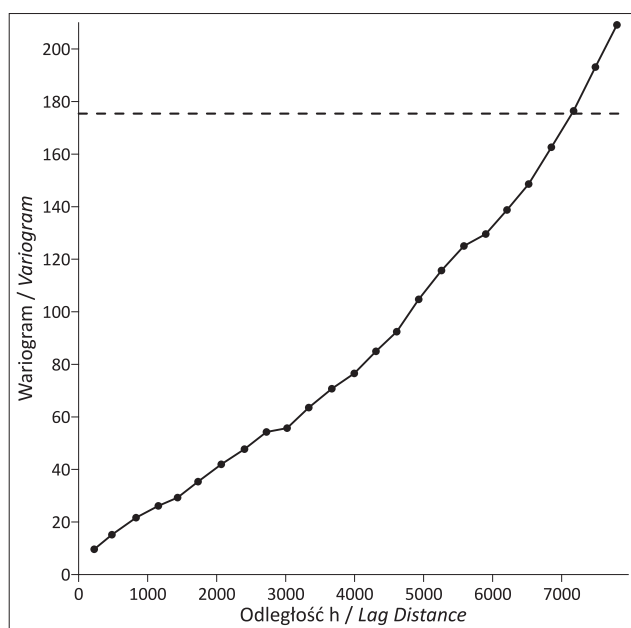
Na ryc. 2 przedstawiłem model powierzchni stropu jednego ze złóż węgla brunatnego. Strop w części południowo-zachodniej osiąga rzędne między 10 a 15 m n.p.m., podczas gdy w części północno-wschodniej dochodzi do -45 m n.p.m. W najbardziej odległych punktach obserwacji różnica wartości rzędnych osiąga 60 m. Pokład wyraźnie zapada w kierunku północno-wschodnim. Wariogram obliczony na podstawie rzędnych zmierzonych w otworach przedstawiłem na ryc. 3. Wykres zachowuje się dziwnie, zupełnie nie tak, jakbyśmy mogli oczekiwać. Zamiast stopniowo zbliżać się do wartości wariancji, wraz z rosnącą odległością  $h$  linia wykresu wznosi się coraz bardziej stromo, niemal wykładniczo. Nawet gdyby dało się dopasować model do wariogramu empirycznego, to na jego podstawie nie da się wyznaczyć zasięgu autokorelacji. Czyżby ten zasięg był nieokreślony? Oczywiście nie. Autokorelacja

zwykle maleje wraz z odległością i w pewnej odległości granicznej całkiem zanika, o czym szerzej pisałem w pierwszym artykule cyklu (Naworyta, 2024a). W kształcie wykresu (ryc. 3) wyraźnie odznacza się obecność tzw. trendu. Regularne zapadanie pokładu w kierunku SW-NE sprawia, że wraz z odległością między parami obserwacji stopniowo rośnie różnica między rzędnymi, wraz z rosnącą odległością różnice są coraz większe. Na wariogramie uwiadczenia się ten wzrost. Czy wobec tego, że nie sposób określić zasięgu autokorelacji, nie da się modelować zjawiska metodą krigingu? Owszem, nie da się, ale czy w związku z tym taki wariogram jest nieprzydatny do dalszych analiz? Nie, ten wariogram w takim właśnie kształcie jest bardzo przydatny. On nam wskazuje, że w zmienności badanego zjawiska występuje trend. To jest bardzo ważna i przydatna informacja.

Aby móc przeprowadzić interpolację metodą krigingu, należy uprzednio wykonać procedurę „odcięcia trendu” od wartości obserwacji. Należy dopasować płaszczyznę, która najlepiej będzie aproksymować wszystkie wartości zbioru.



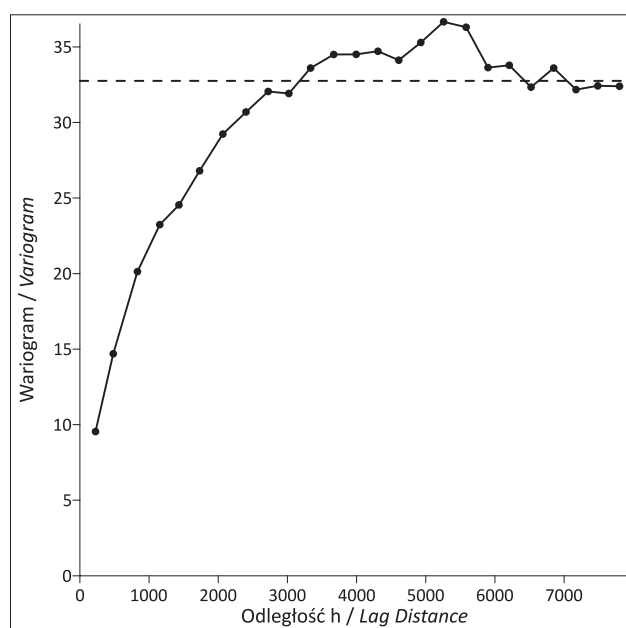
**Ryc. 2.** Model stropu pokładu złoża węgla brunatnego z widocznym wpływem trendu. Model wykonano metodą krigingu z wykorzystaniem funkcji odcięcia trendu  
**Fig. 2.** Model of the top of a lignite deposit with the visible influence of the trend. The model was made by the kriging method using the detrend function



**Ryc. 3.** Wariogram rzędnych stropu pokładu węgla przedstawionego na ryc. 2. Nietypowy kształt wariogramu wskazuje na obecność trendu w zmienności cechy złoża

**Fig. 3.** Variogram of the elevations of the top of the lignite seam shown in Fig. 2. The unusual shape of the variogram indicates the presence of a trend in the variability of the deposit feature

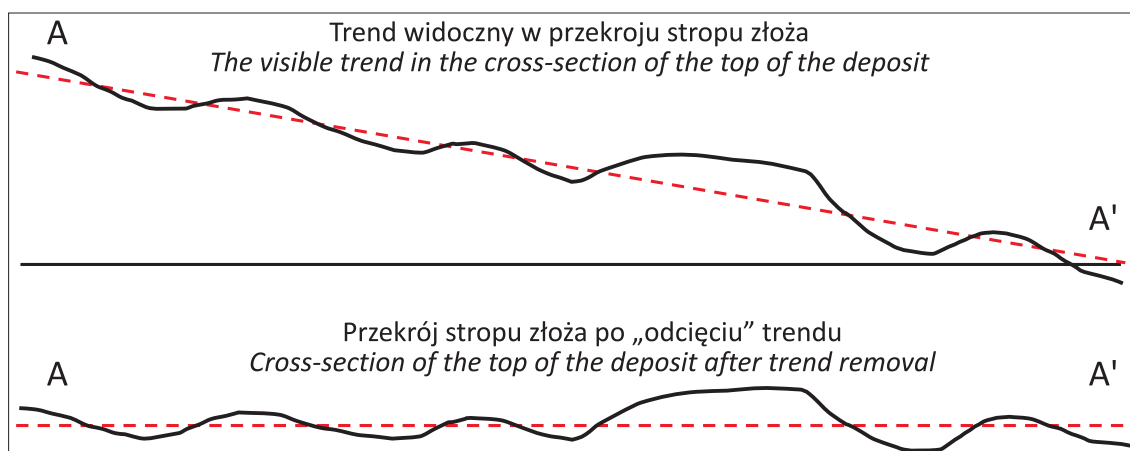
W naszym konkretnym przypadku w przestrzeni ta płaszczyzna będzie pochylona w kierunku północno-wschodnim. Na szczęście nie musimy tego robić „na piechotę”. Pakiety geostatystyczne oferują automatyczną procedurę powstrzymania trendu (*detrend*) albo usuwania trendu (*trend removal*). Wystarczy kliknąć w odpowiednim miejscu. Aby to zrobić, trzeba jednak być świadomym takiej konieczności, czyli należy wykorzystać informację płynącą z kształtu wariogramu (ryc. 3). Chciałbym ponownie podkreślić, jak wielostronnym narzędziem analitycznym jest wariogram. Służy do identyfikacji autokorelacji w zmienności badanego zjawiska, do określenia jej zasięgu, o ile istnienie autokorelacji zostanie uprzednio stwierdzone, do wyznaczenia błędu obserwacji za pomocą wielkości wariancji lokalnej (*nugget effect*) oraz do stwierdzenia, czy w zmienności obserwacji występuje trend.



**Ryc. 4.** Wariogram rzędnych stropu pokładu węgla po zastosowaniu funkcji odciążenia trendu

**Fig. 4.** Variogram of the elevations of the top of a lignite seam after removing the trend

Na ryc. 4 przedstawiłem wariogram rzędnych stropu pokładu z ryc. 2, obliczony po odciążeniu liniowego w formie płaszczyzny dopasowanej do obserwacji. Odciążenie trendu można obrazowo przedstawić jako „wypoziomowanie”. Na schematycznym przekroju A–A' (ryc. 5) przedstawiłem zjawisko trendu w zmienności stropu złoża przed i po odciążeniu komponentu trendu. Dopiero na bazie tak „wypoziomowanych” danych oblicza się wariogram (ryc. 4), który widocznie różni się od pierwotnego (ryc. 3). Przyjrzyjmy się bliżej tym dwóm wykresom. Nie tylko różnią się kształtem, różnią się przede wszystkim wartościami. Wariancja przed odciążeniem trendu (ryc. 3) osiąga wartość ok. 175 m<sup>2</sup> n.p.m. Uwaga, to są rzędne stropu złoża, więc występują w m n.p.m, ale ponieważ w wariogramie z definicji różnice występują w kwadracie, stąd jednostką na wariogramie jest m<sup>2</sup>. Na wariogramie bez trendu (ryc. 4)



**Ryc. 5.** Przekrój A–A' przez strop złoża z ryc. 2 przed i po odciążeniu trendu. Czarną linią oznaczono strop pokładu, linią czerwoną przerywaną przedstawiono linię trendu

**Fig. 5.** Cross-section A–A' through the top of the deposit in Fig. 2 before and after the trend removal. The black line marks the top of the seam, and the red dashed line shows the trend line

wariancja osiąga wartość zaledwie 33 m<sup>2</sup> n.p.m. Po odcięciu trendu różnice między poszczególnymi obserwacjami znacząco się zmniejszyły. Z wariogramu na ryc. 4 wyraźnie można odczytać zasięg autokorelacji  $AR = 3000$  m. Zostanie to wykorzystane w dalszym kroku w procedurze interpolacji metodą krigingu. Uważnego Czytelnika może niepokoić fakt, że przecież po odcięciu trendu pierwotne wartości ulegają zmianie (deformacji), a przecież modelując zjawisko oczekujemy, że będzie ono oddane możliwie najwierniej. Oprogramowanie geostatystyczne na szczęście uwzględnia obecność trendu w procedurze modelowania i ten trend jest ponownie dodany do gridu wynikowego. Wystarczy zaznaczyć odpowiednią opcję w programie. Aby to zrobić, trzeba jednak wiedzieć, że taka konieczność zachodzi.

### WARTOŚĆ ODSTAJĄCA (OUTLIER) W ZBIORZE OBSERWACJI I JEJ WPLYW NA PROCES MODELOWANIA

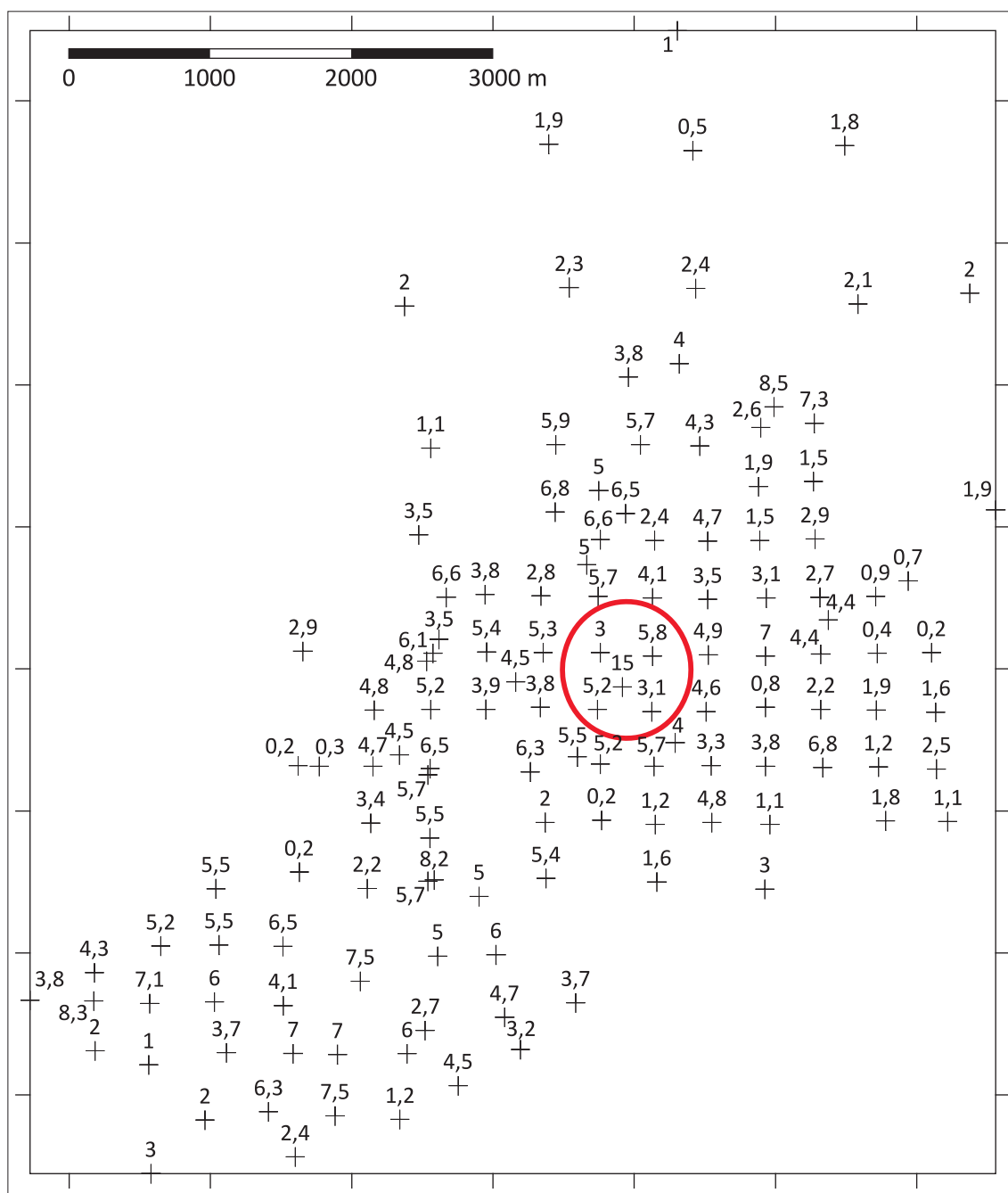
Może się zdarzyć, że wśród obserwacji znajdzie się jedna albo kilka, które pod względem wartości kompletnie nie pasują do reszty. Można je zidentyfikować na histogramie. Diagram wskazuje na zbiór danych, który ma określony rozkład, ale jedna bądź kilka obserwacji wyraźnie odstają od skupionej populacji, są jakby poza głównym zbiorem. Gdyby to była grupa ludzi, to takiego osobnika nazwalibyśmy outsiderem. W statystyce używamy pojęcia *outlier*. Po polsku mówi się o wartościach ekstremalnych albo odstających. Identyfikacja obecności wartości odstających w zbiorze obserwacji jest możliwa na etapie wstępnej analizy danych, o czym pisałem w trzecim odcinku mojego cyklu publikacji (Naworyta, 2024c). To bardzo ważna czynność, jeżeli ją pominiemy i nie uda nam się zidentyfikować wartości ekstremalnych, to ich obecność może nam skutecznie skomplikować, a nawet uniemożliwić dalsze etapy analizy geostatystycznej.

Skąd biorą się takie wartości i jak wpływają na proces modelowania? Istnienie wartości ekstremalnych to często konsekwencja błędu ludzkiego. Załóżmy, że analizujemy zawartość siarki całkowitej w węglu brunatnym. Siarka zwykle towarzyszy pokładowi węgla. Jeżeli na etapie przepisywania danych przy wartości  $S_i = 2,0\%$  pominiemy przecinek, to w zbiorze danych pojawi się wartość  $S_i = 20\%$ . Ktoś, kto zna się na węglach, od razu stwierdzi, że to błąd. Zawartość siarki na poziomie 20% wskazuje bardziej na złoż siarki niż na domieszkę tego pierwiastka w osadach węglowych. Gdy jednak mechanicznie podejmiemy do modelowania, to jego wynik, o ile uda nam się wykonać model, będzie co najmniej podejrzany. Muszę podkreślić, że żaden komputer, nawet najlepszy, nie zastąpi ludzkiego doświadczenia. W przywołanym przykładzie raczej nie może być wątpliwości – geneza tak podejrzane wysokiej zawartości siarki w węglu tkwi w ludzkim błędzie. Można to bez większego ryzyka poprawić, wstawiając brakujący przecinek. Niestety, nie zawsze jest tak łatwo. Do wartości ekstremalnych należy każdorazowo podchodzić indywidualnie. Nie ma jednej, uniwersalnej procedury.

Na przykładzie spróbuję zilustrować, jak wartość ekstremalna potrafi uprzykrzyć modelowanie metodą krigingu. Na ryc. 6 przedstawiłem mapę rozmieszczenia otworów rozpoznawczych z oznaczonymi wartościami miąższości złoża węgla brunatnego. Średnia miąższość tego złoża to  $m_M = 3,96$  m, a odchylenie standardowe  $s_M = 2,28$  m. W centralnej części złoża w jednym z otworów zidentyfikowano

wartość  $M_0 = 15$  m. Miejsce jej występowania oznaczyłem czerwonym okręgiem. Otwór, w którym stwierdzono tak miąższy pokład, leży w bezpośrednim otoczeniu obserwacji o wartościach 3,0; 3,1; 5,2 oraz 5,8 m. Jeżeli nie wykonamy wstępnej analizy statystycznej, najlepiej za pomocą histogramu, to ta wartość może nam umknąć. Nie jest ona widoczna na pierwszy rzut oka. Spójrzmy, jak ta odstająca wartość wpłynie na analizę geostatystyczną. Wariogram wykonany na bazie obserwacji przedstawiłem na ryc. 7. Z jego kształtu wynika, że miąższość pokładu nie wykazuje autokorelacji, jest losowa. Wartości wariogramu oscylują wokół wariancji oznaczonej linią przerywaną. Co więcej, wartości wariogramu są statystycznie wiarygodne. Ta najbliższej osi pionowej została obliczona na bazie aż 81 par obserwacji. Nie ma powodu, aby powątpiewać w jej statystyczną istotność. Świadomy tego operator powinien odstąpić od dalszych analiz geostatystycznych i poprzestać na analizach statystycznych. Jeżeli w zmienności zjawiska nie ma autokorelacji, co wynika z kształtu wariogramu (ryc. 7), to znaczy, że nie ma teoretycznych przesłanek wykonania modelu w postaci mapy izoliniowej. Należałoby uznać, że miąższość tego złoża ma cechy losowe. W tym jednak momencie znowu powinno dojść do głosu doświadczenie geologa. W zmienności miąższości złóż pokładowych najczęściej występuje wyraźny udział nielosowy. Parametr ten zwykle daje się dobrze modelować. Najczęściej w jego zmienności jest wyraźnie widoczna autokorelacja. Dlaczego zatem nie w tym złożu? To powinno nas skłonić do dociekań. Okazuje się, że obserwacja, gdzie stwierdzono wartość ekstremalną  $M_0 = 15$  m, tak znacząco odstaje od pozostałych obserwacji, że biorąc udział w obliczeniach, w parach z wieloma innymi obserwacjami spowodowała podniesienie różnic statystycznych, a więc znacząco wpłynęła na wartości i kształt wariogramu. Jej obecność można porównać do fenomenu samorodka. Wartość  $M_0 = 15$  m na tle obserwacji od 0 do 8,5 m to niemal jak samorodek w złożu złota. A już wiemy, że samorodki powodują podniesienie wartości wariogramu – stąd nazwa *nugget effect* (Naworyta, 2024b).

Jeżeli w procesie wstępnej analizy danych uda nam się zidentyfikować wartość ekstremalną, to możemy spróbować poradzić sobie z jej obecnością i jej negatywnym wpływem. W pierwszym przypadku było to proste. Wartość ekstremalna była wynikiem błędu ludzkiego – zamiast 2,0% wpisano 20%. W kolejnym przykładzie nie możemy być tego pewni. Gdy jednak spojrzymy na najbliższe otoczenie wartości ekstremalnej, to możemy podejrzewać, że i tu wartość została wpisana błędnie. Ponieważ ja sam jestem autorem tego przekłamania, więc wiem, że zamiast  $M_0 = 5,3$  m, co było wartością poprawną, wpisałem do tabeli wartość  $M_0 = 15$  m. W realnym przypadku nie zawsze możemy być pewni genezy błędu. Nie powinniśmy zbyt lekkomyślnie pozbywać się takiej kłopotliwej obserwacji ze zbioru danych. Spróbujmy jednak obliczyć wariogram, pomijając wartość ekstremalną. Możemy to zrobić w tabeli danych przez jej usunięcie, albo wprost w procedurze obliczania wariogramu. *Software* najczęściej umożliwia ograniczenie zakresu wartości w procesie liczenia wariogramu. Na ryc. 8 przedstawiłem wariogram po wyłączeniu ze zbioru danych wartości ekstremalnej. Wygląda on „przyzwoicie” – zgodnie z oczekiwaniem umożliwi identyfikację autokorelacji i jej zasięgu. Wariancja lokalna (*nugget effect*) przybiera relatywnie niską wartość 2,0 m<sup>2</sup>. Możemy dopasować model wariogramu (tu: złożony z *nugget effect* i modelu sferycznego,

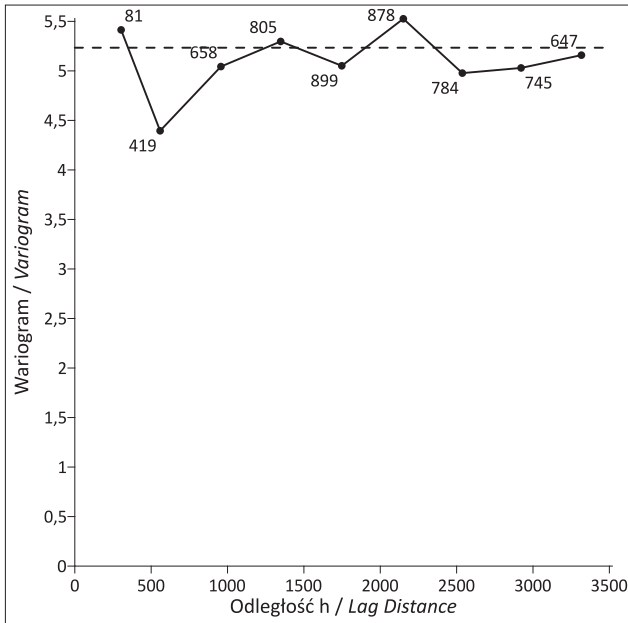


**Ryc. 6.** Otwory rozpoznawcze z etykietą miąższości złoza. W centrum okręgu znajduje się otwór, w którym miąższość osiąga wartość  $M = 15$  m, wartość ta znacznie odstaje od pozostałych

**Fig. 6.** Boreholes with the deposit thickness labels. There is a borehole in the red circle with a thickness of  $M = 15$  m, which is significantly different from the others

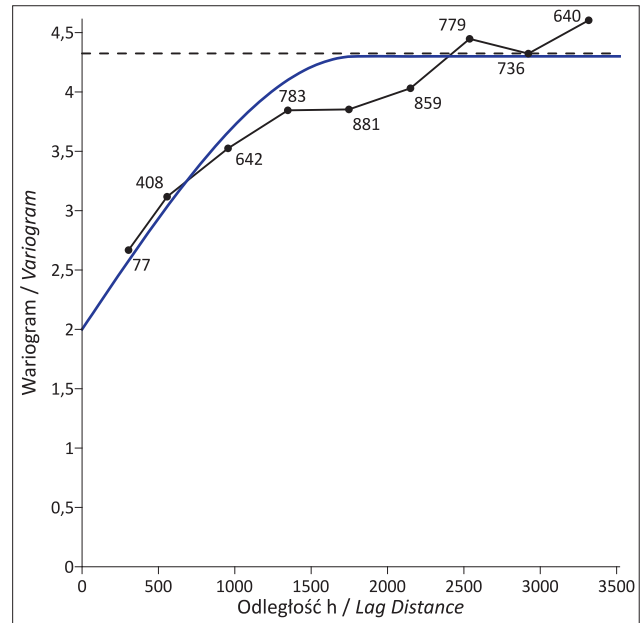
linia niebieska) i wykonać model miąższości złoza, wykorzystując wszystkie zalety procedury krigingu. Zwróćmy uwagę, że wyeliminowanie wartości ekstremalnej tylko nieznacznie zmieniło statystykę zbioru danych. W tabeli 2 zestawilem podstawowe cechy statystyczne. Na wariogramach również można odczytać, że wariancja pokazana linią przerywaną na ryc. 8 nieznacznie się obniżyła z  $5,23 \text{ m}^2$  do  $4,32 \text{ m}^2$ . Różnica ta jednak nie rzuca się w oczy. W miarach statystycznych nie widać wyraźnego wpływu wartości ekstremalnej  $M_0 = 15$  m. Jednak wyraźnie zaznacza się ona w kształcie wariogramu ryc. 7 i 8. Czy koniecznie powinienem ponownie podkreślać, jak ważnym i wielostronnym narzędziem analitycznym jest wariogram?

Wcześniej stwierdziłem, że nie zawsze możemy być pewni co do tego, że *outlier* to wynik błędu popełnionego w procesie dokumentowania złoza lub przepisywania danych z kartki do komputera itp. Nie można pochopnie pozbywać się tych wartości ze zbioru danych tylko dlatego, że to ułatwia modelowanie. Ja rekomenduję, aby w procedurze modelowania chwilowo usunąć tę wartość, co umożliwi zastosowanie metod geostatystycznych. Jeżeli jednak nie jesteśmy pewni co do genezy wartości ekstremalnej, to należałoby ją pokazać osobno na wykonanym modelu. Tak będzie uczciwie. W ten sposób swobodę interpretacji pozostawiamy innym użytkownikom modelu, nie pozbawiając ich informacji pierwotnej.



**Ryc. 7.** Wariogram miąższości pokładu węgla brunatnego. Z jego kształtu wynika, że zmienność badanego parametru ma charakter losowy. Taki mylący kształt wariogramu to efekt obecności jednej tylko wartości ekstremalnej w zbiorze obserwacji

**Fig. 7.** Variogram of the thickness of the lignite seam. Its shape shows the random variability of the examined parameter. This misleading shape of the variogram is the result of the presence of only one outlier in the set of observations



**Ryc. 8.** Wariogram miąższości pokładu złoża po usunięciu wartości ekstremalnej ze zbioru danych. Z jego kształtu można odczytać istnienie autokorelacji i jej zasięg, a wariancja lokalna przyjmuje relatywnie małą wartość

**Fig. 8.** Variogram of the thickness of the deposit after removing the extreme value from the data set. The existence of autocorrelation and its range can be read from its shape, and the nugget effect is relatively low

**Tab. 2.** Podstawowe miary statystyczne zbioru obserwacji w wariancie z wartością odstającą  $M_0 = 15$  m oraz bez tej wartości  
**Table 2.** Basic statistical measures of the set of observations in the variant with and without the outlier value  $M_0 = 15$  m

Cecha statystyczna <i>Statistical feature</i>	Dane z wartością odstającą $M_0 = 15$ m <i>Data with outlier <math>M_0 = 15</math> m</i>	Dane oczyszczone z wartości odstającej <i>Data cleaned of outlier</i>
Liczba obserwacji / <i>Number of observations</i>	131	130
Wartość najmniejsza / <i>Minimum value</i>	0,2 m	0,2 m
Wartość największa / <i>Maximum value</i>	15,0 m	8,5 m
Wartość średnia / <i>Average</i>	3,96 m	3,88 m
Wariancja / <i>Variance</i>	5,23 m <sup>2</sup>	4,32 m <sup>2</sup>

## PODSUMOWANIE

W niniejszym artykule, który stanowi uzupełnienie cyklu trzech artykułów opublikowanych w *Przeglądzie Geologicznym* w 2024 r. pod wspólną nazwą *Geostatystyka dla niematematyków – Przewodnik praktyczny*, opisałem kilka istotnych zagadnień, z którymi można się zetknąć w procesie geostatystycznego modelowania złóż lub innych zjawisk przestrzennych. Brak stacjonarności, obecność trendu czy obecność obserwacji o wartościach odstających to problemy, które mogą skutecznie utrudnić lub wręcz uniemożliwić proces modelowania. Na kilku obrazowych przykładach starałem się pokazać, jak można te problemy zidentyfikować i w jaki sposób można sobie z nimi poradzić.

Chciałbym, aby głównym przesłaniem, jakie z tego właśnie artykułu powinno wypływać, było to, że żaden program komputerowy (nawet najlepszy), żaden algorytm i procedura wciąż jeszcze nie zastąpią ludzkiej wiedzy i doświadczenia. Komputer dla większości z nas jest „czarną skrzynką” i nawet nie wiemy, bo nie musimy wiedzieć, co dzieje się w tym urządzeniu po kliknięciu OK. Dlatego

podkreślam, że import danych do programu komputerowego powinien być poprzedzony wstępną, ale szczegółową, analizą zbioru obserwacji. Ten arcyważny etap modelowania jest żmudny, nieefektywny, jego wynikiem nie są kolorowe obrazy, na które niecierpliwie czekamy. Jednak ta właśnie wstępna szczegółowa ocena danych może nas uchronić przed popełnieniem grubych błędów. Taki właśnie generalny wniosek powinien wypływać z każdego przedstawionego przeze mnie w tym artykule przykładu.

## LITERATURA

- NAWORYTA W., ZAJĄCZKOWSKI M. 2018 – Ocena możliwości eksploatacji zasobów złoża węgla brunatnego Bełchatów – pole Kamięńsk. Biul. Państw. Inst. Geol., 472.  
NAWORYTA W. 2024a – Geostatystyka dla niematematyków. Przewodnik praktyczny – część I. Prz. Geol., 72 (7): 341–349.  
NAWORYTA W. 2024b – Geostatystyka dla niematematyków. Przewodnik praktyczny – część II. Prz. Geol., 72 (8): 391–398.  
NAWORYTA W. 2024c – Geostatystyka dla niematematyków. Przewodnik praktyczny – część III. Prz. Geol., 72 (9): 451–459.

Praca wpłynęła do redakcji 15.11.2024 r.  
Akceptowano do druku 10.01.2025 r.