

Geostatystyka dla niematematyków. Przewodnik praktyczny – część III

Wojciech Naworyta¹



Geostatistics for non-mathematicians. Practical guide – part III. *Prz. Geol.*, 72: 451–459; doi: 10.7306/2024.26

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 would like to introduce potential non-mathematicians to the most important methods and tools from the arsenal of spatial statistics and how to use them properly. I will indicate the areas where they can be used, explain whether they can always be used, show what decisions should be made during calculations and how to interpret the obtained results. It will not be a compendium, but rather a pocket guide facilitating the reader's first contact with geostatistics. In the third article of my series, I will present the kriging procedure – an interpolation method based on geostatistical assumptions. I will list the most important activities and decisions that need to be made when using this method. Creating a model using the kriging method is the basic goal of using geostatistical methods, so in this article I will make extensive use of the content I presented earlier, from interpolation through autocorrelations to the variogram. The latter serves as an important tool in kriging. Using the example of a lignite deposit model, I will show how to interpret the model and the model credibility map.

Keywords: geostatistics, variogram, interpolation, kriging, kriging standard deviation

Geostatystyka to nie filozofia, służy bardzo konkretnym celom. Efektem końcowym jest mapa. Brzmi banalnie, ale to prawda. Od początku chodzi wyłącznie o to, aby z danych, których zwykle jest za mało, pochodzą z nieregularnej sieci otworów albo np. próbek glebowych, a w dodatku nie są bezbłędne, wykonać mapę zmienności interesującego nas zjawiska. Ta mapa ma nieść informację, która może być podstawą do podejmowania ważnych decyzji inwestycyjnych lub naukowych, dlatego powinna przedstawiać zjawisko możliwie dokładnie, bo od tej dokładności będzie zależeć trafność podejmowanych decyzji. Zamiast słowo mapa powinienem właściwie używać pojęcia model. To lepsza nazwa, bo kreślenie mapy na podstawie dyskretnych punktowych próbek to modelowanie, nie mające wiele wspólnego z kartografią.

Model złoża może być przydatny na długo przed podjęciem decyzji o eksploatacji kopaliny. Może posłużyć w procesie wyceny zasobów, a ta może być podwaliną zastawą hipotecznego, dzięki któremu inwestor pozyska kredyt. Na podstawie modelu złoża można opracować dokumenty służące jego zagospodarowaniu: *prefeasibility study*, *feasibility study*, *bussines plan*, projekt zagospodarowania złoża, plan ruchu. To tylko wybrane przykłady z górnictwa, wskazujące na znaczenie modelu złoża. Model, mimo wszystkich wad i ograniczeń wynikających z niedoskonałości danych źródłowych, powinien być możliwie najbardziej dokładny. Dlatego właśnie tracimy tyle czasu, aby opracować najlepszy, najdokładniejszy model interesującego nas zjawiska. Szerzej o meandrach modelowania wypowiedziałem się we wcześniejszej publikacji (Naworyta, 2017).

Ten artykuł nie został pomyślany jako instrukcja obsługi konkretnego programu geostatystycznego, dlatego nie będę przywoływał procedur występujących w poszczególnych programach. Zamiarem moim jest nakreślenie pewnego schematu postępowania, który cechuje metodę a nie konkretny program. Schemat ten jest wspólny dla

wszystkich programów, które korzystają z metod opartych na założeniach geostatystycznych. Ja sam korzystam z programów *Surfer (Golden Software)* i *Isatis (Geovariances)*. Ten pierwszy jest bardziej uniwersalny, a drugi specjalistyczny i wyposażony w wiele typowo geostatystycznych narzędzi, często bardzo wysublimowanych. Jeden i drugi umożliwia przeprowadzenie analiz, o których mowa w moim tekście. Rysunki w tym artykule w większości zostały wykonane w programie *Surfer*.

Pierwszy artykuł mojego cyklu w dużej mierze poświęciłem interpolacji. Dzisiaj temat ten rozszerzę, koncentrując się na interpolacji metodą krigingu, czyli sztandarowej metodzie z arsenału geostatystyki. Jest wiele powodów, dla których kriging można uznać za metodę lepszą niż inne niegeostatystyczne metody interpolacji. Tytuł artykułu zwalnia mnie jednak z konieczności udowadniania tej wyższości przez odwoływanie się do aparatu matematycznego. Czytelnikowi niech wystarczy argument, że w wyniku modelowania metodą krigingu oprócz modelu zmienności parametru otrzymuje się swoisty *bonus* w postaci mapy błędu tego modelu. Otrzymujemy model oraz informację, na ile ten model jest wiarygodny. Ale o tym potem. Jeżeli to ma być przewodnik praktyczny, to zacznijmy od początku.

WSTĘPNA ANALIZA DANYCH

Zanim zaimportujemy dane do programu i klikniemy *OK* albo *skip to end*, powinniśmy się im krytycznie przyjrzeć. To arcyważna czynność, która pozwoli nam uniknąć poważnych błędów już na samym początku pracy. Nie wolno jej pominąć. W swojej praktyce wielokrotnie boleśnie przekonałem się o tym, zwłaszcza wtedy, gdy zachwycając się modelem dostrzegałem kardynalne błędy wynikające z niedostatecznej wstępnej analizy materiału pomiarowego.

Dobrze jest policzyć najważniejsze miary statystyczne, czyli: średnią, odchylenie standardowe, najwyższą i najniższą wartość, ale także liczbę obserwacji, pamiętając o tym,

¹ 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

że w przypadku badań geologicznych nie zawsze każdy parametr złoża został oznaczony w każdym otworze. Można to wykonać w *Excelu*, który oferuje wszystkie te funkcje. To na początek wystarczy. Można też spojrzeć na histogram, z którego na pierwszy rzut oka da się odczytać najważniejsze cechy statystyczne zbioru danych. Wymienione przeze mnie programy do modelowania dysponują modulem statystycznym, za pomocą którego bez większych trudności da się odczytać dziesiątki miar statystycznych. Dla nas w zupełności wystarczą te podstawowe. Oprócz analizy statystycznej dobrze jest się przyjrzeć, jak wygląda sieć otworów (opróbowania) w przestrzeni.

Na ryc. 1 przedstawiłem mapę otworów odwierconych w złożach węgla brunatnego. Skupione są one w dwóch wyraźnie oddzielonych od siebie grupach. Dane były zestawione w jednej tabeli. Gdybym nie wykonał tej mapy, to rozpatrywałbym je jako jedno złożo. Nie wolno mi tego robić. Wyraźnie widać, że chodzi o dwa osobne byty geologiczne. Gdybym pominął analizę danych i wykonał model, to cechy jednego ze złóż byłyby wykorzystane do modelowania drugiego i odwrotnie. A przecież nie ma fizycznych przesłanek, aby takie wzajemne oddziaływanie mogło mieć miejsce.

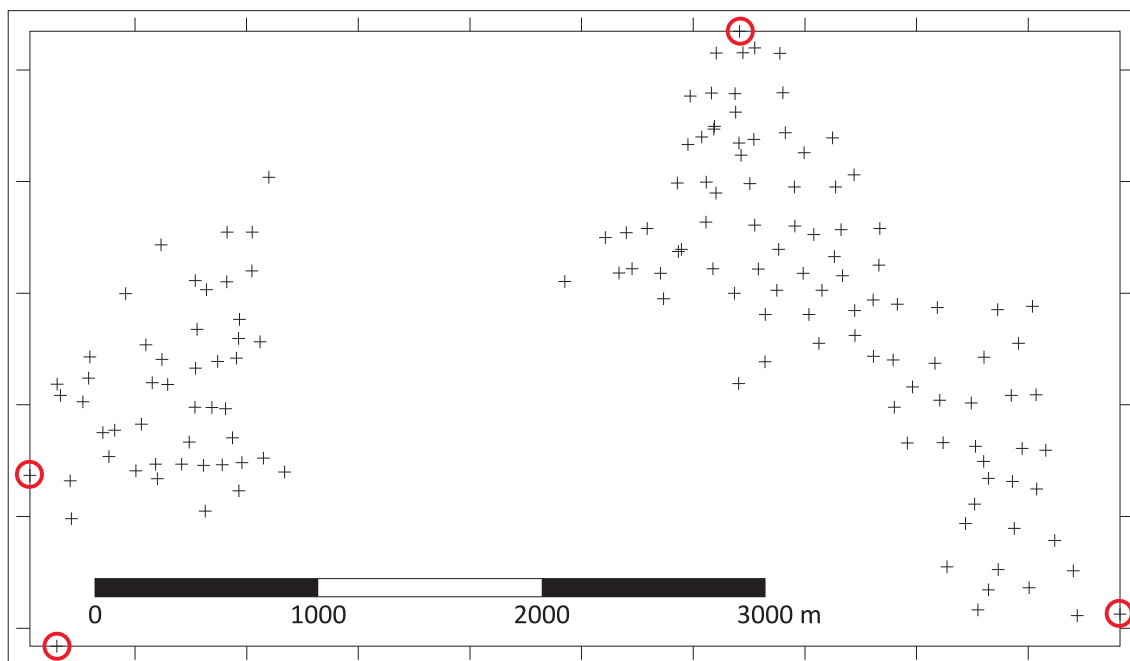
Na ryc. 2. otwory wywiercone w złożu węgla brunatnego są rozmieszczone nieregularnie. Sieć otworów w północnej części złoża jest wyraźnie rzadsza niż ta w części środkowej. Odległość między otworami znacząco odstaje od średniej. Z pewnością jest tego jakiś powód. Gdy się przyjrzy dany bliżej, to okaże się, że w wielu otworach, a szczególnie tych na północy, nie oznaczono parametrów jakościowych węgla. Na mapie pokazałem etykiety wartości opałowej przy otworach, gdzie ten najważniejszy parametr paliwa kopalnego został oznaczony. Tych pomiarów jest znacznie mniej niż otworów. We wszystkich otworach oznaczono tylko parametry strukturalne pokładu, tj. rzędne stropu i spągu. Powód okazał się banalny. W części północnej pokład jest tak cienki, że nie opłaca się go wydobywać, a jeżeli tak, to nie ma powodów, aby w takich otworach

wykonywać kosztowne oznaczenia jakości kopaliny (por. ryc. 6). Takie informacje należy dostrzec już na początku, zanim przedziemy do procesu modelowania.

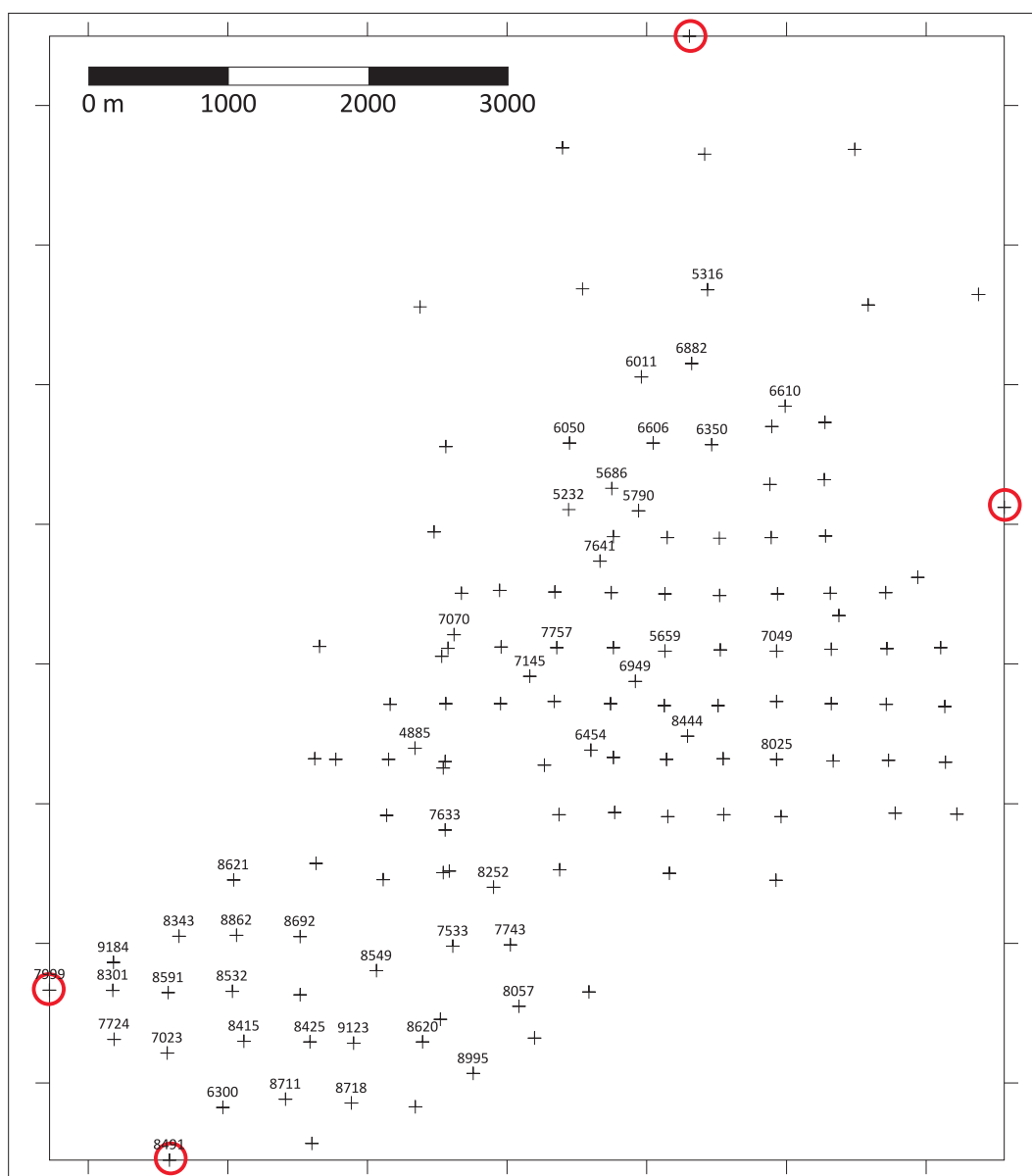
Z przedstawionych map można też odczytać, jaka jest najmniejsza, średnia i największa odległość między otworami. To w analizach geostatystycznych informacja bardzo istotna. Na ryc. 2 tylko kilka otworów jest oddalonych od siebie o odległość mniejszą niż średnia odległość między otworami, która tu wynosi ok. 400 m. Na ich podstawie będzie można oszacować wartość wariacji lokalnej (*nugget effect*). Zwróć na to uwagę, analizując wariogram. W ramach wstępnej analizy danych dobrze jest poznać ich zakres, czyli rozpiętość wartości – od najmniejszych do największych. Oczekujemy, że w wykonanym przez nas modelu te wartości będą reprezentowane. Niestety, efekt wygładzania (*smoothing effect*), typowy dla wszystkich metod interpolacyjnych, sprawi, że w modelu będą one miały mniejszy zakres. Wartość najmniejsza będzie nieco większa niż w danych wyjściowych, a największa będzie miała mniejszą wartość. Efekt wygładzania i jego źródła opisałem w pierwszym artykule cyklu (Naworyta, 2024a). Bez wstępnej obserwacji danych nie będziemy tego wiedzieć, a to przecież bardzo ważna informacja. Model powinien być możliwie wierny danym źródłowym. Nie będzie, ale powinniśmy być tego świadomi.

WARIOGRAM I MODEL WARIOGRAMU

Po wstępnej analizie danych można przejść do dalszych czynności. Wykonujemy wariogram (ryc. 3). Z wykresu dowiadujemy się, czy występuje autokorelacja i czy w związku z tym dalsze czynności związane z modelowaniem mają sens. Jeśli tak, to do wariogramu empirycznego należy dopasować model, z którego odczytamy najważniejsze parametry: wartość wariacji lokalnej (*nugget effect*) oraz zasięg autokorelacji (*autocorrelation range – AR*). Pojęcia te szerzej skomentowałem w drugim artykule mojego cyklu (Naworyta, 2024b). Na wariogramie, który posłuży



Ryc. 1. Mapa otworów rozpoznawczych dwóch sąsiadujących złóż
Fig. 1. Map of geological boreholes for two adjacent deposits



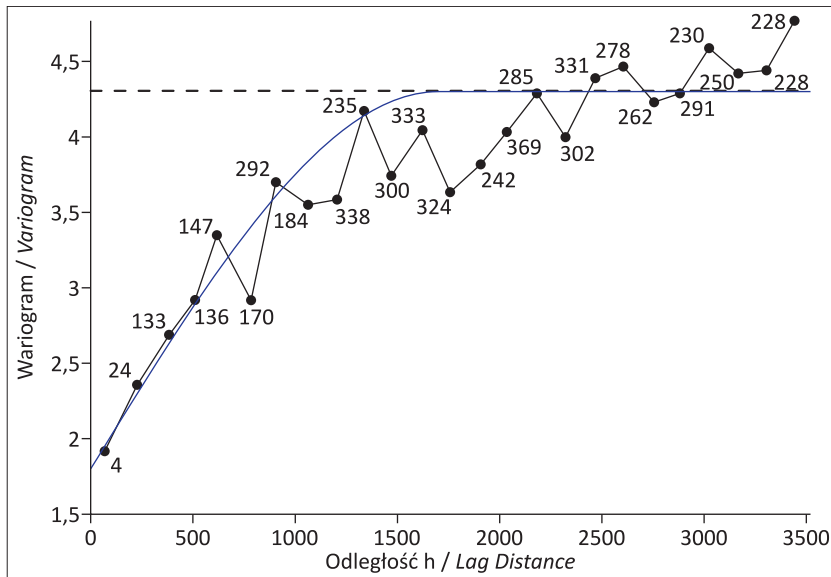
Ryc. 2. Przykład rozmieszczenia otworów rozpoznawczych w złożu węgla brunatnego
Fig. 2. Example of the arrangement of geological boreholes on a lignite deposit

do modelowania miąższości pokładu złoża węgla brunatnego (ryc. 3) wartość wariancji lokalnej wynosi $\sigma^2 = 1,8 \text{ m}^2$, a zasięg autokorelacji $AR = 1700 \text{ m}$. Obok punktów reprezentujących wartości wariogramu empirycznego występują liczby. Informują one, ile par obserwacji wzięło udział w obliczaniu tych konkretnych wartości. Zwróćmy uwagę na punkt najbliższy osi pionowej. Tylko 4 pary zostały wykorzystane do obliczenia wartości wariogramu dla odległości wynoszącej w przybliżeniu $h = 100 \text{ m}$. To bardzo mało i statystyczna wartość takiego obliczenia jest znikoma. Mimo to w tym konkretnym przypadku punkt ten doskonale wpisuje się w trend. Cztery punkty z zakresu odległości od $h = 0 \text{ m}$ do $h = 500 \text{ m}$ leżą niemal w linii prostej. Dopasowany model sferyczny prawie dokładnie pokrywa się z położeniem tych punktów. Nie jest to jednak reguła. Właśnie tu, na początku wariogramu, gdzie występuje niewiele par reprezentujących najmniejsze odległości między obserwacjami, wartość wariogramu często odbiega od głównego trendu. Nie należy się tym szczególnie przejmować. Model wariogramu trzeba dopasować tak, aby

reprezentował główny trend wyznaczony przez punkty wariogramu empirycznego, obliczone na podstawie większej liczby par. Dobrze jest jednak wyświetlić na wariogramie liczbę par, aby wiedzieć, jak statystycznie istotne są poszczególne wartości wariogramu, szczególnie te dla małych odległości h , bo bliskich obserwacji zwykle jest niewiele.

INTERPOLACJA METODĄ KRIGINGU

Teraz można przejść do właściwej interpolacji metodą krigingu. Jest wiele odmian tej metody. Opisywana przeze mnie to kriging zwyczajny (*ordinary kriging*). W pierwszym kroku należy zdefiniować siatkę interpolacyjną (*grid*). Programy geostatystyczne najczęściej podpowiadają operatorowi geometrię gridu, wzorując się na rozmieszczeniu danych w przestrzeni. Najczęściej zasięg gridu odpowiada skrajnym współrzędnym obserwacji. W wyniku automatycznego dopasowania siatki do danych skrajne obserwacje będą leżały na krawędziach mapy.



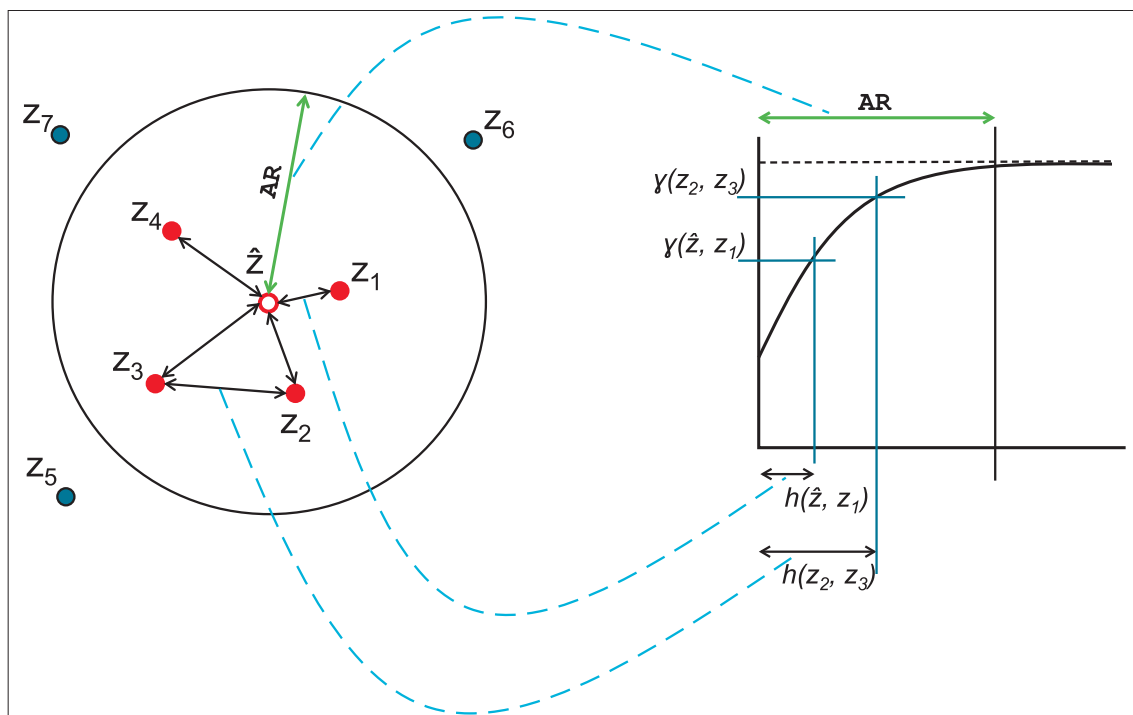
Ryc. 3. Wariogram empiryczny miąższości pokładu złoża węgla brunatnego z dopasowanym modelem (niebieska linia)

Fig. 3. Empirical variogram of the thickness of the lignite deposit with the fitted model (blue line)

Na ryc. 1 i 2 kolorem czerwonym oznaczyłem obserwacje, które leżą na krawędziach mapy. Aby tego uniknąć, można rozszerzyć siatkę interpolacyjną na wschód, zachód, północ i południe. Wtedy wszystkie miejsca poboru próbek będą leżały wewnątrz wykonywanej przez nas mapy. Dobrze jest też zadbać o to, aby odległość w kierunkach X i Y była wielokrotnością rozmiaru oczka siatki interpolacyjnej. Jeżeli rozmiar pojedynczego oczka siatki interpolacyjnej ustalimy na 50×50 m, to rozpiętość siatki w obydwu kierunkach powinna być podzielna przez 50. To kosmetyka, ale przy-

datna w praktyce modelowania. Jeżeli będziemy chcieli wykonywać operacje matematyczne na gridach, to siatki interpolacyjne biorące udział w operacjach muszą być pod względem rozmiarów identyczne. Więcej informacji na temat gridu zamieściłem w pierwszym artykule cyklu (Naworyta, 2024a).

Mamy siatkę interpolacyjną, wariogram i model wariogramu – możemy przejść dalej. Importowanie danych zostało już wykonane, bo na ich podstawie program zaproponował geometrię siatki interpolacyjnej. W następnym kroku trzeba zaimportować wariogram, a raczej model wariogramu, który wcześniej został dopasowany do wariogramu empirycznego. Program najczęściej sam o to zapyta, ale można to przeoczyć. Jeżeli nie zaimportujemy naszego modelu, to program zrobi to za nas, ale będzie to model liniowy, nie oddający rzeczywistej zmienności danych źródłowych. W kolejnym kroku padnie ważne pytanie o otoczenie (*neighbourhood, search ellipse*). Niezależnie od tego, jak to zostanie nazwane, chodzi o otoczenie, najczęściej w granicach okręgu, rzadziej elipsy, wokół pojedynczego węzła siatki interpolacyjnej. Do estymacji wartości w węzle siatki interpolacyjnej program wykorzysta tylko obserwacje znajdujące się wewnątrz zdefiniowanego przez nas otoczenia. To zagadnienie powinienem dokładniej wyjaśnić. Wybranie zakresu otoczenia wymaga od nas świadomej decyzji. Na ryc. 4 pokazałem ideę krigingu. Rysunek niesie wiele informacji, ale skoncentrujemy się teraz na zdefiniowaniu samego otoczenia. Nie wchodząc w szczegóły matematycznej



Ryc. 4. Schemat ilustrujący zastosowanie informacji z wariogramu w procedurze szacowania metodą krigingu

Fig. 4. A diagram illustrating the use of information from a variogram in the kriging estimation procedure

kuchni krigingu, czego na łamach niniejszego artykułu szczególnie nie wolno mi robić, idea krigingu sprowadza się do tego, aby wyliczyć wartości wag w_i , które przemnożone przez wartości obserwacji z_i pozwolą na oszacowanie wartości w węźle siatki interpolacyjnej \hat{z} według prostego wzoru:

$$\hat{z} = w_1 z_1 + w_2 z_2 + \dots + w_n z_n \quad [1]$$

$$\sum_1^n w_i = 1 \quad [2]$$

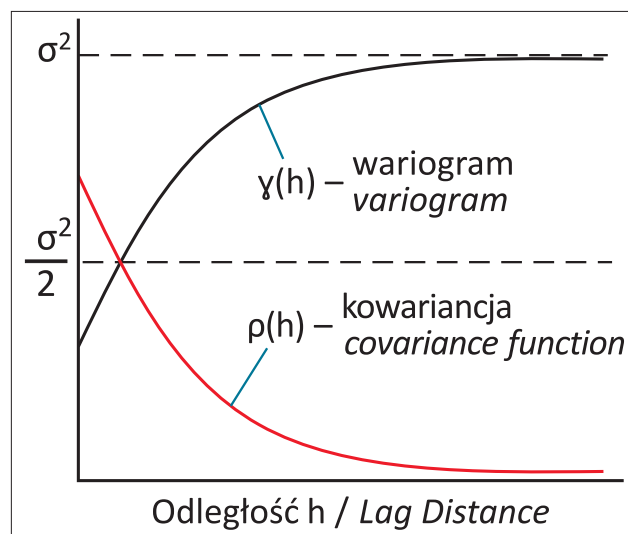
Należy zadać pytanie, kiedy oszacowanie wartości będzie dokładniejsze – czy gdy do interpolacji wykorzystamy więcej obserwacji z_i , czy gdy będzie ich mniej? Pytanie to można postawić inaczej, bardziej precyzyjnie – które obserwacje powinny brać udział w obliczeniach? Otóż tylko te, które leżą w zasięgu autokorelacji $h = AR$. Tylko one przecież mają jakiś związek, jakąś relację z wielkością szacowaną. Jeżeli obserwacje z_i w stosunku do wartości szacowanej leżą poza zasięgiem autokorelacji, to w stosunku do niej są niezależne, są losowe. Więcej o autokorelacji napisałem w pierwszym artykule cyklu (Naworyta, 2024a). Promień otoczenia należy odczytać z modelu wariogramu, jest to zasięg autokorelacji $h = AR$, czyli odległość, którą sami wybraliśmy, aby dopasować model do wariogramu empirycznego. W wariogramie na ryc. 3 zasięg autokorelacji wynosi $AR = 1700$ m. W programach, które stosuję, wielkość otoczenia nie przenosi się automatycznie z wariogramu do procedury krigingu. Bardzo dobrze, bo w tym obszarze operator powinien mieć wolną rękę. Definiując otoczenie krigingu, wpisuję na ogół wielkość mniejszą od AR przynajmniej o ok. 30%. Dlaczego? Bo zależność par obserwacji odległych od siebie o odległość równą AR jest bliska losowej. Spójrzmy na wariogram na ryc. 3. Wartości wariogramu dla odległości od $h = 1200$ do $h = 1700$ m są bliskie wariancji wyznaczonej poziomą, przerywaną linią. Lepiej będzie, gdy do szacowania wartości w węzłach siatki interpolacyjnej zostaną użyte obserwacje bliższe. Tu trzeba wypracować swój własny złoty środek. Może się okazać, że tych obserwacji bliższych nie jest zbyt wiele, więc duże zawężanie zakresu otoczenia też nie jest wskazane. Jedno jest pewne – obserwacje odległe od miejsca szacowania o odległość większą od AR nie powinny być w ogóle brane pod uwagę.

A co się stanie, gdy zdefiniujemy zasięg bardzo szeroko albo użyjemy wszystkich obserwacji do szacowania każdego węzła siatki? Czy wynik procedury krigingu będzie całkiem zły? Nie, będzie nieco gorszy, ale nie będzie całkiem zły. Zawsze przecież największe wagi będą przypisane do obserwacji najbliższych, pozostałe, dalsze obserwacje będą miały znikomy wpływ, ale jednak niezzerowy. Z logicznego założenia metody wynika, że suma wag zawsze powinna wynieść 1 [2]. Gdyby była większa od 1 to wartość oszacowana byłaby przeszacowana, a gdyby suma wag była mniejsza od 1, to wartość byłaby niedoszacowana. Wartość 1 ma tu wymiar 100%. Każda obserwacja przemnożona przez wagę wnosi jakiś udział do oszacowania wartości, razem wnoszą 100%. Wynika z tego, że te odległe i liczne obserwacje błędnie albo nieświadomie użyte do estymacji, leżące poza zasięgiem autokorelacji AR będą miały przypisane małe wartości wag, bliskie zeru, bo suma wszystkich wag musi być równa 1. Jeżeli tak, to po co je w ogóle uwzględniać w obliczeniach? Efektem

negatywnym nadmiernego rozszerzenia otoczenia będzie pewne dodatkowe wygładzenie modelu, wystąpi rodzaj uśrednienia wartości szacowanych. W opozycji do tego, jeśli ograniczymy promień otoczenia do odległości mniejszej nawet niż AR , to model będzie bardziej zróżnicowany. Wyraźniejsze będą lokalne różnice. A przecież w całej tej geostatystyce chodzi właśnie o to, aby model możliwie dokładnie odzwierciedlał zmienność wynikającą z danych źródłowych.

Zanim klikniemy OK w oczekiwaniu na wynik, jeszcze kilka słów na temat, w jaki sposób są obliczane wagi w_1, w_2, \dots, w_n . Jak w wielu innych algorytmach interpolacyjnych zależą one od odległości obserwacji z_i od miejsca wartości szacowanej \hat{z} . To dość oczywiste. Jeżeli u podstaw interpolacji leży zjawisko autokorelacji, to zależność wag od odległości jest tego konsekwencją. Istotne pytanie brzmi – jaka jest ta zależność, bo nie jest ona po prostu liniowa. Otóż ta zależność jest wyrażona kształtem wariogramu, który został obliczony na bazie obserwacji i użyty w procedurze krigingu, a ściślej mówiąc – modelem wariogramu. Zakłada się, że statystyczna zależność stwierdzona pomiędzy obserwacjami, która została wyrażona wariogramem empirycznym, powinna zachodzić również w miejscach, w których nie zostały wykonane obserwacje, czyli w węzłach siatki interpolacyjnej.

Wagi oblicza się, rozwiązując równanie macierzowe. W tym równaniu są wykorzystane nie tylko zależności między obserwacjami z_1, z_2, \dots, z_n a wartością szacowaną \hat{z} , ale również wzajemne zależności między wszystkimi obserwacjami w zdefiniowanym otoczeniu. W procedurze krigingu są wykorzystywane wszystkie dostępne informacje – wartości w miejscach obserwacji, ich położenie oraz ich wzajemne relacje. Więcej informacji z materiału pomiarowego nie da się uzyskać. Na ryc. 4 przedstawiłem schemat ideowy zależności pomiędzy modelem wariogramu i otoczeniem węzła siatki interpolacyjnej, w którym przebiega szacowanie wartości \hat{z} . Zasięg autokorelacji AR wyznacza promień otoczenia. W szacowaniu wartości biorą udział obserwacje: z_1, z_2, z_3 i z_4 , ale już nie obserwacje: z_5, z_6 i z_7 , bo leżą poza otoczeniem, poza zasięgiem AR . Waga w_1 potrzebna do obliczenia wpływu obserwacji z_1 na



Ryc. 5. Schemat zależności między wariogramem i kowariancją
Fig. 5. Diagram of the relationship between variogram and covariance function

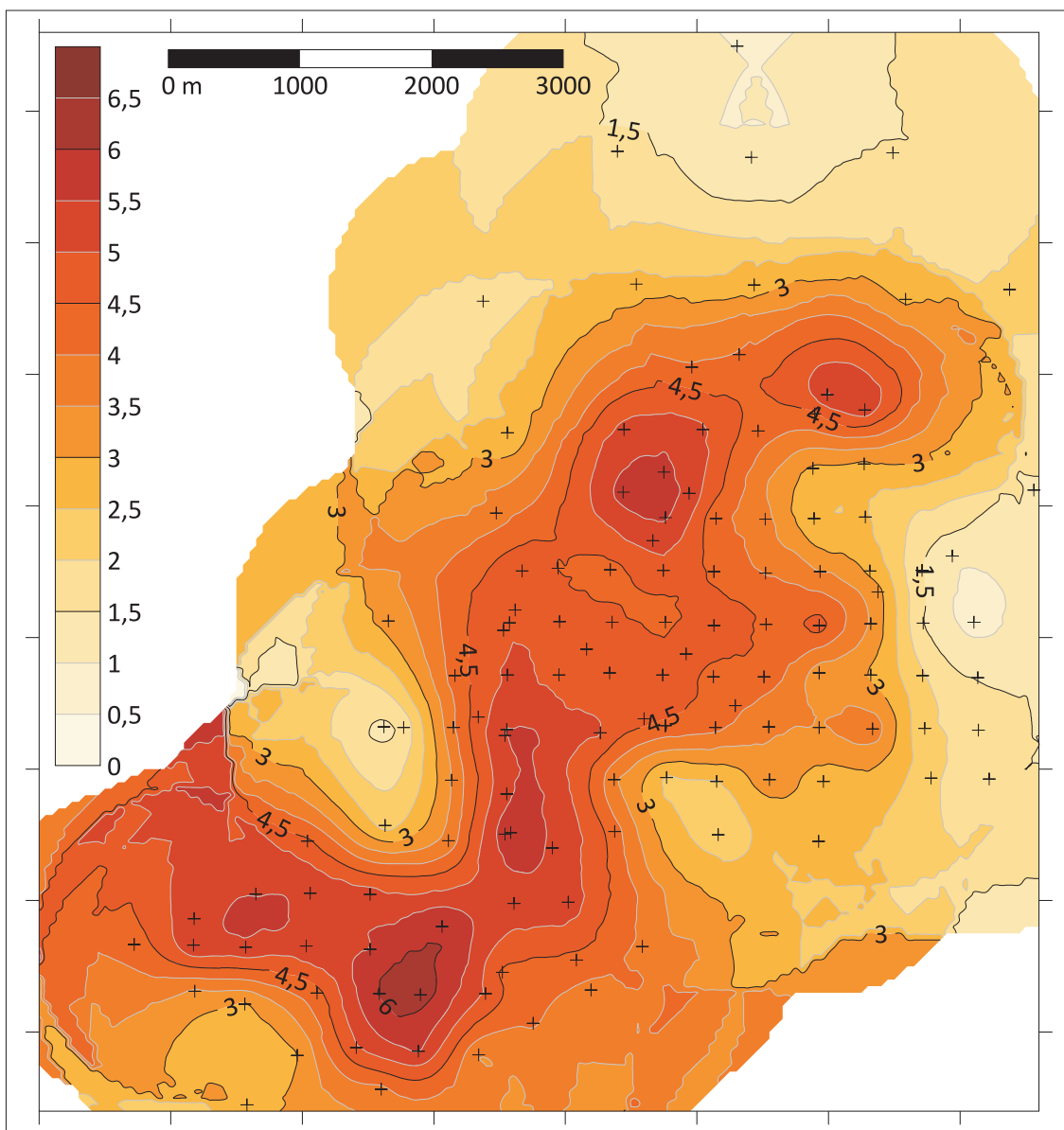
wartość szacowaną zostanie wyznaczona (niebezpośrednio) z wartości wariogramu $\gamma(\hat{z}, z_1)$, a ta jest funkcją odległości $h(\hat{z}, z_1)$ (ryc. 4). Ten sam schemat dotyczy wartości $\gamma(z_2, z_3)$. W procedurze krigingu będzie ona wyznaczona z modelu wariogramu na podstawie odległości $h(z_2, z_3)$.

Uważnego obserwatora może niepokoić fakt, że gdyby dosłownie brać to, co napisałem o obliczaniu wag, to można by dojść do błędnego wniosku, że im dalsze są obserwacje, tym większe wagi, czyli większy wpływ obserwacji na wartość szacowaną. Jest to sprzeczne z rozsądkiem – przecież wartości wariogramu rosną wraz z odległością, bo rośnie statystyczna różnica między obserwacjami. Aby można było poprawnie obliczyć wagi krigingu, należałoby raczej zastosować jakąś odwrotną zależność, czyli funkcję, której wartości maleją wraz z rosnącą odległością. Nic prostszego. Wymaga to tylko operacji matematycznej, której efektem będzie odwrócenie wariogramu względem linii poziomej wyznaczonej przez połowę wariancji. W wyniku tej operacji uzyskuje się wykres funkcji (ryc. 5), która jest niczym innym jak tylko lustrzanym odbiciem wariogramu.

To właśnie wartości odczytane z tej funkcji są wykorzystywane do obliczania wektora wag w równaniu krigingu. Niech nam takie wyjaśnienie wystarczy.

WYNIKI I ICH INTERPRETACJA

Możemy wreszcie uruchomić procedurę krigingu. Równania macierzowe dla tysięcy węzłów siatki interpolacyjnej zostaną obliczone w ułamku sekundy. W moim przykładzie grid ma 165 rzędów i 153 kolumny, czyli 25 245 węzłów siatki interpolacyjnej. Program wykonał obliczenia dla 19 201 węzłów. Pozostałe węzły (6044) nie zostały uwzględnione w obliczeniach, bo w ich otoczeniu nie dokonano żadnych obserwacji. W pamięci komputera zapisał się grid składający się z 19 201 rzędów. W każdym rzędzie za pomocą współrzędnych x, y i oszacowanej wartości $z(x, y)$ jest zapisany jeden węzeł siatki interpolacyjnej. To jest oczekiwany wynik naszej pracy. Na ryc. 6 przedstawiłem model miąższości złoza węgla brunatnego zwizualizowany za pomocą mapy izolinowej (*contour map*). Mapę wy-

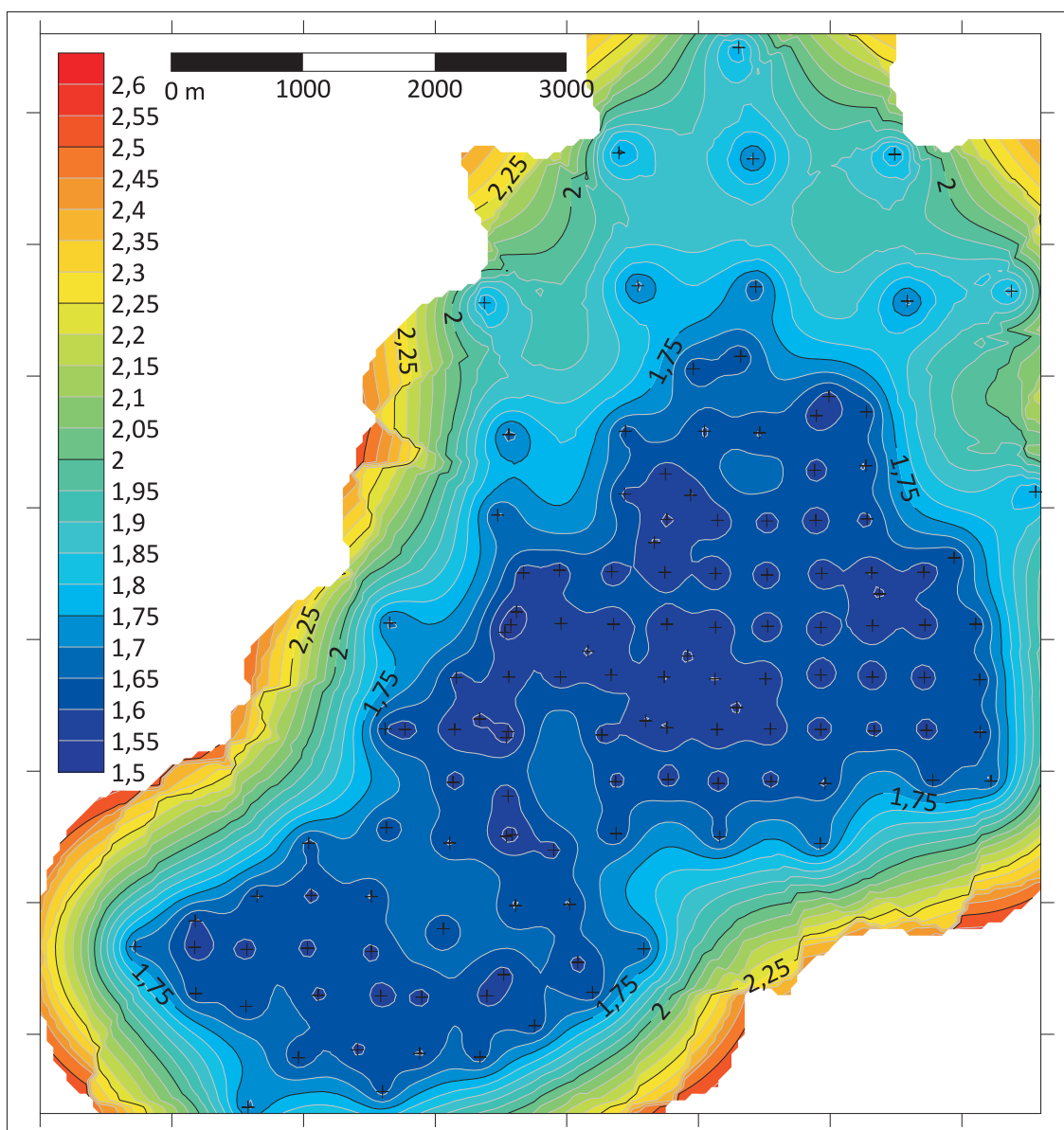


Ryc. 6. Model miąższości złoza węgla brunatnego [m]
Fig. 6. Model of the thickness of the lignite deposit [m]

pełniłem kolorami, korzystając z jednej z dostępnych w programie palet kolorów. Wstawiłem barwną skalę oraz podziałkę. To wyłącznie kosmetyka mająca na celu ułatwienie interpretacji modelu. Dla właściwej interpretacji modelu ważne są miejsca, w których dokonano obserwacji (tu: otwory). Oznaczyłem je krzyżykami. Skomentuję szerzej ten model. Na mapie (ryc. 6) są widoczne białe plamy oraz dziwne koliste kształty, szczególnie w północnej części mapy. Z doświadczenia wiemy, że złoża tak nie wygląda. Białe miejsca są tam, gdzie program, nie mając danych z obserwacji, nie oszacował żadnej wartości. Jednobarwne, koliste kształty to też efekt braku wystarczającej ilości danych. Promień koła jest równy promieniowi otoczenia (tu: 1200 m). Program zgodnie z algorytmem w całym kole oszacował wartość równą jednej obserwacji. Biorąc pod uwagę wzór [1] nie mógł zrobić nic innego. Zastanówmy się czy te koła i białe plamy to wada mapy albo procedury? Nie, wręcz przeciwnie, to zaleta. W tych miejscach program nie obliczył wartości, bo nie miał żadnych podstaw do takich obliczeń. Dzięki naniesionym

krzyżykom widać, gdzie występują braki obserwacji i gdzie model jest przez to niedoskonały. Ta niedoskonałość modelu jest dla nas ważną informacją, wskazującą na deficyty rozpoznania. Gdyby program w całym obszarze wygenerował ładny model odpowiadający naszym oczekiwaniom, to byłby to model niewiarygodny, czyli zły.

Spójrzmy teraz na drugą mapę (ryc. 7). Jest to mapa odchylenia standardowego modelu (*kriging standard deviation*) przedstawionego na ryc. 6. To ten bonus, o którym wspominałem na początku, który odróżnia kriging od niegeostatystycznych procedur interpolacyjnych. Można ją też nazwać mapą wiarygodności modelu. Uwaga, jest to błąd teoretyczny. Wartości przedstawione na mapie wynikają z obliczeń, które są oparte na obserwacjach. Rzeczywisty błąd, tak jak rzeczywista zmienność modelowanego parametru, jest nieznaną i pozostanie nieznaną aż do momentu eksploatacji złoża. Dla ułatwienia interpretacji zastosowałem tęczaową skalę barw. Tak jak na ryc. 6 tu również krzyżykami oznaczyłem otwory rozpoznawcze. Da się zauważyć zależność między gęstością opróbowania



Ryc. 7. Odchylenie standardowe krigingu miąższości złoża węgla brunatnego [m]

Fig. 7. Kriging standard deviation of the lignite deposit thickness model [m]

a wielkością błędu modelu. Tam, gdzie obserwacji jest więcej, błąd jest mniejszy i odwrotnie. Różnice widać szczególnie w środkowej i północnej części złoża. Największe wartości błędu obserwuje się na zewnątrz modelu. To oczywiste, tu nie ma otworów, a wartości zostały oszacowane w wyniku ekstrapolacji. Błąd szacowania na mapie przyjmuje wartości od 1,5 do 2,6 m. To odchylenie standardowe zatem błąd jest pokazany w jednostkach modelu (tu: w metrach).

Błąd o najmniejszej wartości (w zakresie 1,5–1,6 m) występuje w obszarach ciemnoniebieskich, czyli tam, gdzie sieć otworów jest najgęstsza. Pokład węgla ma w tym obszarze średnio 5 m grubości, tak wynika z modelu. Jeżeli odchylenie standardowe modelu ma w tym miejscu wartość ok. 1,5 m, to znaczy, że grubość pokładu węgla mieści się w granicach od 3,5 do 6,5 m. Czy jesteśmy tego pewni? Statystycznie tak, na poziomie niecałych 70%, bo przyjąłem, że wartość oszacowana mieści się w granicach $\pm 1 \sigma$, czyli w granicach \pm jednego odchylenia standardowego σ . Gdybyśmy chcieli mieć pewność stuprocentową, to należałoby na podstawie normalnego rozkładu błędu (a nie ma żadnych powodów, aby uznać, że błąd rozkłada się inaczej) przyjąć zakres $\pm 3 \sigma$, bo w rozkładzie normalnym w tych granicach mieszczą się prawie wszystkie wartości. No to proszę bardzo – w miejscu, gdzie model wskazuje $M = 5$ m, a odchylenie standardowe tego oszacowania wynosi $\sigma = 1,5$ m, możemy stwierdzić niemal ze stuprocentową dokładnością, że grubość pokładu mieści się w zakresie od 0,5 m do 9,5 m. Teraz jesteśmy wreszcie pewni, że nie popełniliśmy błędu! Owszem, ale jaka to jest właściwie informacja, że grubość pokładu wynosi od 0,5 do 9,5 m? To nie jest dobra informacja. To tak, jakby kupować mieszkanie o powierzchni średnio 60 m², ale w granicach od 10 do 150 m². Dlatego porzucmy stuprocentową pewność i niech nam starczy informacja, że w miejscu, gdzie model wskazuje $M = 5$ m z prawdopodobieństwem ok. 70% miąższość pokładu mieści się w granicach od 3,5 do 6,5 m, średnio ma 5 m. To już bardzo dużo informacji. Chciałbym, aby to mocno wybrzmiało. Żaden model nie pokazuje wartości pewnych. Są to wartości średnie i obciążone błędem wynikającym z niedoskonałości obserwacji i z samej metody modelowania.

W drugim artykule mojego cyklu szeroko rozpisałem się na temat wariancji lokalnej (*nugget effect*). Wróćmy do tego problemu. Najmniejsza wartość błędu na mapie (ryc. 7) wynosi $\sigma = 1,5$ m. Jest to błąd modelu w obszarach, gdzie sieć obserwacji jest bardzo gęsta. Model jest w tych miejscach najbardziej dokładny, a mimo to nie jest dokładny. Błąd rzędu 1,5 m dla pokładu o grubości od 3 do 6,5 m to duży błąd! Dlaczego jest taki wielki? Odpowiedź tkwi w wariogramie. Z modelu wariogramu odczytałem wielkość wariancji lokalnej na poziomie 1,8 m². To wariancja, więc występuje w jednostkach kwadratowych. Po spierwiastkowaniu otrzymamy odchylenie standardowe, którego wartość wynosi $\sigma = 1,34$ m. Tyle wyniósłby teoretyczny błąd modelu, gdyby sieć obserwacji była bardzo gęsta, znacznie bardziej gęsta niż ta, którą dysponujemy. Tymczasem obserwacje użyte do modelowania są oddalone od siebie średnio o 400 m. Z wariogramu (ryc. 3) można odczytać, że tylko 28 par obserwacji reprezentuje odległość mniejszą od 300 m. Dlatego najmniejszy błąd modelu ma wartość większą niż wariancja lokalna, którą odczytaliśmy z wykresu wariogramu. Ergo – najmniejszy teoretyczny błąd modelu

będzie zawsze większy niż odchylenie standardowe obliczone z wariancji lokalnej. A zatem, informację o najmniejszym błędzie modelu mamy już na etapie obliczania wariogramu! Już w tym miejscu możemy zdecydować, czy model o takim błędzie jest dla nas interesujący, czy lepiej odstąpić od wykonywania mapy. Nieco szerzej pisałem o tym w drugim artykule cyklu (Naworyta, 2024b).

A właściwie dlaczego ten błąd modelu jest tak duży? Jest on prostą konsekwencją błędów, jakie popełniono podczas wykonywania obserwacji (oznaczania parametrów na podstawie rdzeni z otworów rozpoznawczych). Jeżeli obserwacje, na bazie których wykonujemy model, są obciążone błędem, to oczywiście jest, że ten błąd będzie miał odbicie w modelu. Co więcej, w modelu nie będzie on nigdy mniejszy niż wariancja lokalna, będzie zawsze tylko większy. Innymi słowy, nie możemy się spodziewać, że posługując się danymi, które zawierają w sobie błąd, uda nam się wykonać model, który będzie bezbłędny, albo którego błąd będzie mniejszy niż błąd obserwacji. To bardzo ważny wniosek.

Na zakończenie powtórzę czynności, bez których nie wyobrażam sobie poprawnego modelowania metodą krigingu. Są to:

- analiza wstępna danych źródłowych – podstawowe miary statystyczne;
- analiza rozmieszczenia obserwacji w przestrzeni (np. otworów);
- obliczenie wariogramu empirycznego i stwierdzenie istnienia autokorelacji;
- dopasowanie modelu wariogramu do wariogramu empirycznego oraz wyznaczenie wariancji lokalnej i zasięgu autokorelacji (*AR*);
- zdefiniowanie geometrii siatki interpolacyjnej – zasięgu i wielkości oczka siatki;
- w procedurze krigingu wyznaczenie otoczenia na podstawie zasięgu autokorelacji (*AR*);
- zdefiniowanie w programie plików wynikowych gridu modelu i gridu odchylenia standardowego modelu;
- wizualizacja gridów wynikowych w formie mapy modelu i błędu modelu;
- wnioskowanie o modelowanym zjawisku na podstawie modelu, błędu modelu i wariogramu.

W tabeli 1 zestawiałem podstawowe miary statystyczne modelu w porównaniu do odpowiednich miar danych źródłowych. Warto zwrócić uwagę na efekt wygładzania zaznaczający się w wartościach skrajnych, wariancji i odchyleniu standardowym po interpolacji. Takie zestawienie pozwala oszacować, z jaką dokładnością model odzwierciedla zmienność danych źródłowych. W tabeli 1 wyraźnie różnią się miary rozrzutu – wariancja i odchylenie standardowe. Duże różnice mogą wynikać z niedoskonałości modelu w obszarach o niewielkiej liczbie danych oraz w miejscach, gdzie model został wykonany metodą ekstrapolacji, czyli na zewnątrz, poza obszarem otworów rozpoznawczych.

PODSUMOWANIE

W trzydziściowym cyklu artykułów, który w zamyśle miał być przewodnikiem praktycznym, nakreśliłem tylko najważniejsze problemy stosowania metod geostatystycznych. Celowo i świadomie pominąłem pewne zagadnienia, od których naukę geostatystyki zwyczajowo się zaczyna. Nie wspominałem o założeniu stacjonarności, o występo-

Tab. 1. Porównanie wybranych miar statystycznych danych źródłowych i modelu grubości pokładu węgla brunatnego
Table 1. Comparison of selected statistical measures of source data and the lignite seam thickness model

Miara Measure	Dane źródłowe Source Data	Model Model
Liczba otworów/węzłów siatki / Number of holes/mesh nodes	131	19 201
Wartość minimalna [m] / Minimum value [m]	0,2	0,2
Wartość maksymalna [m] / Maximum value [m]	8,5	6,24
Średnia [m] / Average [m]	3,89	3,199
Mediana [m] / Median [m]	3,9	3,07
Wariancja [m ²] / Variance [m ²]	4,31	1,68
Odchylenie standardowe [m] / Standard deviation [m]	2,08	1,297
Zakres [m] / Range [m]	8,3	6,04

waniu i wpływie trendu w danych źródłowych, występowaniu i znaczeniu wartości odstających (*outliers*) w zbiorze danych. Nie wspominałem ani słowem o zmienności kierunkowej oraz o innych formach krigingu, a przecież jest jeszcze kriging uniwersalny, cokriging, kriging inductorowy i inne. Cały osobny artykuł można poświęcić samej tylko symulacji geostatystycznej.

Obawiam się, że pominięcie założeń stacjonarności będzie mi poczytane przez konserwatystów za błąd kardynalny. Mam tego świadomość i biorę to ryzyko na siebie. Celowo to ominąłem, podkreślając tylko praktyczne zagadnienia, z jakimi amator modelowania musi się zaznajomić, aby mógł świadomie sterować programem i nie musiał korzystać z opcji *skip to end*, którą autorzy oprogramowania przygotowali jako koło ratunkowe dla laików. Samodzielne poznanie podstaw geostatystyki z podręczników jest dość trudne i czasochłonne. Nie da się tego również nauczyć, korzystając z instrukcji obsługi programów geostatystycznych. Sam kilkakrotnie uczestniczyłem w dedykowanych szkoleniach w firmie *Geovariances* we Francji i jestem świadomy, że niektóre wysublimowane opcje pakietu *Isatis* poznałem zaledwie powierzchownie. Stoję

twierdząc na stanowisku, że nauka nie jest po to, aby utrudniać ludziom życie, a rolą naukowca nie jest udowadnianie innym, że nauka i wiedza to nie jest coś dla *rozmaitych Puchatków i Prosiaczków*, jak to w monologu Kłapouchego wspaniale wyraził Milne (1926). Wręcz przeciwnie – narzędzia wypracowane przez naukowców powinny się nadawać do stosowania również przez laików. Mam nadzieję, że mój cykl artykułów przyda się komuś i pozwoli zrozumieć pewne, nie do końca intuicyjne, meandry geostatystyki. Marzy mi się, aby Czytelnik, podobnie jak ja, gdy to wszystko wreszcie zrozumiałem, mógł zakrzyknąć: *Ależ to proste!*

LITERATURA

- MILNE A.A. 1926 – Kubuś Puchatek; w tłumaczeniu Ireny Tuwim. Nasza Księgarnia.
 NAWORYTA W. 2017 – Meandry modelowania złóż – na podstawie doświadczeń i obserwacji. *Gór. Odkr.*, 4/2017: 4–9.
 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.

Praca wpłynęła do redakcji 3.09.2024 r.
 Akceptowano do druku 12.09.2024 r.