ARTYKUŁY INFORMACYJNE

Geostatystyka dla niematematyków. Przewodnik praktyczny – część II

Wojciech Naworyta[1]

Geostatistics for non-mathematicians. Practical guide – Vol. II. Prz. Geol., 72: 391–398; doi: 10.7306/2024.21

A b s t r a c t. 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 this, the second article of a three-part series, I will focus on the variogram – the most important tool in geostatistics. I will show what is hidden behind the mathematical formula and what information about the analysed spatial phenomenon can be obtained only using a variogram. In the third article, I will present another useful function of the variogram. It will act as a tool in the process of geostatistical interpolation using the kriging method.

Keywords: *geostatistics, variogram, local variance, nugget effect, variogram model*

Narzędzia geostatystyczne są przydatne, a nawet niezbędne, w wielu dziedzinach, nie tylko w geologii. Na przeszkodzie do ich powszechnego stosowania stoją pozornie trudne podstawy matematyczne, od których zwykło się zaczynać nauczanie przedmiotu. Wieloletnie doświadczenia i obserwacje upoważniają mnie do twierdzenia, że nawet jeżeli metody geostatystyczne są stosowane, to często w sposób niepełny i niedoskonały. W cyklu trzech artykułów chciałbym przybliżyć potencjalnym użytkownikom niematematykom najważniejsze przydatne metody i narzędzia z arsenału statystyki przestrzennej oraz sposoby ich właściwego wykorzystania. Wskażę dziedziny, w których mogą mieć zastosowanie. Wyjaśnię, czy zawsze można je stosować. Pokażę, jakie decyzje należy podjąć w toku obliczeń i jak należy interpretować otrzymane wyniki. Nie będzie to kompendium, raczej kieszonkowy przewodnik ułatwiający pierwszy kontakt z geostatystyką. W niniejszym artykule, drugim w trzyczęściowym cyklu, skupię się na wariogramie – najważniejszym narzędziu geostatystyki. Pokażę, co kryje się we wzorze Matherona oraz jakie informacje o analizowanym zjawisku przestrzennym można pozyskać wyłącznie za pomocą wariogramu. W trzecim artykule przedstawię kolejną użyteczną funkcję wariogramu. Będzie pełnił funkcję narzędzia w procesie interpolacji geostatystycznej metodą krigingu.

WARIOGRAM

Wariogram to narzędzie wielorakiego użytku, coś jak kombinerki w warsztacie geostatystyki. Przydatność tego narzędzia będę udowadniał stopniowo. *Wariogram, variogram, semiwariogram* – to synonimy. W literaturze polskiej i anglojęzycznej, zwłaszcza tej wcześniejszej, stosuje się słowo *semiwariogram* względnie *semivariogram*, w programach komputerowych, tych współczesnych, jest już stosowany zwrot *wariogram* albo *variogram*. Ja też wolę tę ostatnią, uproszczoną wersję i taką będę tu stosował. *Semiwariogram* znaczy właściwie pół wariogramu, tak jak *semidolce* i *semisec* – w odniesieniu do wina to odpowiednio półsłodkie i półwytrawne. Znaczenie tego *pół* wyjaśnię, opisując wzór, jeden z niewielu, który w moim cyklu artykułów o geostatystyce musi się pojawić. Słowo *wariogram*, tak jak i słowo *wariograf,* odnosi się do różnicy. Przedrostek *wario* od łacińskiego *varius* znaczy różny, różnorodny, rozmaity. Zatem chodzi o różnienie się, o różnicę. W przypadku wariografu chodzi o wykazanie, czy to, co mówi osoba badana wariografem, nie różni się od prawdy. W przypadku wariogramu, też chodzi o różnice. Drogi Czytelniku, jeżeli znudziłem Cię słowem wariogram, to lepiej już teraz porzuć ten artykuł, bo dalej nie będzie wcale lepiej. Bez wariogramu nie ma geostatystyki!

Cykl moich artykułów zatytułowałem *Geostatystyka dla niematematyków.* Chcąc być konsekwentnym muszę wzór Matherona bardzo dokładnie objaśnić (Matheron, 1963):

$$\gamma(h) = \frac{1}{2n} \sum_{i=1}^{n} (z_{i+h} - z_i)^2 \qquad [1]$$

Jest on kluczowy i bez zrozumienia jego sensu nie ma co brnąć dalej. Wartości wariogramu liczymy na podstawie zbioru obserwacji *Z*. Zacznijmy więc od tego, że mamy obserwacje, czyli dane, które zostały pozyskane w przestrzeni jakiegoś zjawiska, np. złoża kopaliny. Te dane są ściśle określone w przestrzeni za pomocą współrzędnych

[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

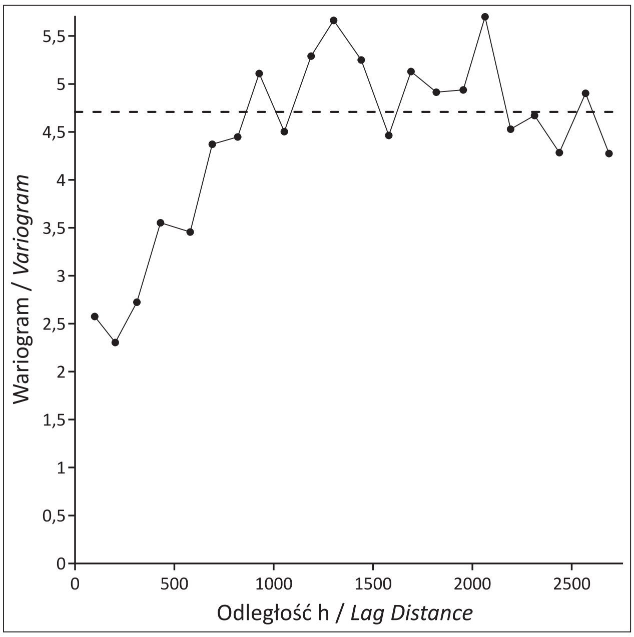
płaskich X i Y albo przestrzennych X, Y, Z. Wystarczy w zupełności, jak na tym etapie ograniczymy się do współrzędnych płaskich określających położenie pobrania próby na powierzchni geodezyjnej albo, jak kto woli, na mapie. Obserwacje są rozrzucone w mniej lub bardziej regularnej sieci. Są obserwacje względem siebie bliskie i są odległe. Można ze zbioru obserwacji wybrać podzbiór par, które są oddalone od siebie o h = 100 m, drugi podzbiór obserwacji prowadzonych w większej odległości, np. h = 200 m itd. Takie wyróżnianie podzbiorów zakończymy na niewielkiej liczbie par obserwacji, które są odległe od siebie o największą możliwą odległość wynikającą z rozmiarów analizowanego zjawiska. Te podzbiory nie wykluczają się wzajemnie. Każda obserwacja będzie należeć właściwie do każdego podzbioru, bo można z nią tworzyć dowolne pary bliskie i odległe. W tym wyróżnianiu podzbiorów chodzi o wszystkie możliwe kombinacje par. Obserwacje obok współrzędnych mają też jakąś cechę oznaczoną liczbowo, jakiś parametr. Załóżmy, dla ułatwienia, że sprawa dotyczy rozpoznania złoża pokładowego otworami wiertniczymi, a parametr, który badamy, to grubość pokładu zmierzona na wyciągniętym z górotworu rdzeniu.

Wróćmy do wzoru i zacznijmy od jego lewej strony. Wartość wariogramu $\gamma(h)$ liczymy każdorazowo dla konkretnej odległości h pomiędzy parami obserwacji, która wynosi np. h = 100 m, czyli dla par, które są odległe od siebie o pewien dystans – tu: ok. 100. Dlaczego około? Przecież matematyka jest dziedziną ścisłą, skąd zatem to około? Po pierwsze, to nie matematyka tylko geostatystyka, a więc bardziej statystyka niż matematyka. Po drugie, w przyjętej odległości h musi być jakaś tolerancja, ponieważ w zbiorze obserwacji trudno byłoby odszukać otwory odległe od siebie o odcinek równy dokładnie h = 100 m. W programach komputerowych ta tolerancja jest płynna i można ją swobodnie definiować w parametrach wariogramu. Nie zaprzątajmy sobie teraz tym głowy. Zatem ze zbioru obserwacji wyciągamy wszystkie pary, które są odległe o ok. 100 m (np. między 80 i 120 m, albo nawet między 50 a 150 m). Każdą taką parę bierzemy osobno pod lupę i obliczamy różnicę pomiędzy wartościami pomierzonego parametru, czyli w tym przypadku różnice w miąższości złoża. Zadajemy pytanie – jak różnią się wartości pomierzone blisko siebie, czyli tu w odległości ok. 100 m. Interesuje nas to, o ile się różnią – czyli już wiemy dlaczego to narzędzie nazywa się wariogramem. Wyrażenie $z_{(i)} - z_{(i+h)}$ to zwykła różnica wartości parametru między tym zmierzonym w miejscu (i) oraz tym zmierzonym w miejscu odległym od niego o odległość h czyli w miejscu ($i + h$). Ze wzoru wynika, że każdą taką różnicę należy podnieść do kwadratu. Po co? Jednym z powodów jest to, aby różnica nie zależała od kolejności odejmowania. Idźmy dalej – przed nawiasem mamy sumę podzieloną przez $2n$. Czyli sumę wszystkich kwadratów różnic podzieloną przez podwojoną liczbę tych par. Zostawmy na razie tę dwójkę w mianowniku. Suma podzielona przez liczbę obserwacji n to nic innego jak przepis na średnią arytmetyczną (por. wzór 2). Zatem wzór na wariogram pozwala nam obliczyć, jaka jest średnia kwadratowa różnica pomiędzy obserwacjami. Jeżeli na chwilę pominiemy nawet to potęgowanie, to w gruncie rzeczy chodzi o to, jak średnio różnią się obserwacje oddalone od siebie o odległość h. Ten kwadrat nie powinien nas deprymować. Będziemy o nim pamiętać. Czyli już teraz wiemy, że chodzi o wielkość średnią, czyli statystyczną i chodzi o statystyczną różnicę pomiędzy obserwacjami prowadzonymi w danej odległości h. A ta dwójka w mianowniku? To właśnie ona jest winna zamieszania w nazewnictwie. To przez nią niektórzy mówią semiwariogram, czyli pół wariogramu. Ona ma swój sens, ale tu i teraz trudno mi będzie go wyjaśnić. Tymczasem zatrzymajmy się jeszcze chwilę przy samej nazwie. Wspomniałem już, że semiwariogram czy wariogram to jedno i to samo. Gdyby upierać się przy tej pierwszej nazwie, która podkreśla połowę, czyli tę nieszczęsną dwójkę w mianowniku, to znaczyłoby, że wariogramem należy określić wzór bez tej dwójki. Taki wzór w geostatystyce nie funkcjonuje, nie jest nigdzie stosowany. Zatem spokojnie możemy używać określenia wariogram zamiast konserwatywnego semiwariogram, odnosząc się do tego, co zostało wyrażone wzorem [1].

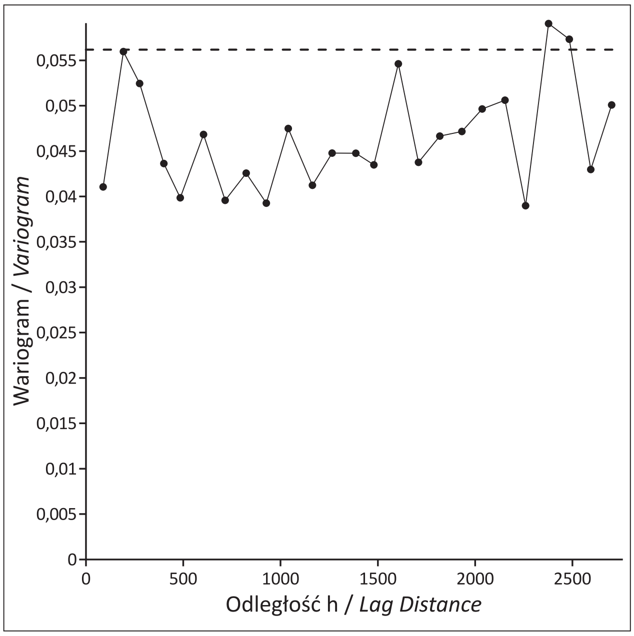
Na razie obliczyliśmy wartość wariogramu $\gamma(h)$ dla h = 100 m. Ze wzoru otrzymaliśmy jedną wartość liczbową. To teraz tę samą czynność powtarzamy dla innych odległości między parami obserwacji, np. h = 200, h = 300 itd. aż do największej możliwej wartości, która wynika z rozciągłości przestrzennej zjawiska. Wartości nanosimy na wykres, gdzie w osi poziomej są odległości między parami h (*Lag Distance*), a na osi pionowej występują wartości wariogramu $\gamma(h)$. To te kropki na wykresie (ryc. 1). Przypominam, że obliczaliśmy wariogram miąższości złoża. Wielkość ta jest wyrażona w metrach, tymczasem wariogram jest wyrażony w jednostkach do kwadratu, czyli tu w tym konkretnym przypadku w metrach kwadratowych. To oczywiste, bo we wzorze wszystkie policzone różnice w parach najpierw podnosimy do kwadratu, a potem je sumujemy i dzielimy tę sumę przez podwójną liczbę par $2n$. Kwadrat nie znika. Ten kwadrat nie ma znaczenia logicznego, tak jak w przypadku np. obliczania powierzchni – ma znaczenie wyłącznie matematyczne. A zatem pamiętajmy, że wariogram wyraża nam statystyczną różnicę między parami obserwacji, ale jest to różnica kwadratowa!

Spróbujmy zinterpretować wykres przykładowego wariogramu miąższości złoża węgla brunatnego (ryc. 1). Dla h = 200 wartość wariogramu, czyli to, jak statystycznie różnią się obserwacje bardzo bliskie, wyniosła ok. 2,2 m^2. Po spierwiastkowaniu to ok. 1,48 m. Jest to średnia różnica miąższości złoża dla obserwacji odległych o ok. 200 m. To znaczy, że najczęściej tak się różnią wyniki obserwacji, ale nie zapominajmy o tym, że jest to wartość średnia, czyli statystyczna. W rzeczywistości różnice mogą osiągać nawet 7 m, ale mogą być także bliskie zera. Jednak statystycznie rzecz biorąc obserwacje w parach odległych od siebie o ok. 200 m różnią się o ok. 1,48 m.

Dla h = 500 m wariogram przyjmuje wartość znacznie większą. Pary są bardziej odległe i też wartości zaobserwowane w parach bardziej się od siebie różnią. Wartość wariogramu wynosi ok. 3,6 m^2, czyli różnią się o 1,9 m. Obserwacje wraz z rosnącą odległością różnią się coraz bardziej. Rośnie statystyczna różnica. Ten wzrost obserwujemy aż do odległości bliskiej h = 1000 m. Po osiągnięciu tego dystansu wzrost wykresu wyhamowuje i jego kształt zaczyna się robić mniej przewidywalny. Wraz ze wzrostem odległości obserwacje w parach różnią się od siebie coraz bardziej, aż do odległości 1000 m. Do tej odległości obserwuje się statystyczną regularność, czyli wyraźną zależność między odległością i statystyczną różnicą między obserwacjami. Ta zależność to autokorelacja, o której szeroko



Ryc. 1. Przykładowy wariogram miąższości złoża węgla brunatnego
Fig. 1. Example of a variogram of the thickness of a lignite deposit



Ryc. 2. Wariogram zawartości siarki całkowitej w złożu węgla brunatnego
Fig. 2. Variogram of total sulphur content in a lignite deposit

wypowiedziałem się w pierwszym artykule tego cyklu (*Prz. Geol*., 72 (7): 341–349), a wariogram jest narzędziem, które umożliwia stwierdzenie jej istnienia. To pierwsza przydatna funkcja wariogramu. Z jego pomocą możemy dowiedzieć się, czy analizowane zjawisko przestrzenne (tu: miąższość złoża) wykazuje autokorelację. Jeżeli tak, to możemy pokusić się o wykonanie interpolacji oraz mapy zmienności tego zjawiska.

Dla odległości większych, czyli $h > 1000$ m, zachowanie wariogramu staje się chaotyczne. Ta odległość wyznacza bardzo istotną granicę. Jest to zasięg autokorelacji (*autocorrelation range*). Badany parametr Z wykazuje autokorelację o zasięgu $h = 1000$ metrów (ryc. 1). Możliwość oceny zasięgu autokorelacji to druga ważna funkcja wariogramu.

Dla kontrastu na przykładzie wariogramu zawartości siarki w złożu węgla brunatnego (ryc. 2) pokażę, kiedy wariogram wskazuje brak istnienia autokorelacji. Wartości tego wariogramu oscylują wokół pewnej wartości poziomej. Nie ma wyraźnego trendu ani rosnącego ani malejącego, a zmiany są losowe. To znaczy, że różnica pomiędzy obserwacjami nie zależy od odległości h. Niezależnie od tego, jakie weźmiemy pary obserwacji, czy te bliższe czy dalsze, wartości na tych parach statystycznie różnią się podobnie. Nie ma autokorelacji. W takim przypadku jesteśmy zmuszeni odstąpić od dalszych analiz z użyciem narzędzi geostatystycznych, te już nam nie pomogą. Nie narysujemy wiarygodnej mapy zmienności przestrzennej tego zjawiska. Możemy dalej analizować zbiór obserwacji, ale z użyciem narzędzi statystycznych – policzyć średnią, wariancję, odchylenie standardowe, wykreślić histogram. I to właśnie jest doskonały moment, aby przypomnieć znaczenie niektórych wielkości, bo przynajmniej jedna z nich bardzo nam się przyda. Nie są to nowe wzory w tym artykule. Zaraz się okaże, że przynajmniej w części pojawiły się one już wcześniej i nie bez powodu.

$$m = \frac{1}{n} \sum_{i=1}^{n} (z_i) \qquad [2]$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (z_i - m)^2 \qquad [3]$$

Pierwszy z nich to oczywiście wzór na średnią arytmetyczną [2]. W statystyce jest to tzw. wartość oczekiwana. Niby to samo, a jednak nie to samo. Jeżeli na podstawie miąższości złoża ze 100 otworów obliczymy średnią arytmetyczną i wyjdzie nam 10 m to w odniesieniu do tych 100 otworów będzie to średnia arytmetyczna, ale jeżeli na tej podstawie wnioskujemy o średniej grubości całego pokładu badanego złoża, to musimy to nazwać wartością oczekiwaną. Gdybyśmy odwiercili kolejnych 200 otworów, to wartość liczbowa obliczona wzorem [2] z pewnością będzie inna. Jak widać, częściej będziemy mówić o wartości oczekiwanej niż o średniej arytmetycznej. Wzór na średnią jest banalny, wstawiając go tutaj nie miałem zamiaru obrażać inteligencji Czytelnika. Musiał się tu znaleźć, bo wartość m posłuży nam do obliczenia wariancji [3], a ta jest równie ważna jak wariogram.

Wariancja to wielkość statystyczna, która jest średnią sumy różnic podniesionych do kwadratu. W nawiasie jest różnica między obserwacją z_i i średnią z obserwacji m. Różnice po podniesieniu do kwadratu sumujemy, wykorzystując wszystkie dostępne obserwacje, a sumę dzielimy przez liczbę obserwacji n – no niezupełnie... przez n, ale pomniejszone o 1. Wytłumaczenie sensu tej jedynki w estymatorze statystycznym, jakim jest wariancja, wykracza poza ramy tego artykułu. Właściwie ma ona sens liczbowy wyłącznie wtedy, gdy dysponujemy niewielką liczbą obserwacji, powiedzmy $n = 10$. Wtedy przez uwzględnienie tej jedynki w mianowniku sumę dzielimy przez 9, a nie przez 10. Niezależnie od tego, co dzielimy, różnica jest wyraźna. W sytuacji, gdy mamy 1000 obserwacji, to czy sumę różnic dzielimy przez 1000 czy przez 999, nie ma większego wpływu na wartość obliczonej w ten sposób wariancji. Wzór na wariancję [3] bardzo przypomina wzór na wariogram [1]. W jednym i drugim jest średnia z różnic podniesionych do kwadratu. Tyle, że we wzorze na wariancję [3] w nawiasie jest różnica między obserwacją z_i i wartością średnią z obserwacji *m*, a w wariogramie liczymy różnice pomiędzy obserwacjami w parach. Zaraz porównamy te wielkości bliżej.

Jedno jest pewne – wariancja i wariogram muszą mieć ze sobą coś wspólnego oprócz nazwy. Mają i będziemy to wykorzystywać. Wariancja jest statystyczną miarą rozrzutu obserwacji wokół średniej *m*. Im bardziej wartości obserwacji odbiegają od wartości średniej, tym większa jest wartość wariancji. I odwrotnie jeżeli wartości są bardzo skupione wokół wartości średniej *m,* to i wariancja przyjmuje niewielką wartość.

Przyjrzyjmy się jeszcze raz wzorom na wariancję i wariogram. Wariancja operuje na liczbach niezależnie od ich umiejscowienia w przestrzeni. Nie jest istotne, która wartość gdzie została zmierzona. Wariancja to wielkość statystyczna. W wariogramie też liczymy różnice, ale tu istotne jest wzajemne położenie obserwacji. Obserwacje mają współrzędne. Wariogram to wielkość geostatystyczna. W przypadku wariancji mamy do czynienia z jedną wartością, podczas gdy wariogram to wykres złożony w wielu wartości obliczonych dla różnych odległości *h* między parami obserwacji. Jedna i druga wartość występuje w jednostkach kwadratowych. Obliczmy więc wariogram na danym zbiorze obserwacji oraz wariancję na tym samym zbiorze i zestawmy wyniki na jednym wykresie. Właściwie już to zrobiłem. Pozioma linia przerywana na ryc. 1 i 2 to właśnie nic innego jak wartość wariancji. Obowiązuje dla każdego *h*, bo wariancja jest liczona dla wszystkich obserwacji, więc od *h* w żaden sposób nie zależy. Wartości wariogramu rosną do pewnej odległości *h.* Osiągnąwszy poziom wyznaczony linią wariancji wariogram przestaje rosnąć i zaczyna się jego nieregularny przebieg. Często, choć nie zawsze, zdarza się, że wariogram osiąga poziom wariancji dla odległości *h* równej zasięgowi autokorelacji $h = AR$ (*AR* – *autocorelation range*). Nie jest to reguła, ale najczęściej tak właśnie się dzieje. Dlatego, rysując wykres wariogramu, należałoby zestawić go z wariancją.

Jak w praktyce interpretować to, co widać na wykresach? Jeżeli wariogram osiąga niskie wartości, co najczęściej ma miejsce dla par obserwacji bliskich sobie i wraz z rosnącą odległością między parami obserwacji wartości wariogramu rosną, to obserwujemy zależność zwaną autokorelacją. Wartości wariogramu rosną do pewnej odległości $h = AR$. Dla odległości $h > AR$ mówimy, że badane zjawisko jest losowe. Rzeczywiście, nie widać żadnej zasady, która by tym przebiegiem rządziła. W opozycji do tego, gdy $h < AR$, zmienność zjawiska po części jest nielosowa (przewidywalna), a po części losowa (nieprzewidywalna). Im wyższą wartość osiąga wariogram, im bardziej zbliża się do poziomu wariancji, tym większy jest udział losowy w zmienności badanego zjawiska. Jeżeli wariogram przyjmuje wartości bliskie poziomu wariancji, to znaczy, że dla tych odległości *h* między parami udział losowy w zmienności jest dominujący albo wręcz całkowity.

Nadeszła pora, aby wrócić do nieszczęsnej dwójki w mianowniku wzoru na wariogram, która odpowiada za przedrostek *semi* w starym, konserwatywnym nazewnictwie. Gdyby nie było tej dwójki, to po zestawieniu wariogramu z wariancją poziom tej ostatniej wypadłby w połowie wartości wykresu wariogramu. Takie zestawienie wykresów byłoby mało praktyczne. Z dwójką w mianowniku wariogram i wariancja obliczona na tej samej populacji obserwacji świetnie obrazuje nielosowy i losowy charakter zmienności zjawiska. Nie jest mi znany inny powód, dla którego ta dwójka została wprowadzona do przepisu na wariogram [1].

Wróćmy jeszcze raz do zjawiska, które nie wykazuje autokorelacji. W jego zmienności obserwuje się wyłącznie udział losowy. Wariogram dla takiego przypadku pokazałem już na ryc. 2. Linią przerywaną oznaczyłem poziom wariancji. Niezależnie od odległości *h* pomiędzy parami obserwacji statystyczna różnica kwadratowa pomiędzy tymi obserwacjami jest bliska wartości wariancji, jest losowa.

NUGGET EFFECT, EFEKT SAMORODKÓW, WARIANCJA LOKALNA

Pozostajemy przy wariogramie. Zwrócę uwagę na kolejną niezwykle ważną jego cechę. Pisząc o tym, jak się oblicza wartość wariogramu, zacząłem rachunki od przykładowej wartości $h = 100$. A właściwie dlaczego $h = 100$, a nie od początku, czyli dla $h = 0$? No właśnie, dlaczego nie od zera? Przypominam, że wartość wariogramu liczy się dla par obserwacji. To znaczy, że dla $h = 0$ te pary składałyby się z tych samych obserwacji, bo odległość pomiędzy obserwacjami wynosiłaby $h = 0$. Gdybyśmy rzeczywiście tak to liczyli to wartość wariogramu $\gamma(h = 0)$ teoretycznie zawsze wynosiłaby 0. A gdyby wziąć obserwacje bardzo sobie bliskie, ale nie pokrywające się – dla *h* bliskiego 0? Czy w ogóle dysponujemy takimi obserwacjami? W obszarze geologii raczej nie. Odwiercenie otworu to spory koszt, dlatego geologiczne prace poszukiwawcze zaczyna się od wierceń otworów oddalonych od siebie nieraz o wiele kilometrów, potem, gdy wyniki tych prac rokują nadzieje na występowanie złoża, zagęszcza się sieć rozpoznawczą i tak po kilku etapach rozpoznania geologicznego dochodzi się do siatki otworów, w której zależnie od rodzaju złoża i kopaliny, średnia odległość między otworami może wynieść od 100 do 300 m. Nie ma powodu, aby wiercić otwory blisko siebie. Nie robi się tego z racjonalnych powodów ekonomicznych. Czasem, wyjątkowo rzadko, zdarza się, że znajdą się otwory sobie bliższe, będą to jednak pojedyncze przypadki, prawdopodobnie pochodzące z różnych kampanii dokumentowania złoża. To samo wytłumaczenie można zastosować do innych dziedzin, np. badania zanieczyszczenia gleby. Jeżeli badamy konkretną powierzchnię, to z powodów ekonomicznych nie pobieramy próbek odległych od siebie o $h = 1$ m, tylko w siatce otworów regularnie rozmieszczonych w całym obszarze badań, który może obejmować powierzchnię wielu dziesiątków hektarów.

Wariogramu $\gamma(h = 0)$ nie oblicza się, bo nie ma to logicznego sensu, a jednak, paradoksalnie, wartość wariogramu w tym miejscu ma kolosalne znaczenie dla wszystkich analiz geostatystycznych. Trzeba ją oszacować i robi się to przez aproksymację, przedłużając krzywą wariogramu w kierunku osi pionowej. W znakomitej większości przypadków linia taka przetnie oś pionową w miejscu, gdzie wariogram przyjmuje wartości znacznie większe od zera. Zatem wbrew temu, czego można byłoby oczekiwać, wartość $\gamma(h = 0)$ najczęściej wcale nie wynosi zero.

Wartość wariogramu $\gamma(h = 0) > 0$ określa się mianem efektu samorodków (*nugget effect*). Co ona oznacza? Dlaczego ta wartość jest tak ważna? Skąd wzięła się nazwa? Zacznę od końca, czyli od nazwy. Swoje pierwsze zastosowania geostatystyka znalazła w rozpoznaniu złóż złota w Republice Południowej Afryki. Złoto występuje w złożach w postaci rudy o określonej, niewielkiej zawartości pierwiastka Au. Czasem jednak znajdzie się skupienie złota w postaci bryłki, określanej mianem samorodka (*nugget*). Gdy więc wiercimy otwory w złożu, to w rdzeniach najczęściej będziemy mieli do czynienia z określoną, niewielką zawartością złota w rudzie, czasem jednak natrafimy na samorodek, czyli wartość zdecydowanie większą niż pozostałe. Jeżeli użyjemy takich obserwacji do obliczenia wariogramu, to okaże się, że istnienie tego samorodka silnie wpływa na wyniki obliczeń. Różnice wartości pomiędzy parami obserwacji, niezależnie od odległości pomiędzy nimi, wychodzą znacznie większe, niż gdyby tego samorodka nie było. Przypominam, że licząc wariogram, każda obserwacja bierze udział w obliczeniach wielokrotnie, bo wielokrotnie wchodzi w kombinacje par odległych od siebie o różne odległości *h.* Dla każdego *h* ta wartość zostanie podwyższona, bo różnica między samorodkiem a próbą z ubogiej rudy będzie wyraźna. To podniesienie wartości wariogramu dla $h = 0$ będzie widoczne na pierwszy rzut oka, bo aproksymowany wariogram przetnie oś pionową w miejscu wartości dużo większej od zera. Być może przykład ten nie jest aż tak oczywisty. Przecież w naszych polskich warunkach raczej nieczęsto badamy złoża złota, a mimo to efekt samorodka w kształcie wariogramów występuje niemal zawsze. W analizach zjawisk geologicznych jest to efekt typowy, a jego brak jest sytuacją nader wyjątkową.

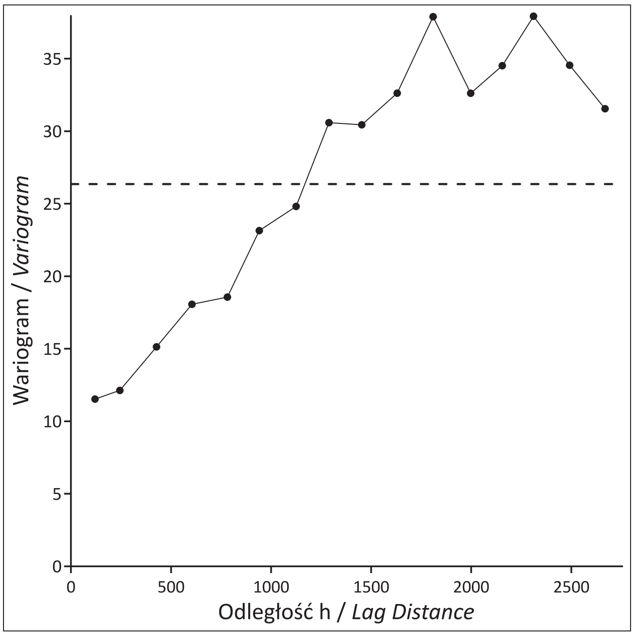
Obok zmienności samego zjawiska, jak to ma miejsce w złożach złota, w wartości $\gamma(h = 0) > 0$ uwidacznia się jeszcze jedna bardzo ważna z punktu widzenia analiz geostatystycznych cecha obserwacji. To błąd oznaczenia. Cóż to takiego? Każdy geodeta wie, co to jest błąd pomiaru, bo jest z tym zaznajamiany już od pierwszego roku studiów. Niegeodetom winny jestem wyjaśnienie. Przypuśćmy, że w sali, gdzie siedzą studenci, każdy otrzymuje linijkę o długości np. 1 m i ma do wykonania zadanie. Należy zmierzyć długość sali dydaktycznej. Szybko okaże się, że mamy tyle różnych wyników, ilu studentów wykonywało pomiar. Co więcej, każda osoba ponawiająca swój pomiar dojdzie do innego wyniku. Tak, bo mierzenie odległości, zwłaszcza linijką, nie jest najlepszą metodą pomiaru. Każde przyłożenie linijki niesie za sobą jakąś niedoskonałość, która przekłada się na większy lub mniejszy błąd całego procesu pomiaru. Ktoś powie, że byłoby lepiej, gdybyśmy wykonali pomiar dalmierzem laserowym zamiast przestarzałą metodą z użyciem przymiaru liniowego. Tak, błąd na pewno będzie mniejszy, ale różnice pomiędzy poszczególnymi pomiarami też wystąpią, bo ściana nie jest równa, bo linia pomiaru nie jest idealnie prostopadła do ściany itd.

JESZCZE RAZ O NIEDOSKONAŁOŚCI OBSERWACJI GEOLOGICZNYCH I ICH ZNACZENIU

Wróćmy do geologii. Tu wykonujemy pomiary czymś znacznie mniej doskonałym niż linijka albo dalmierz laserowy. Zacznijmy od pomiaru miąższości złoża, np. węgla brunatnego. Po wyciągnięciu rdzenia odmierzamy odległość między górną granicą występowania pokładu a dolną. Czy na pewno potrafimy dokładnie wyznaczyć granice między pokładem węgla a skałą płonną? Warstwy przystropowe mogą być właściwościami (kolorem, strukturą) zbliżone do węgla, tak jak osady przy spągu pokładu. Już tu możemy się pomylić. Taki pokład może mieć grubość 3 m albo 15 m, a nawet 45 m. Tę grubość można zmierzyć choćby taśmą, dalmierzem, ale przecież rdzeń wyciąga się z górotworu w częściach. Widać zatem, że pomiar grubości pokładu ani nie jest prosty, ani nie da się go wykonać w sposób bezbłędny. A jest to pomiar bezpośredni. Przykładamy linijkę i otrzymujemy wartość, która jest wartością wynikową i taką wpisujemy do tabeli.

To teraz zróbmy eksperyment myślowy – wbrew rozsądkowi wykonujemy otwory blisko siebie, załóżmy w odległości 1 m. Taka odległość w geologii jest niemal pomijalna. Oczekujemy, że wartość parametru pomierzona na tych dwóch sąsiadujących ze sobą rdzeniach będzie niemal równa, prawie tożsama. To intuicyjne i zgodne z doświadczeniami geologicznymi i górniczymi. Nie ma powodu, aby w tak bliskich obserwacjach występowały widoczne różnice. Po wykonaniu pomiaru okazuje się, że nasze liczby wcale nie są tożsame. Różnica, która objawiła się w pomiarach, najprawdopodobniej nie wynika z cechy samego badanego fenomenu tylko właśnie z błędu oznaczenia tej cechy. Taki błąd popełnimy w trakcie oznaczania wszystkich otworów w złożu i spowoduje on, że wariogram nie zacznie swojego biegu od wartości zero, tylko znacznie powyżej. Wartość wariogramu w miejscu przecięcia z osią pionową to będzie wartość błędu oznaczenia, średniego błędu, który popełniliśmy, oznaczając parametr we wszystkich otworach. Efekt tego błędu ma identyczne konsekwencje jak efekt samorodka w przywołanym przykładzie analiz złóż złota. Na ryc. 1 jest pokazany wariogram miąższości pokładu złoża węgla brunatnego z wyraźnie widocznym efektem samorodka.

Jeżeli jeszcze nie jesteś przekonany, co do wielkości błędu, jaki można popełnić w czasie oznaczania parametrów geologicznych i nie tylko, to mam dla Ciebie jeszcze jeden przykład. Weźmy ten sam węgiel brunatny, ale tym razem będzie to parametr niezwykle istotny dla paliw kopalnych, jakim jest wartość opałowa, nazywana też kalorycznością [kJ/kg]. Jest to ilość ciepła, jaką można uzyskać po spaleniu jednostkowej masy paliwa. Aby ją oznaczyć, należy pobrać z rdzenia wiertniczego, czasem o długości 15–20 m, reprezentatywną próbkę węgla, następnie trzeba ją zmielić, dokładnie wymieszać, czyli uśrednić, pobrać pewną próbkę z tej homogenicznej masy, następnie trzeba ją spalić w odpowiednim piecu i dokonać pomiaru ciepła, jakie się w tym procesie wydzieliło. Na każdym z etapów takiego pomiaru nawet świetny laborant nie tylko może, ale na pewno popełnia błędy. Opisana analiza w odróżnieniu od pomiaru miąższości pokładu jest analizą wielostopniową i niebezpośrednią. Począwszy od pobrania próby węgla z rdzenia, poprzez mielenie, mieszanie, palenie –



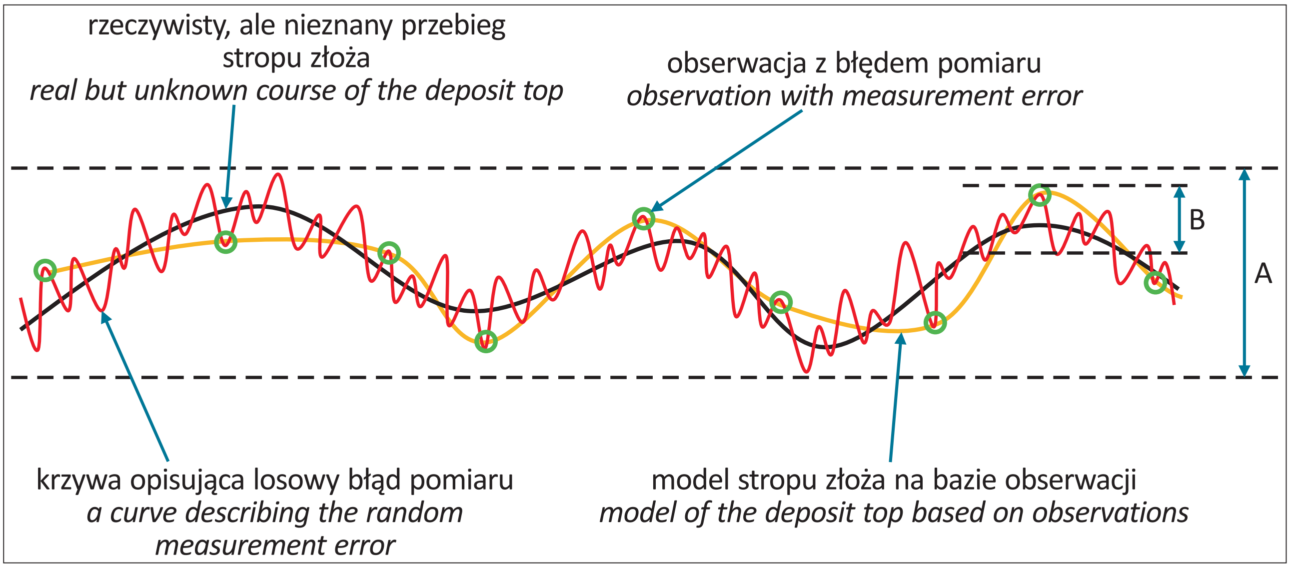
Ryc. 3. Wariogram popielności węgla *A* [%] w złożu węgla brunatnego
Fig. 3. Variogram of ash content *A* [%] in the coal of a lignite deposit

każdy z etapów wykonuje się z określoną dokładnością albo z określonym błędem. Błędy czasem się sumują, czasem odejmują, czasem niwelują, ale najczęściej nie pozostają bez wpływu na wynik pomiaru.

Innym parametrem, który bada się w węglach, jest popielność *A* [%]. To jest ilość popiołu, która powstanie po spaleniu próbki węgla. Popielność wylicza się w procentach w odniesieniu do masy węgla. Tak samo jak w pomiarze wartości opałowej należy najpierw pobrać próbkę z rdzenia, wysuszyć, wymieszać, wybrać część, spalić i dokonać pomiaru masy popiołu. Ta analiza również jest wielostopniowa i każdy jej etap generuje pewien błąd. Na ryc. 3 pokazałem wariogram popielności węgla brunatnego, na którym jest wyraźnie widoczny wysoki udział efektu samorodków, wynikający z błędu oznaczenia. Jego wartość to ok. 12%^{2}, czyli w jednostkach pomiaru ok. 3,5%. O tyle różnią się wartości popielności węgla w próbach pobranych bardzo blisko siebie. Dla porównania, wariancja odczytana z wykresu (pozioma przerywana linia) na poziomie 27%^{2} wskazuje, że wszystkie obserwacje różnią się statystycznie o wartość 5,2%. To pierwiastek z wariancji s^2, czyli odchylenie standardowe s.

Odznaczający się na wykresach wariogramów (ryc. 1, 3) błąd pomiaru będzie miał kolosalne znaczenie dla dalszych analiz. O tym jednak dalej. Wyjaśniając konstrukcję wzoru na wariogram, mówiłem o pewnych podobieństwach pomiędzy wariogramem i wariancją. Rzeczywiście, pojęcia te mają ze sobą dużo wspólnego. W literaturze przedmiotu możemy spotkać pojęcie wariancja lokalna jako równoznaczne z efektem samorodków (*nugget effect*). Mnie ta nazwa jest znacznie bliższa niż ta pochodząca od złóż złota. Wariancja lokalna to statystyczna kwadratowa różnica pomiędzy bardzo bliskimi obserwacjami. Tak tłumaczę sobie tę nazwę, która nie oznacza nic innego jak opisany wcześniej efekt samorodków. Chciałbym teraz zilustrować wpływ wariancji lokalnej na obserwację zjawiska przestrzennego, przykładowego parametru złoża, np. rzędnej stropu pokładu.

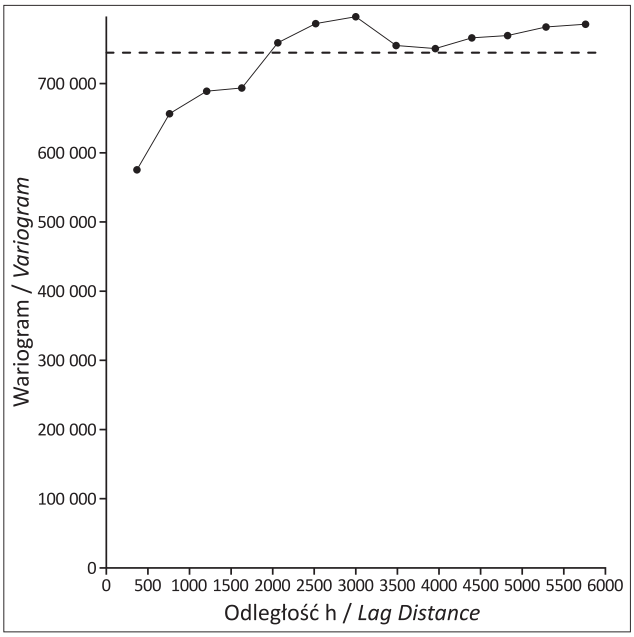
Na ryc. 4 czarną linią pokazałem przebieg stropu złoża, gdybyśmy, wbrew naturze, mogli go zobaczyć. Zielonymi punktami pokazałem obserwacje wykonane w trakcie wierceń. One nie pokrywają się z przebiegiem stropu złoża, bo podczas oznaczenia popełniono błąd. Błąd, jak to błąd, ma charakter losowy, czasem jest duży, czasem mały, czasem dodatni, czasem ujemny. Linia czerwona ma ilustrować zjawisko błędu. On towarzyszy wszystkim obserwacjom, bo oznaczenie parametru nigdy nie jest bezbłędne. Na podstawie obserwacji w procesie interpolacji wykonałem model złoża (żółta linia). Model opisuje strop, jednak przez błąd pomiaru model nie pokrywa się z rzeczywistym, ale niewidocznym dla nas przebiegiem stropu złoża. A teraz uwaga! Zakres zmienności parametru złożowego oznaczyłem literą *A*, zakres błędu pomiaru literą *B*. Dla modelowania zjawisk przestrzennych kluczową



Ryc. 4. Ilustracja wpływu błędu pomiaru na efekt interpolacji zjawiska na przykładzie stropu złoża pokładowego
Fig. 4. Illustration of the influence of measurement error on the effect of interpolation of the phenomenon, exemplified by the top of a bed deposit

rolę ma stosunek B do A. Im będzie on większy, tym mniej dokładny będzie model. Ten stosunek będzie widoczny na wariogramie. Za wielkość A odpowiada poziom wariancji na wariogramie, a za B wariancja lokalna (*nugget effect*). Im bardziej ta ostatnia zbliża się do wariancji ogólnej, tym gorszy będzie efekt modelowania. W skrajnym przypadku, gdy wariancja lokalna jest równa wariancji ogólnej, to mamy do czynienia ze zjawiskiem losowym i o modelowaniu możemy zapomnieć, bo w zmienności zjawiska nie ma autokorelacji. Uwaga, wielkości A, B nie są tożsame z wariancją ogólną i wariancją lokalną. Te ostatnie to wielkości kwadratowe, a A i B określają zakresy w jednostkach pierwotnych, np. tu konkretnie w metrach nad poziomem morza, bo mierzymy rzędne stropu pokładu złoża.

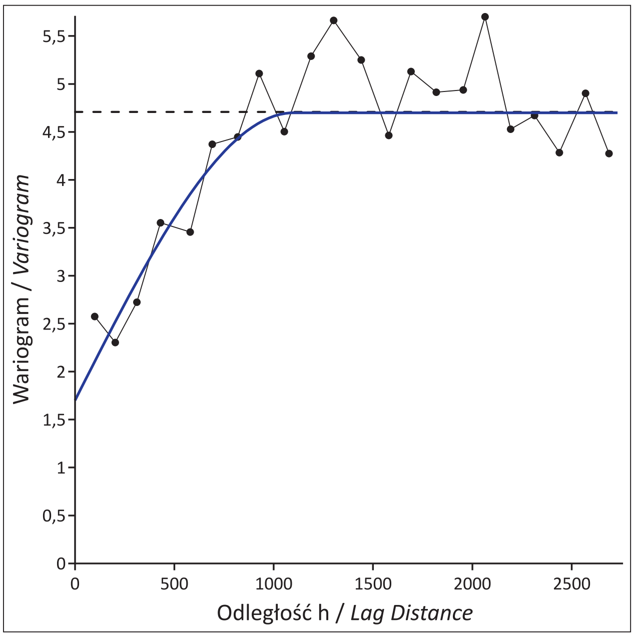
Do opisu zjawiska przedstawionego na ryc. 4 można zastosować język radiowy. Duże fale o dużej amplitudzie i dużej długości (linia czarna), to jest rzeczywista zmienność parametru złoża. Nazwijmy je sygnałem. Niewielkie falki (linia czerwona) o mniejszej amplitudzie i bardzo małej długości to jest wyobrażenie błędu pomiaru. Te nazwiemy szumem. Błąd pomiaru to jest właśnie taki szum, który zakłóca odbiór głównego sygnału, jakim są te duże fale oznaczające faktyczną, geologiczną zmienność badanego przez nas parametru złoża. Od błędu pomiaru, czyli od tego szumu bardzo zależy dokładność naszych analiz. Jeżeli udział szumu w stosunku do sygnału będzie znaczący, to dokładność naszych badań będzie nikła, i odwrotnie, jeśli udział tego szumu w stosunku do sygnału będzie znikomy, to dokładność badań będzie wysoka. Błąd pomiaru zakłóca obserwację, tak jak brudne szkło okularów utrudnia obserwację świata. Szum albo błąd pomiaru uwidacznia się w wielkości wariancji lokalnej. Możliwość jej wyznaczenia to trzecia arcyważna, choć jeszcze nie ostatnia funkcja wariogramu. W trzecim artykule, opisując interpolację metodą krigingu, wyjaśnię, jak ogromny wpływ na dokładność modelu analizowanego zjawiska przestrzennego ma wariancja lokalna. Na ryc. 5 pokazałem przykładowy wariogram, w którym udział wariancji lokalnej w stosunku do wariancji ogólnej jest bardzo duży. W przebiegu zjawiska jest widoczna autokorelacja, owszem, ale mając tak istotną wartość błędu pomiaru albo, stosując język radiowy, tak duży udział szumu w stosunku do sygnału, nie możemy się spodziewać, że model, który wykonamy metodą krigingu będzie dokładny. Ja osobiście w takim przypadku odstąpiłbym od modelowania zjawiska. Dokładność modelu będzie znikoma.



Ryc. 5. Przykładowy wariogram z dużym udziałem wariancji lokalnej (*nugget effect*) w stosunku do wariancji ogólnej
Fig. 5. An example of a variogram with a high share of local variance (nugget effect) in relation to the general variance

MODELOWANIE WARIOGRAMU

Wszystkie wykresy na rycinach 1, 2, 3 i 5 to tzw. wariogramy empiryczne – doświadczalne, obliczone na podstawie obserwacji, czyli danych pomiarowych. Wariogram empiryczny nie jest funkcją – jest to zbiór wartości dyskretnych dla określonych, przyjętych przez operatora wartości h, czyli odległości między parami obserwacji. Te wartości są pokazane w postaci kropek na wykresach, a linie, które je łączą, właściwie nie mają żadnego matematycznego znaczenia. Ułatwiają tylko interpretację wykresu. Aby ten wariogram mógł być przydatny w procedurze interpolacji metodą krigingu, musi być określony w całej dziedzinie, dla każdego h. I tu właśnie konieczny jest świadomy udział operatora, czyli Twój. Do wariogramu empirycznego trzeba dopasować funkcję – nazwiemy ją modelem wariogramu. W programach komputerowych jest dostępna opcja automatycznego dopasowania modelu do wariogramu empirycznego (*autofitting*). Moje doświadczenia wskazują, że efekty tej automatyzacji są dalece niedoskonałe. Manualna czynność modelowania wariogramu jest prosta, wręcz intuicyjna. Do wyboru są gotowe funkcje – modele wariogramu, spośród których należy wybrać najlepiej pasującą i manewrując kilkoma parametrami dopasować ją do wariogramu empirycznego. Ja to robię na oko. Modelowanie wariogramu to pierwsza ważna czynność w analizie geostatystycznej, w której operator musi wykazać inicjatywę i świadomie wykonać pewne czynności. Tu już kończy się etap przypadkowego klikania. Czuję się w obowiązku wyjaśnić kilka ważnych czynności. Na ryc. 6 przedstawiłem wariogram empiryczny z dopasowaną funkcją – modelem wariogramu. W kształcie wariogramu empirycznego jest widoczny *nugget effect* (wariancja lokalna), wartości wariogramu rosną (zatem wiemy już, że istnieje zjawisko autokorelacji) i po osiągnięciu h = 1000 m załamują się i przybierają przebieg zbliżony do wariancji, która jest wyznaczona linią przerywaną. Linia niebieska to model wariogramu, który aproksymuje wariogram empiryczny. Na wykresie widać jedną linię, ale formalnie model składa się z dwóch komponentów. Pierwszym jest *nugget effect,* a dopiero drugim model wyznaczony linią niebieską. Ten pierwszy jest określony przez jedną wartość (tu: 1,7 m^2), drugi to funkcja dopasowana za pomocą dwóch parametrów. Jeden określa jej wysokość, a raczej wartość maksymalną, przy której wariogram przestaje rosnąć (*sill*), a drugi to zasięg (*lenght, range*) i jest to wartość h, dla której wariogram załamuje swój przebieg i dalej biegnie poziomo. Już wiemy, że ta wartość oznacza zasięg autokorelacji. W modelowaniu wariogramu nie ma żadnej filozofii. Wymienione parametry są dobierane intuicyjnie, na oko, iteracyjnie. Ważne jest, aby wiedzieć, że na największą wartość modelu, która jest określona przez wariancję, składa się suma wariancji lokalnej (*nugget*



Ryc. 6. Wariogram empiryczny z dopasowanym modelem (linia niebieska)
Fig. 6. Empirical variogram with a fitted model (blue line)

effect) oraz wysokość drugiego komponentu modelu (*sill*). Istotny jest też wybór funkcji spośród dostępnych w programie opcji. Oprogramowanie geostatystyczne ma wiele modeli wariogramów do wyboru: wykładniczy, model Gaussa, model de Wijsa, sferyczny, liniowy, losowy i wiele innych. Są to funkcje określone wzorami, których wcale nie musimy znać. Program będzie z nich korzystał poza naszymi plecami. Ważny jest jednak wybór właściwej funkcji. Moje wieloletnie doświadczenia wskazują, że wariogramy zjawisk geologicznych w znakomitej większości można modelować, używając modelu sferycznego, rzadziej modelu Gaussa. W wielu przypadkach możliwe jest też zastosowanie wariogramu liniowego, ale wyłącznie w ograniczonym zakresie zasięgu *h*. Wykresy modeli – sferycznego i Gaussa – różnią się przede wszystkim w przedziale małych wartości *h*. Sferyczny zaczyna swój bieg ostro i od razu pnie się w górę, podczas gdy model Gaussa zaczyna swój wzrost bardzo łagodnie, niemal poziomo. Nazwano go modelem Gaussa, bo jest to *de facto* odwrócona połowa krzywej Gaussa, opisująca normalny rozkład statystyczny. Model Gaussa i sferyczny różnią się również w przebiegu górnych wartości wariogramu, w sposobie załamania dla *h* oznaczających zasięg autokorelacji.

Na ryc. 6 pokazałem wariogram empiryczny miąższości pokładu węgla brunatnego z dopasowanym modelem sferycznym. Model ułatwia oszacowanie najważniejszych parametrów zmienności. Wariancja lokalna, wyznaczona przecięciem modelu z osią pionową, ma wartość 1,7 m^2, a zasięg autokorelacji, czyli załamanie modelu, następuje dla wartości h = 1100 m.

PODSUMOWANIE

Wariogram to niezwykle przydatne narzędzie geostatystyczne. Z analizy obserwacji za pomocą wariogramu można dowiedzieć się o istnieniu lub braku istnienia autokorelacji w zmienności analizowanego zjawiska przestrzennego. Jeżeli autokorelacja istnieje, to korzystając z wariogramu da się określić jej zasięg. Jeżeli jednak stwierdzimy, że autokorelacja nie występuje to ta wiedza pozwoli nam uniknąć błędów, jakie popełnilibyśmy, uparcie i wbrew logice modelując zjawisko. W zidentyfikowanej wariogramem wartości wariancji lokalnej odbija się zmienność lokalna zjawiska, która bardzo często ma swoje źródło w błędzie pomiaru. Oprócz wymienionych funkcji analitycznych wariogram empiryczny i jego model pełnią ważną funkcję użytkową w procesie interpolacji metodą krigingu. Bez wariogramu nie zastosujemy krigingu. W procesie modelowania wariogramu kluczowe jest możliwie najlepsze dopasowanie modelu w obszarze wariancji lokalnej oraz w części wariogramu empirycznego, w której jest widoczny regularny wzrost, czyli w obszarze istnienia autokorelacji. Od tego będzie zależeć dokładność modelu, czyli wartość informacyjna mapy modelowanego zjawiska. Wszystko, co do tej pory napisałem, jest podporządkowane temu właśnie celowi, bo geostatystyka służy analizie i modelowaniu zjawisk przestrzennych. O wykorzystaniu wariogramu w procesie modelowania metodą krigingu, a także o błędzie tego modelu wypowiem się szeroko w trzecim artykule cyklu pt. *Geostatystyka dla niematematyków.*

LITERATURA

MATHERON G. 1963 – Principles of geostatistics. Economic geology, 58 (8): 1246–1266.

Praca wpłynęła do redakcji 17.07.2024 r.
Akceptowano do druku 4.08.2024 r.