

Zastosowanie metod geostatystyki w doświadczalnictwie polowym¹

Janusz Gołaszewski

Katedra Hodowli Roślin i Nasiennictwa, Akademia Rolniczo-Techniczna
w Olsztynie, Pl. Łódzki 3, 10-724 Olsztyn

STRESZCZENIE

Doświadczenia polowe prowadzone w warunkach znacznej heterogenności gleby są zwykle obciążone dużym błędem doświadczalnym, co w efekcie może prowadzić do słabej oceny porównywanych obiektów. Od stopnia znajomości charakteru zmienności glebowej danego pola doświadczalnego zależą, w dużej mierze, racjonalne przedsięwzięcia eksperymentalne związane zarówno z metodyką planowania, zakładania i prowadzenia doświadczenia, jak również ze sposobem opracowania danych. Względnie nowe w szacowaniu zmienności przestrzennej jest podejście geostatystyczne. Pozwala ono m.in. na określenie wielkości wariancji losowej i wariancji strukturalnej, przypisanej przestrzennemu skorelowaniu obserwacji.

W pracy przedstawiono założenia teoretyczne oraz algorytm obliczeniowy podstawowych metod statystyki przestrzennej: semiwariogramu i krigingu. Wskazano na możliwe kierunki zastosowania tych metod w doświadczalnictwie polowym. Stwierdzono, że szacowanie zmienności glebowej pola doświadczalnego metodami statystyki przestrzennej może ułatwić optymalizowanie metodyki doświadczenia polowego, stanowić narzędzie diagnostyczne w wyborze odpowiedniej metody statystyki klasycznej do opracowania danych z doświadczenia oraz dokonanie analizy wyników wieloletnich doświadczeń z dziedziny nauk o nawożeniu i żywności.

SŁOWA KLUCZOWE: doświadczalnictwo polowe, geostatystyka, przestrzenne skorelowanie obserwacji, semiwariogram, kriging

1. Wstęp

W doświadczalnictwie polowym jednym z czynników losowych jest zmienność glebowa. Sprawia ona, że w pozornie jednolitych warunkach glebowych, uzyskujemy plony różniące się, niekiedy nawet dość znacznie. Dlatego też, racjonalnie zaplanowane doświadczenie polowe musi pogodzić teoretyczne rozważania związane z wybo-

¹ Praca wykonana w ramach projektu badawczego KBN 5 PO6A 00711

rem układu doświadczalnego oraz określeniem wielkości poletka i liczby bloków ze stanem faktycznej zmienności przestrzennej pola doświadczalnego.

Podstawą efektywnego doświadczenia polowego jest wybór właściwej wielkości poletka. Decyzja odnośnie do jego wielkości powinna być uzasadniona ekonomicznie, a równocześnie – w celu poprawnej interpretacji danych z doświadczenia – powinna gwarantować “najlepszą” ocenę błędu doświadczalnego (Binns, 1982). Kumulatywne efekty mikrozmienności glebowej między poletkami mogą maskować rzeczywiste różnice obiektowe. Wykrywanie tych efektów, a następnie ich eliminowanie dzięki odpowiednim rozwiązaniom technicznym (kształt poletek, pojemność bloku, zorientowanie poletek itp.) bądź (oraz) zastosowanie odpowiednich metod statystycznych umożliwiają ostatecznie dokładne porównania międzyobektowe.

Precyzyjną ocenę charakteru naturalnej zmienności glebowej pola doświadczalnego zapewnia analiza wyników doświadczenia beczynnikowego (ślepego). Doświadczenie to obejmuje gęstą sieć mikropoletek, tzw. jednostek podstawowych (np. 1 m^2), obsianych jednym gatunkiem rośliny. Im mniejsze są jednostki podstawowe, tym precyzyjniejsza jest ocena zmienności glebowej. Plony zbierane z każdej jednostki podstawowej wykazują przestrzenne zróżnicowanie, odzwierciedlając charakter zmienności glebowej. Łącząc sąsiednie jednostki ze sobą symuluje się coraz większe poletka. Z każdą z kolejnych wielkości poletek związana jest pewna zmienność. Ujęcie kolejnych wielkości poletek i odpowiadających im wariacji w formułę równania regresji logarytmicznej prowadzi do estymacji wskaźnika zmienności glebowej b (Smith, 1938). Jest to summaryczna dla danego doświadczenia wielkość, określająca stopień skorelowania między sąsiednimi jednostkami. Wartość b zawiera się w przedziale $(0, 1)$; im bliższa jest jedności, tym większa jest niezależność informacji z sąsiednich poletek oraz większe wyrównanie gleby. Jednakże, biorąc pod uwagę koszty związane z prowadzeniem doświadczenia beczynnikowego, a także z uwagi na fakt, że uzyskane wyniki odnoszą się tylko do określonego pola, doświadczenia tego typu zakłada się sporadycznie.

Alternatywną w stosunku do doświadczeń beczynnikowych, jest analiza zmienności glebowej na podstawie niektórych układów doświadczeń polowych, w których, z założenia metody, występuje kilka wielkości poletek (Koch i Rigney, 1951). Przykładowo, w układzie losowanych bloków z jednym czynnikiem występują poletka wielkości jednostkowej i poletka wielkości bloku, natomiast w układzie split-plot z dwoma czynnikami występują poletka wielkości jednostkowej, poletka wielkości podbloku i poletka wielkości bloku. Efektem analizy tych doświadczeń może być, podobnie jak w przypadku doświadczeń beczynnikowych, oszacowanie wskaźnika zmienności glebowej. Dodatkowe informacje dotyczące zmienności glebowej, możliwe do uzyskania z doświadczeń czynnikowych a pomijane zwykle przez doświadczalników, mogłyby stanowić poszerzone tło dla interpretacji wyników. Przykłady praktycznej estyma-

cji wskaźnika zmienności glebowej z doświadczeń czynnikowych, jak również wady i zalety takiego podejścia, prezentuje praca Gołaszewskiego (1996).

Jak dotychczas, wskaźnik zmienności glebowej wykorzystywano tylko w pracach nad optymalizacją wielkości jednostek eksperymentalnych dla różnych gatunków roślin uprawnych (np. Trętowski, 1975; Rejman i Szczepański, 1978; Gołaszewski, 1995, 1996). Brak jest w literaturze innych przykładów wykorzystania tego wskaźnika.

Kolejną możliwość oceny zmienności glebowej stwarza zastosowanie analizy geostatystycznej pola. Metody geostatystyczne umożliwiają m.in. określenie przestrzennej zależności obserwacji w różnych kierunkach geograficznych, zakresu skorelowania obserwacji i predykcji wyników na podstawie parametrów zmienności przestrzennej (kriging). Jakkolwiek metody te są powszechnie stosowane w kopalnictwie złóż mineralnych, kartografii i badaniach pedologicznych, to w odniesieniu do doświadczalnictwa polowego wciąż jest mało przykładów ich praktycznego wykorzystania. W polskojęzycznym piśmiennictwie praca Gołaszewskiego (1996) jest próbą przybliżenia idei tych metod oraz ich aplikacji w badaniach związanych z optymalizacją metodyki doświadczeń z roślinami strączkowymi. Innymi przykładami zastosowania metod geostatystyki w naukach przyrodniczych przez polskich autorów są ponadto prace Krajewskiego i in. (1994, 1995), dotyczące analizy przemieszczania się zanieczyszczeń w powietrzu oraz Usowicz i in. (1995), którzy wykorzystali metody geostatystyczne w kwantyfikacji przestrzennych zmian wybranych parametrów właściwości fizyko-chemicznych gleby w różnych uprawach rolniczych.

Niniejsza praca ma na celu przegląd ważniejszych metod geostatystycznych z punktu widzenia możliwości ich wykorzystania w doświadczalnictwie polowym, oraz prezentację niektórych z nich na przykładzie doświadczenia bezczynnikowego z łubinem żółtym.

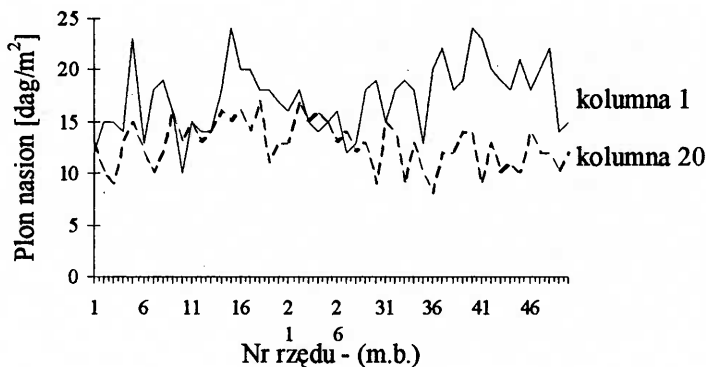
2. Definicja i podstawowa interpretacja geostatystyki

Etymologicznie, pojęcie „geostatystyka” określa statystyczne badanie zjawisk przyrodniczych. Definicja podana w 1962 r. przez francuskiego matematyka G. Matherona – prekursora formalistycznego podejścia do zagadnień statystyki przestrzennej – mówi: „Geostatystyka jest aplikacją sformalizowanych funkcji losowych w celu rozpoznawania i estymacji zjawisk przyrodniczych”.

Teoretyczne podstawy geostatystyki przedstawił Matheron (1962, 1971) w pracy „*Traité de Géostatistique Appliquée*”, a następnie w angielskojęzycznej jej wersji „*The theory of regionalized variables and its application*”. Formułując swoją teorię Matheron wykorzystał doświadczenia D.G. Krige, który do poszukiwania złóż złotonośnych w Afryce Południowej zastosował własną metodę interpolacyjną, określaną (od nazwiska autora) jako „kriging” (Krige, 1966). Pierwotny kierunek wykorzystania metod

statystyki przestrzennej w kopalnictwie złóż naturalnych znalazł swoje odzwierciedlenie w nazewnictwie niektórych parametrów estymowanych tymi metodami, jak np. wariancja i efekt samorodka¹.

Zjawiska przyrodnicze mogą być opisane przestrzennym rozkładem jednej lub wielu zmiennych losowych. Z matematycznego punktu widzenia wartość plonu w dowolnym punkcie pola jest funkcją jego pozycji na polu. Jednakże nie ma jednoznacznej formuły matematycznej opisującej dane zjawisko, ponieważ niekoniecznie rozkład zmiennej musi być regularny. Rozpatrzmy plony z jednostek podstawowych doświadczenia beczynnikowego, rejestrowane zgodnie z kierunkiem przebiegu wybranych kolumn (Ryc.1).



Ryc. 1. Zmienność plonów z jednostek podstawowych w wybranych kolumnach doświadczenia beczynnikowego z hubinem żółtym

Plony w kolumnie 1 wykazują wyraźne ukierunkowanie zmienności, natomiast plony w kolumnie 20 fluktuują wokół pewnej wartości przeciętnej. Zarówno w jednym, jak i w drugim przypadku istnieje określona struktura zmienności, związana z podobieństwem obserwacji sąsiadujących ze sobą, tym silniejszym im mniejszy jest dzielący je dystans. Zatem można wnioskować, że zmienność danej właściwości przyrodniczej opisywanej w przestrzeni musi być rozpatrywana w kategoriach statystycznych. W geostatystyce zmienna losowa zwana jest "zmienną regionalizowaną", natomiast zjawisko przyrodnicze, które na podstawie rozkładu zmiennej regionalizowanej jest opisywane określa się terminem "regionalizacja", np.

– plony z poletek w doświadczeniu, w którym bloki umieszczono w jednym pasie

¹ Znalazienie złotego samorodka w warstwie złoża jest zdarzeniem przypadkowym, mającym charakter losowy. Średnica znalezionych samorodków jest zwykle niewspółmiernie mniejsza aniżeli odległość dzieląca kolejne pomiary.

doświadczalnym, można rozpatrywać jako rozkład tej zmiennej (plonów) w przestrzeni jednowymiarowej,

– zmienność glebową pola doświadczalnego mogą opisywać plony doświadczenia bezczynnikowego, rejestrowane zgodnie z układem rzędów i kolumn (przestrzeń dwuwymiarowa),

– zjawiska i przemiany zachodzące w glebie można opisywać za pomocą zmiennych regionalizowanych takich jak np. zawartość N , skład granulometryczny gleby, gęstość gleby, itp., które wykazują zmienność w przestrzeni trójwymiarowej.

Niech $z(x)$ będzie wartością zmiennej losowej stacjonarnej $Z(x)$ w punkcie x . Klasyczna statystyka zakłada, że wartość pewnej właściwości z w dowolnym punkcie x przestrzeni, znajdującym się w obrębie analizowanej powierzchni, wynosi:

$$z(x) = \mu + \varepsilon(x), \quad (1)$$

gdzie μ jest średnią populacyjną lub wartością oczekiwaną z , a $\varepsilon(x)$ określa losowe, nieskorelowane przestrzennie, rozproszenie wartości wokół średniej. Zakłada się, że odchylenia od średniej populacyjnej mają rozkład normalny z wartością oczekiwaną zero i wariancją σ^2 .

Problem natury geostatystycznej polega na przedstawieniu zmienności funkcji $Z(x)$ w przestrzeni, gdy zmienia się x . Określenie tej funkcji pozwala m.in. na predykcję wartości $z(x_0)$ w punkcie x_0 , dla którego brak jest danych.

3. Teoria zmiennych regionalizowanych

Geostatystyka opiera się na koncepcji zmiennych regionalizowanych, funkcjach losowych i stacjonarności. Skrótowy przegląd tych koncepcji z punktu widzenia ich praktycznego wykorzystania w doświadczałnictwie polowym będzie przedmiotem poniższych rozważań. Całościowe ujęcie zagadnień związanych ze statystyką przestrzenną można znaleźć w książkach Journela i Huijbregtsa (1978) i Clarka (1979) (kopalnictwo) oraz Webstera i Olivera (1990) (gleboznawstwo).

Załóżmy, że $z(x_1)$ reprezentuje pewną obserwację (np. zawartość przyswajalnego potasu w glebie) dokonaną w danym punkcie przestrzeni x_1 . Obserwację taką możemy uważać jako szczególną realizację pewnej zmiennej losowej $Z(x_1)$ w punkcie x_1 . Z kolei, zestaw obserwacji $z(x)$ dla wszystkich punktów analizowanej powierzchni, czyli zmienna regionalizowana $z(x)$, może być uważana jako szczególna realizacja zmiennej losowej $Z(x)$.

Stąd też, lokalnie, w punkcie x_1 , $Z(x_1)$ jest zmienną losową.

Zmienna losowa $Z(x)$ jest również funkcją losową, w tym rozumieniu, że dla każdej pary punktów w przestrzeni x_1 i x_2 odpowiednie zmienne losowe $Z(x_1)$ i $Z(x_2)$ nie

są niezależne, lecz skorelowane, odzwierciedlając tym samym przestrzenną strukturę zmiennej regionalizowanej $z(x)$.

Charakterystyka zmienności odnosząca się do całej populacji, a nie tylko określonego zestawu danych, wymaga wprowadzenia pojęcia stacjonarności. Oznaczmy dwa punkty w przestrzeni jako x i $x + h$, gdzie x oznacza współrzędne punktu w jednym, dwu lub trzech kierunkach, a h jest wektorem zarówno odległości jak i kierunku, określonym w literaturze przedmiotu jako lag.

Funkcja losowa $Z(x)$ jest stacjonarną pierwszego stopnia, gdy jej wartość oczekiwana nie zależy od x , tzn. we wszystkich punktach analizowanej powierzchni jest taka sama:

$$E[Z(x)] = \mu \quad (2)$$

oraz

$$E[Z(x) - Z(x + h)] = 0, \quad (3)$$

gdzie h jest wektorem odległości między punktami pobierania prób.

Funkcja losowa $Z(x)$ jest stacjonarną drugiego stopnia gdy spełniony jest warunek (2), a kowariancja $C(h)$ dla każdej pary $\{Z(x), Z(x + h)\}$ jest postaci:

$$C(h) = E[(Z(x) - \mu)(Z(x + h) - \mu)] = E[(Z(x) \cdot Z(x + h)) - \mu^2], \quad (4)$$

i zależy od h , a nie zależy od x . Kowariancja przestrzenna zbliża się do wariancji próby gdy odległość dzieląca punkty pobierania prób zbliża się do zera.

W odniesieniu do próby, wariancja m -par obserwacji dla wektora h wynosi

$$\bar{s}^2 = \frac{1}{2m} \sum_{i=1}^m [z(x_i) - z(x_i + h)]^2, \quad (5)$$

gdzie $i = 1, 2, \dots, m$, przy czym m określa liczbę par obserwacji $\{z(x), z(x + h)\}$, które dzieli wektor odległości h .

Słabsza forma stacjonarności, określana terminem "hipoteza wewnętrzna (intrinsic hypothesis)" (Journal i Huijbregts, 1978), zakłada, że $E[Z(x)] = \mu$ oraz dla wszystkich wektorów h , wariancja różnicy $z(x) - z(x + h)$ jest skończona i niezależna od położenia w obrębie analizowanej powierzchni, tj.

$$\text{var}[Z(x) - Z(x + h)] = E[Z(x) - Z(x + h)]^2 = 2\gamma(h). \quad (6)$$

W przypadku gdy skończonej wariancji i kowariancji nie można określić, jak ma to miejsce w warunkach obecności trendu żyzności gleby, a więc ukierunkowanych zmian żyzności, wtedy stacjonarność drugiego stopnia nie jest zachowana. Obecność trendu (np. żyzności gleby) w obrębie analizowanej powierzchni, wymaga jeszcze innego podejścia, mianowicie oceny reszt $\varepsilon(x)$ z równania (1), gdzie $\varepsilon(x)$ jest przestrzennym

komponentem losowym ze średnią zero i wariancją

$$\text{var}[\varepsilon(x) - \varepsilon(x+h)] = E[\varepsilon(x) - \varepsilon(x+h)]^2 = 2\gamma(h). \quad (7)$$

Jednakże, biorąc pod uwagę fakt, że ostateczne oszacowania zmienności przestrzennej wykorzystywane do predykcji z w punktach nie objętych pomiarami, odnoszą się do powierzchni, dla której odległości dzielące próby są relatywnie małe (zwykle od $3h$ do $5h$), zakłada się stacjonarność w obrębie tej powierzchni. Warunek ten jest określany jako "quasi-stacjonarność".

Funkcja $2\gamma(h)$ jest wariogramem, natomiast funkcja $\gamma(h)$ odpowiednio semiwariogramem. Wyrażenie $\gamma(h)$ w (6) jest wartością oczekiwaną s^2 dla danego h . Graficzny rozkład semiwariancji $\hat{\gamma}(h)$ wyznaczonych dla kolejnych odległości h jest najczęściej stosowanym sposobem prezentacji semiwariogramu.

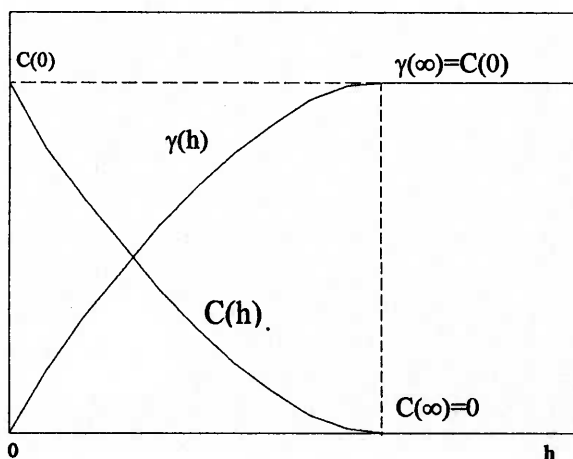
Relacje między kowariancją a semiwariogramem przedstawia Ryc. 2 (za Journel i Huibregts, 1978). Ponieważ korelacja między dwiema zmiennymi $z(x)$ i $z(x+h)$ zmniejsza się zwykle ze wzrostem odległości między nimi, dlatego też funkcja kowariancji zmierza do zera ($C(h) \rightarrow 0$, gdy $|h| \rightarrow \infty$).

Jak wynika z (4), dla lag $h = 0$,

$$C(0) = E[Z^2(x)] - \mu^2 = \sigma^2. \quad (8)$$

Iloraz $C(h)$ i $C(0)$, oznaczony przez $\rho(h)$, jest współczynnikiem autokorelacji dla lag h ,

$$\rho(h) = \frac{C(h)}{C(0)}. \quad (9)$$



Ryc. 2. Kowariancja i semiwariogram

Jego wartość zawiera się w przedziale od -1 do 1 , a o silniejszej korelacji mówią wartości zbliżone do 1 lub -1 . Podobnie jak w przypadku semiwariogramu, który opisuje relację między semiwariancją a wektorem h , korelogram odnosi kowariancję i autokorelację do h . Wymienione trzy parametry $\gamma(h)$, $C(0)$ i $\rho(h)$ są ekwiwalentne, a wzajemne relacje między nimi opisuje poniższa zależność:

$$\gamma(h) = C(0) - C(h) = \sigma^2[1 - \rho(h)]. \quad (10)$$

Jak podają Webster i Oliver (1990), często w praktyce zdarza się taka sytuacja, że nie można przyjąć założenia o stałej (dla danego pola) i skończonej wariancji, np. w sytuacji, gdy semiwariancja wyznaczana dla określonego zakresu danych nie osiąga granicy, przy której następuje jej stabilizacja. Wtedy autokorelacja, w przeciwieństwie do semiwariancji, nie ma żadnego praktycznego znaczenia. Dlatego też, podstawą teorii zmiennych regionalizowanych jest przede wszystkim semiwariogram oraz jego parametry.

4. Semiwariogram i jego parametry

Funkcja losowa $Z(x)$ ujmuje zarówno losowe, jak i strukturalne aspekty zmiennej regionalizowanej. Semiwariogram $\gamma(h)$ opisuje ten komponent funkcji losowej $Z(x)$, który jest związany z przestrzenną zależnością obserwacji:

$$\gamma(h) = \frac{1}{2} E[Z(x) - Z(x+h)]^2. \quad (11)$$

Wyznaczona dla danego h jednokierunkowa semiwariancja (np. NS lub EW lub NW-SE, itd.) jest średnim kwadratem różnic między wszystkimi obserwacjami, które dzieli wektor h :

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i) - z(x_i+h)]^2, \quad (12)$$

gdzie $N(h)$ jest liczbą par obserwacji dla danego h .

Rozkład wartości semiwariancji w analizowanej próbie, wykorzystanych do wyznaczenia semiwariogramu, jest izotropowy, jeśli semiwariancje estymowane dla różnych kierunków w przestrzeni są jednakowe, w przeciwnym razie mamy do czynienia z anizotropią. W zależności od stwierdzonego rozkładu, różnią się metody wyznaczenia semiwariogramu. Anizotropia danych wymusza najczęściej konieczność transformacji danych i stosowanie zaawansowanego aparatu statystycznego, natomiast w przypadku izotropii semiwariancje liczone dla różnych kierunków w przestrzeni można wykorzystać do określenia ogólnego – dla analizowanej powierzchni pola – semiwariogramu. Oczywiście, we wszystkich przypadkach, gdy rozkład zmiennej (cechy) jest inny niż

normalny, transformacja danych jest konieczna.

Dla dwukierunkowego rozkładu danych, pogrupowanych w r rzędów i c kolumn semiwariancję opisują następujące równania:

$$\hat{\gamma}(p, q) = \frac{1}{2(r-p)(c-q)} \sum_{i=1}^{r-p} \sum_{j=1}^{c-q} [z(i, j) - z(i+p, j+q)]^2, \quad (13)$$

gdzie p i q są odpowiednio wektorem p i wektorem q w dwu kierunkach. Równania te umożliwiają wyznaczenie połowy wariogramu w odniesieniu do odległości od $-q$ do q oraz od 0 do p . Wariogram jest symetryczny względem centrum analizowanej powierzchni, stąd pełny zestaw semiwariancji uzupełniają:

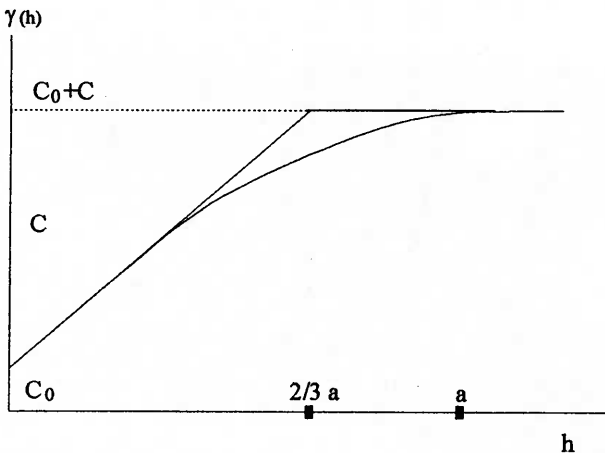
$$\hat{\gamma}(-p, q) = \hat{\gamma}(p, -q) \quad \text{oraz} \quad \hat{\gamma}(-p, -q) = \hat{\gamma}(p, q). \quad (14)$$

Jak wspomniano, graficzny rozkład semiwariancji w zależności od h , reprezentuje semiwariogram. Kształt semiwariogramu może być różny, zależnie od rodzaju danych wykorzystywanych w analizie oraz przyjętej "siatki" dokonywania pomiarów.

Typowy semiwariogram (Ryc. 3) określają następujące parametry:

- zakres (a) - określa granicę przestrzennej zależności obserwacji; zakłada się, że semiwariancje poza zakresem a są równe s^2 i tłumaczy je zmienność losowa; zakres określa ponadto maksymalny promień, w zasięgu którego pobiera się sąsiednie obserwacje do krigingu,

- wariancja samorodka (efekt samorodka) (C_0) - nieciągłość wariogramu względem środka układu współrzędnych, wynikająca z błędów pomiarowych oraz mikrozmienności w zakresach odległości mniejszych niż dokonywano pomiary; procentowe wyrażenie



Ryc. 3. Teoretyczny semiwariogram

udziału wariancji samorodka w wariancji progowej umożliwia porównanie względnej wielkości efektu samorodka różnych zmiennych regionalizowanych,

- wariancja strukturalna (C) - prawdziwy przestrzenny komponent wariancji próby, określa tę część wariancji, która wynika z przestrzennej zależności obserwacji,
- próg ($C_0 + C$) - określa wartość, przy której następuje stabilizacja wariancji; odpowiada on w przybliżeniu wariancji s^2 danych stacjonarnych; w warunkach braku komponenta C_0 , próg jest określony wielkością C .

5. Modele wariogramów

Model wariogramu jest funkcją opisującą rozkład semiwariancji w zależności od h , dla której odchylenia wartości teoretycznych semiwariancji od empirycznych są najmniejsze. Określony dla danego rozkładu semiwariancji model umożliwia jednoznaczne określenie parametrów wariogramu. Może on być liniowy, logarytmiczny, wykładniczy, sferyczny, itd., jednakże wybór odpowiedniego modelu jest dość trudny i wymaga od badacza pewnego doświadczenia. Problem polega na tym, że model jest jedynie przybliżeniem prawdziwego wariogramu; niewłaściwie zdefiniowany może stać się głównym źródłem zniekształcenia wyznaczanych krigingiem wyników. Z szerokiego spektrum możliwych do wyboru modeli wariogramu praktyczne znaczenie w doświadczalnictwie połowym wydają się mieć jedynie dwa modele: liniowy i sferyczny (Trangmar i in., 1985; Perrier i Wilding, 1986; Stroup i in., 1994) (Ryc. 3).

Model liniowy semiwariogramu z progiem ma postać:

$$\gamma(h) = \begin{cases} C_0 + C \left(\frac{h}{a}\right) & \text{dla } 0 < h \leq a \\ C_0 + C & \text{dla } h > a \end{cases}, \quad (15)$$

natomiast model sferyczny odpowiednio:

$$\gamma(h) = \begin{cases} C_0 + C \left(\frac{3}{2} \frac{h}{a} - \frac{1}{2} \frac{h^3}{a^3}\right) & \text{dla } 0 < h \leq a \\ C_0 + C & \text{dla } h > a \end{cases}. \quad (16)$$

6. Kriging

Kriging umożliwia uzyskanie optymalnych, nieobciążonych predyktorów zmiennej regionalizowanej w punktach nie objętych pomiarami, w oparciu o strukturalne właściwości semiwariogramu i pierwotny zestaw danych. Najprostsza forma krigingu obejmuje predykcję wartości w danym punkcie powierzchni (kriging punktowy), względnie wartości dla określonej powierzchni (kriging blokowy), przy założeniu rozkładu normalnego i stacjonarności badanej zmiennej losowej. W przypadku niestacjonarności,

rozkładu innego niż normalny, występowania trendów czy anizotropii predykcja wartości jest utrudniona (Henley, 1981).

W krigingu punktowym wartość zmiennej regionalizowej z w punkcie x_0 jest średnią ważoną wartości obserwowanych w sąsiedztwie tego punktu:

$$\hat{z}(x_0) = \sum_{i=1}^n w_i z(x_i), \quad (17)$$

gdzie n oznacza liczbę punktów wykorzystanych w predykcji, natomiast w_i są wagami związanymi z obserwowanymi wartościami zmiennej $z(x_i)$. Wyznaczony predyktor jest nieobciążony

$$E[\hat{z}(x_0) - z(x_0)] = 0, \quad (18)$$

gdy suma wag jest równa 1

$$\sum_{i=1}^n w_i = 1. \quad (19)$$

Wtedy wyznaczona wariancja błędu predykcji σ_k^2 przyjmuje wartość najmniejszą

$$\sigma_k^2 = \text{var}[\hat{z}(x_0) - z(x_0)] = \min. \quad (20)$$

Kombinację wag związanych z danym punktem można uzyskać z relacji

$$\sum_{j=1}^n w_j \gamma(x_i, x_j) + \lambda = \gamma(x_i, x_0), \quad \text{gdzie } j = 1, 2, \dots, n. \quad (21)$$

Wartości $\gamma(x_i, x_j)$ i $\gamma(x_i, x_0)$ są wartościami semiwariogramu między obserwacjami w punktach x_i a x_j oraz między obserwacją w punkcie x_i a predykowaną wartością w punkcie x_0 , natomiast λ jest mnożnikiem Lagrange'a związanym z minimalizacją σ_k^2 . Rozwiązanie $n + 1$ równań dla każdego w_i i λ umożliwia oszacowanie $\hat{z}(x_0)$ – wzór (17), a wariancja może być wyznaczona ze wzoru

$$\sigma_k^2 = \sum_{j=1}^n w_j \gamma(x_i, x_0) + \lambda. \quad (22)$$

Kriging punktowy jest szczególnym przypadkiem krigingu blokowego, gdzie wyznaczana jest wartość dla pewnej powierzchni; krigowana wartość zmiennej regionalizowanej z dla dowolnego bloku (panelu) A jest średnią ważoną wartości x_i obserwowanych w sąsiedztwie tego bloku.

$$\hat{z}(A) = \sum_{i=1}^n w_i z(x_i). \quad (23)$$

Minimalna wariancja wyznaczana dla bloku A , postaci

$$\sigma_A^2 = 2w_i\gamma(x_i, A) - w_iw_j\gamma(x_i, x_j) - \gamma(A, A), \quad (24)$$

wymaga zdefiniowania trzech kategorii semiwariogramów:

$\gamma(x_i, A)$ – średni semiwariogram między każdym punktem x_i sąsiadującym z panelem A a każdym punktem w panelu A ,

$\gamma(x_i, x_j)$ – średni semiwariogram między każdą parą punktów x_i, x_j z zestawu n punktów wykorzystanych w predykcji wartości panelu A ,

$\gamma(A, A)$ – średni semiwariogram między wszystkimi punktami w panelu A .

Średnie semiwariogramy wyznacza się za pomocą tzw. funkcji pomocniczych, których standaryzowane wartości dla pospolitych modeli wariogramu o paramerach $a = 1$ i $C + C_0 = 1$ można odczytać z tabel (np. Clark, 1979; Perrier i Wilding, 1986) lub wykresów (Journal i Huijbregts, 1978), lub też wyznaczyć za pomocą rachunku różniczkowego (np. Webster i Oliver, 1990). Ponadto, wzory umożliwiające wyznaczenie funkcji pomocniczych dla różnych modeli semiwariogramu można znaleźć w pracy Guibala (1973), natomiast w odniesieniu do najczęściej wykorzystywanych w badaniach przyrodniczych modeli liniowego i sferycznego także w pracy Webstera i Burgessa (1984).

Rozwiązanie układu równań (21) umożliwia wyznaczenie wag dla każdej z obserwacji

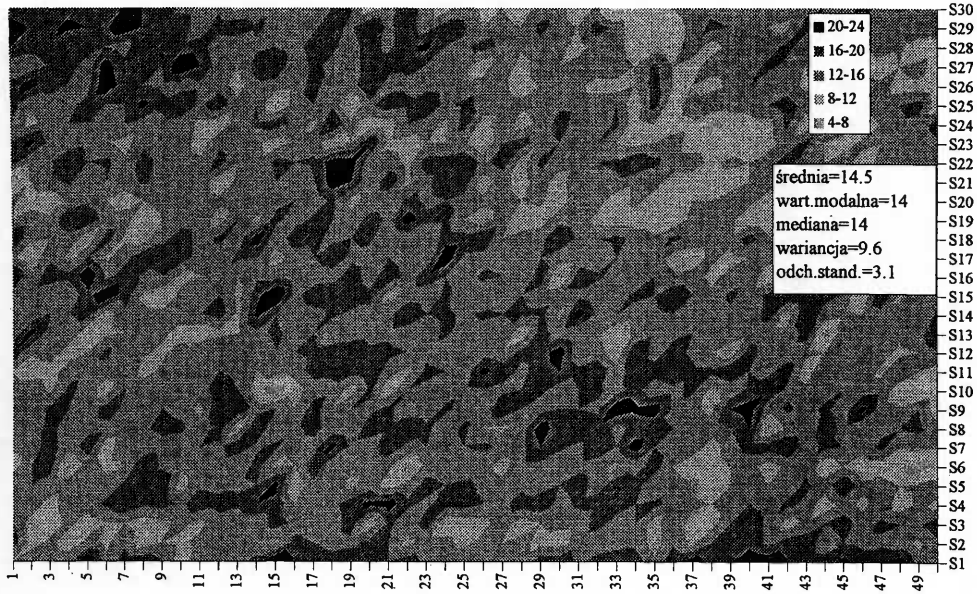
$$\begin{aligned} w_1\gamma(x_1, x_1) + w_2\gamma(x_1, x_2) + \dots + w_n\gamma(x_1, x_n) + \lambda &= \gamma(x_1, A) \\ w_1\gamma(x_2, x_1) + w_2\gamma(x_2, x_2) + \dots + w_n\gamma(x_2, x_n) + \lambda &= \gamma(x_2, A) \\ \vdots & \\ w_1\gamma(x_n, x_1) + w_2\gamma(x_n, x_2) + \dots + w_n\gamma(x_n, x_n) + \lambda &= \gamma(x_n, A) \\ w_1 &+ w_2 &+ \dots + w_n &+ 0 = 1 \end{aligned}, \quad (25)$$

a następnie średniej i wariancji:

$$\begin{aligned} \hat{z}_A &= w_1x_1 + w_2x_2 + \dots + w_nx_n \\ \hat{s}_A^2 &= w_1\gamma(x_1, A) + w_2\gamma(x_2, A) + \dots + w_n\gamma(x_n, A) + \lambda - \gamma(A, A) \end{aligned} \quad (26)$$

7. Przykład liczbowy

Podstawą dalszych rozważań będą wyniki doświadczenia bezczynnikowego (50 m × 30 m) z łubinem żółtym o jednostkach podstawowych wielkości 1 m² (1 m × 1 m). Rozkład plonów z jednostek podstawowych przedstawiono na Ryc. 4. Relacja między statystykami położenia wskazuje na rozkład normalny badanej cechy, jednakże wzdłuż przekątnej obserwuje się minimalną tendencję kierunkowej zmiany plonu. Początkowy fragment macierzy danych z 48 jednostek podstawowych w układzie 8 × 6 (Ryc. 5)



Ryc. 4. Rozkład zmienności plonów lubinu żółtego z jednostek podstawowych w dag/m^2 (dane empiryczne)

posłuży jako podstawa ilustrująca technikę wyznaczania semiwariogramów i krigingu pierwszej wartości.

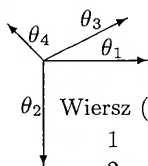
Średnie semiwariogramy dla analizowanego fragmentu danych, wyznaczone dla izotropii rozkładu danych, zestawiono w Tabeli 1.

Tab. 1. Średnie semiwariogramy izotropowe

lag h	1	1	2	2
Odległość w m	1	$\sqrt{2}$	2	$2\sqrt{2}$
$N(h)$	82	68	70	48
$\gamma(h)$	9.4	10.7	11.8	14.8

Semiwariogramy kierunkowe oraz wyznaczony na ich podstawie model semiwariogramu dla kompletnego zestawu danych z doświadczenia prezentuje Ryc. 6.

Wyznaczone zgodnie z modelem sferycznym [wzór (12)] parametry semiwariogramu wynoszą: $a = 9$ m, $C_0 = 7.2$ (dag/m^2)², $C = 2.1$ (dag/m^2)². Można stwierdzić, że zakres skorelowania obserwacji sięga 9 m, natomiast udział wariancji samorodka, a więc wariancji losowej, stanowi 75% zmienności ogólnej. Gdy $h \geq 9$ wtedy semiwariogram ustala się, przyjmując wartość $\gamma(h) = 7.2 + 2.1 = 9.3$ (dag/m^2)².

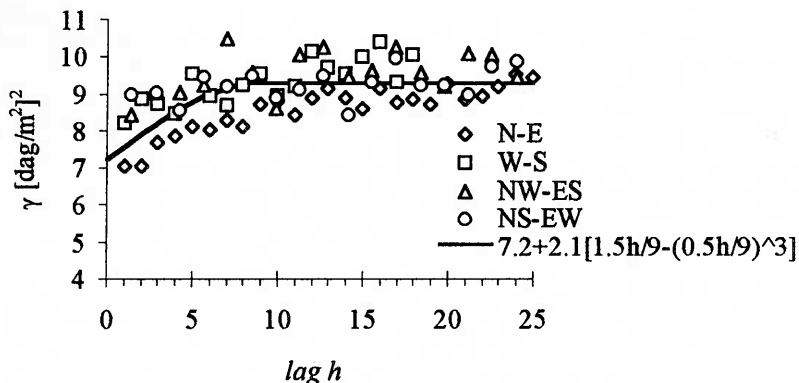


Wiersz (i)	Kolumna (j)						$\theta_1, (\text{lag } h = 1),$ $[z(i, j) - z(i, j + 1)]^2$					$\theta_1, (\text{lag } h = 2),$ $[z(i, j) - z(i, j + 2)]^2$			
	1	2	3	4	5	6									
1	12	18	16	12	15	10	36	4	16	9	25	16	36	1	4
2	15	10	10	15	16	16	25	0	25	1	0	25	25	36	1
3	15	13	9	10	15	17	4	16	1	25	4	36	9	36	49
4	14	10	13	15	14	7	16	9	4	1	49	1	25	1	64
5	23	13	10	14	16	14	100	9	16	4	4	169	1	36	0
6	13	11	12	17	13	11	4	1	25	16	4	1	36	1	36
7	18	15	12	20	17	12	9	9	64	9	25	36	25	25	64
8	19	9	10	17	19	20	100	1	49	4	1	81	64	81	9
							$N(1) = 40, \gamma(1) = 9.0$					$N(2) = 32, \gamma(2) = 16.1$			
$\theta_2, (\text{lag } h = 1),$ $[z(i, j) - z(i + 1, j)]^2$							$\theta_3, (\text{lag } h = 1),$ $[z(i, j) - z(i + 1, j + 1)]^2$					$\theta_4, (\text{lag } h = 1),$ $[z(i, j + 1) - z(i + 1, j)]^2$			
9 64 36 9 1 36							4 64 1 16 1					9 36 4 0 36			
0 9 1 25 1 1							4 1 0 0 1					25 9 36 36 1			
1 9 16 25 1 100							25 0 36 16 64					1 1 9 0 9			
81 9 9 1 4 49							1 0 1 1 0					169 0 25 0 81			
100 4 4 9 9 9							144 1 49 1 25					0 1 4 1 1			
25 16 0 9 16 1							4 1 64 0 1					49 9 25 49 36			
1 36 4 9 4 64							81 25 25 1 9					16 9 100 0 49			
$N(1) = 42, \gamma(1) = 9.7$							$N(1) = 35, \gamma(1) = 9.5$					$N(1) = 35, \gamma(1) = 11.9$			
$\theta_2, (\text{lag } h = 2),$ $[z(i, j) - z(i + 2, j)]^2$							$\theta_3, (\text{lag } h = 2),$ $[z(i, j) - z(i + 2, j + 2)]^2$					$\theta_4, (\text{lag } h = 2),$ $[z(i, j + 2) - z(i + 2, j)]^2$			
9 25 49 4 0 49							9 64 1 25					1 1 36 0			
1 0 9 0 4 81							4 25 16 64					16 25 9 1			
64 0 1 16 1 9							25 1 49 16					196 9 25 9			
1 1 1 4 1 16							4 49 0 16					0 16 4 100			
25 4 4 36 1 4							121 49 49 4					64 1 16 36			
36 4 4 0 36 81							9 36 49 9					49 64 9 36			
$N(2) = 36, \gamma(2) = 8.1$							$N(2) = 24, \gamma(2) = 14.5$					$N(2) = 24, \gamma(2) = 15.1$			

Ryc. 5. Schemat wyznaczenia semiwariogramów kierunkowych dla pierwszych dwu odległości (lag $h = 1$ i lag $h = 2$)

Liczba sąsiednich obserwacji wymagana do wiarygodnego krigingu wartości średniej i wariancji nie powinna przekraczać 25 przy regularnej siatce pomiarów (Burgess i Webster, 1980) oraz 40 dla danych rozmieszczonych nieregularnie (Vieira i in., 1983).

W prezentowanym przykładzie jako podstawę predykcji przyjęto $n = 12$ sąsiednich obserwacji w układzie 4×3 (Ryc. 5). Do wyznaczenia średnich semiwariogramów zastosowano uproszczone podejście, polegające na wykorzystaniu tzw. dwukierunkowych funkcji pomocniczych $H(d, b)$ i $F(d, b)$ dla izotropowego modelu sferycznego (Tab. 2).



Ryc. 6. Semiwariogramy kierunkowe oraz model sferyczny semiwariogramu

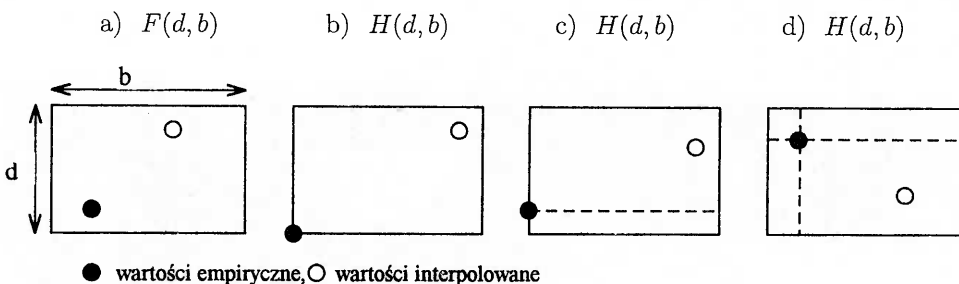
Tab. 2. Funkcje pomocnicze $H(d, b)$ i $F(d, b)$ dla izotropowego modelu sferycznego, przy założeniu, że $a = 1$ i $C = 1$ (fragment, Clark 1981)

d/a	$H(d, b)$					$F(d, b)$				
	b/a					b/a				
	0.1	0.2	0.3	0.4	0.5	0.1	0.2	0.3	0.4	0.5
0.1	0.114	0.177	0.243	0.310	0.374	0.078	0.120	0.165	0.211	0.256
0.2	0.177	0.227	0.285	0.346	0.406	0.120	0.155	0.196	0.237	0.280
0.3	0.243	0.285	0.336	0.390	0.445	0.165	0.196	0.231	0.270	0.309
0.4	0.310	0.346	0.390	0.439	0.489	0.211	0.237	0.270	0.305	0.342
0.5	0.374	0.406	0.445	0.489	0.535	0.256	0.280	0.309	0.342	0.376

W takim przypadku odległości d (wzdłuż kolumn) i b (wzdłuż wierszy) podano jako wartości względne w stosunku do zakresu a , tzn. $H(d, b)$ z zakresem a odpowiada $H(d/a, b/a)$ z zakresem 1. Istotę funkcji pomocniczych $H(d, b)$ i $F(d, b)$ przedstawiono na Ryc. 7.

Pierwsza funkcja (7a) służy do wyznaczenia średniego semiwariogramu $\gamma(A, A)$, natomiast trzy pozostałe (7b-d) odpowiednio – $\gamma(x_i, A)$. Prześledźmy zatem kolejne etapy obliczeń prowadzące do estymacji wag związanych z każdą obserwacją (Tab.3).

Oznaczmy kolejne obserwacje numerami od 1 do 12, a następnie określmy ich współrzędne (w metrach), przypisując pierwszej obserwacji współrzędne (0, 0), ostatniej zaś (3, 2). Dla porządku, w kolumnie 4 Tabeli 3 przedstawiono wartości plonów odpowiadające kolejnym obserwacjom. W zależności od przypisanej pozycji inny jest sposób szacowania średnich semiwariogramów $\gamma(x_i, A)$. Pierwsza obserwacja leży na rogu prostokąta, stąd odpowiada funkcji pomocniczej ilustrowanej na Ryc. 7b. Sze-



Ryc. 7. Dwukierunkowe funkcje pomocnicze

Tab. 3. Wyznaczenie średnich semiwariogramów $\gamma(x_i, A)$

Nr	d	b	$z(i)$	d/a	b/a	$H(d, b)$	d/a	b/a	$H(d, b)$	$\gamma(x_i, A)$
1	0	0	12	0.33	0.22	0.3162				7.86
2	0	1	18	0.33	0.11	0.2698				7.77
3	0	2	16	0.33	0.22	0.3162				7.86
4	1	0	15	0.11	0.22	0.1970	0.22	0.22	0.2524	7.69
5	1	1	10	0.11	0.11	0.1278	0.22	0.11	0.1970	7.57
6	1	2	10	0.11	0.22	0.1970	0.22	0.22	0.2524	7.69
7	2	0	15	0.11	0.22	0.1970	0.22	0.22	0.2524	7.69
8	2	1	13	0.11	0.11	0.1278	0.22	0.11	0.1970	7.57
9	2	2	9	0.11	0.22	0.1970	0.22	0.22	0.2524	7.69
10	3	0	14	0.33	0.22	0.3162				7.86
11	3	1	10	0.33	0.11	0.2698				7.77
12	3	2	13	0.33	0.22	0.3162				7.86

rokość bloku A (w tym przypadku prostokąta) wynosi $b = 2$ m, natomiast długość $d = 3$ m. Z Tabeli 2 odczytujemy, że

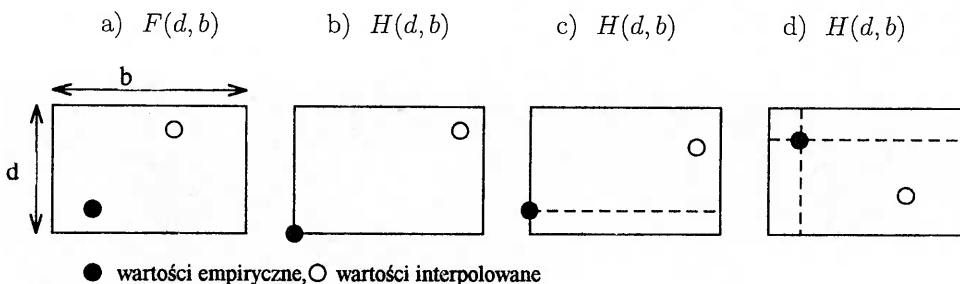
$$H(d, b) = H(3/9, 2/9) = H(0.33, 0.22) = 0.3162;$$

po podstawieniu wartości oszacowań parametrów semiwariogramu z modelu sferycznego otrzymujemy

$$H(d, b) = H(d, b)C + C_0 = (0.3162)(2.1) + 7.2 = 7.86.$$

Stąd semiwariogram między pierwszą wartością a wszystkimi wartościami bloku A wynosi:

$$\gamma(x_1, A) = \frac{d \cdot b \cdot H(d, b)}{d \cdot b} = \frac{3 \cdot 2 \cdot 7.86}{3 \cdot 2} = 7.86.$$



Ryc. 7. Dwukierunkowe funkcje pomocnicze

Tab. 3. Wyznaczenie średnich semiwariogramów $\gamma(x_i, A)$

Nr	d	b	$z(i)$	d/a	b/a	$H(d, b)$	d/a	b/a	$H(d, b)$	$\gamma(x_i, A)$
1	0	0	12	0.33	0.22	0.3162				7.86
2	0	1	18	0.33	0.11	0.2698				7.77
3	0	2	16	0.33	0.22	0.3162				7.86
4	1	0	15	0.11	0.22	0.1970	0.22	0.22	0.2524	7.69
5	1	1	10	0.11	0.11	0.1278	0.22	0.11	0.1970	7.57
6	1	2	10	0.11	0.22	0.1970	0.22	0.22	0.2524	7.69
7	2	0	15	0.11	0.22	0.1970	0.22	0.22	0.2524	7.69
8	2	1	13	0.11	0.11	0.1278	0.22	0.11	0.1970	7.57
9	2	2	9	0.11	0.22	0.1970	0.22	0.22	0.2524	7.69
10	3	0	14	0.33	0.22	0.3162				7.86
11	3	1	10	0.33	0.11	0.2698				7.77
12	3	2	13	0.33	0.22	0.3162				7.86

rokość bloku A (w tym przypadku prostokąta) wynosi $b = 2$ m, natomiast długość $d = 3$ m. Z Tabeli 2 odczytujemy, że

$$H(d, b) = H(3/9, 2/9) = H(0.33, 0.22) = 0.3162;$$

po podstawieniu wartości oszacowań parametrów semiwariogramu z modelu sferycznego otrzymujemy

$$H(d, b) = H(d, b)C + C_0 = (0.3162)(2.1) + 7.2 = 7.86.$$

Stąd semiwariogram między pierwszą wartością a wszystkimi wartościami bloku A wynosi:

$$\gamma(x_1, A) = \frac{d \cdot b \cdot H(d, b)}{d \cdot b} = \frac{3 \cdot 2 \cdot 7.86}{3 \cdot 2} = 7.86.$$

Identyczne wartości semiwarogramu będą miały również obserwacje z numerami 4,10 i 12. Czwarta obserwacja o współrzędnych (1, 0) jest położona w drugim wierszu i pierwszej kolumnie danych pierwotnych, a więc na krawędzi prostokąta. Funkcja pomocnicza $H(d, b)$ odpowiadająca wariantowi przedstawionemu na Ryc. 7c, jest średnią dwu paneli o wymiarach 1×2 m i 2×2 m. W związku z tym mamy dwie funkcje pomocnicze:

$$H(d_1, b) = H(1/9, 2/9) = H(0.11, 0.22) = 0.1970$$

oraz

$$H(d_2, b) = H(2/9, 2/9) = H(0.22, 0.22) = 0.2524;$$

po podstawieniu parametrów semiwariogramu z modelu sferycznego otrzymujemy

$$H(d_1, b) = (0.1970)(2.1) + 7.2 = 7.61$$

oraz

$$H(d_2, b) = (0.2524)(2.1) + 7.2 = 0.2524.$$

Stąd semiwariogram między czwartą wartością a wszystkimi wartościami bloku A

$$\gamma(x_1, A) = \frac{d_1 \cdot b \cdot H(d_1, b) + d_2 \cdot b \cdot H(d_2, b)}{d_1 \cdot b + d_2 \cdot b} = \frac{1 \cdot 2 \cdot 7.61 + 2 \cdot 2 \cdot 7.73}{1 \cdot 2 + 2 \cdot 2} = 7.69.$$

Podobny sposób liczenia zastosujemy w przypadku obserwacji o numerach 2,6,7,9 i 11.

Ostatni wariant funkcji pomocniczej $H(d, b)$ (Ryc. 7d) dotyczy sytuacji gdy obserwacja znajduje się wewnątrz bloku A. W naszym przypadku dotyczy to punktów o numerach 5 i 8. Teraz w bloku mamy cztery panele: dwa o wymiarach 1×1 m i dwa o wymiarach 2×2 m. Daje to cztery funkcje pomocnicze, jednak sposób ich wyznaczenia, a następnie semiwariogramów jest analogiczny jak w powyższym przykładzie. Z uwagi na symetryczność paneli w bloku A, w Tabeli 3 przedstawiono tylko dwie funkcje $H(0.11, 0.22)$ i $H(0.22, 0.22)$. Analiza wyznaczonych semiwariogramów $\gamma(x_i, A)$ pokazuje, że ich wielkość zależy od odległości od centrum bloku A; im bliżej centrum położona jest dana obserwacja, tym mniejsza jest wartość semiwariogramu.

Wyznaczenie semiwariogramów między poszczególnymi obserwacjami $\gamma(x_i, x_j)$ wymaga jedynie podstawienia odpowiednich odległości między nimi (h) do wyznaczonego modelu sferycznego semiwariogramu. Przykładowo semiwariogram między punktami o numerach 1 i 2, oddalonych o $h = 1$ m, będzie wynosił

$$\gamma(x_1, x_2) = 7.2 + 2.1 \left[\frac{3}{2} \frac{1}{9} - \frac{1}{2} \left(\frac{1}{9} \right)^3 \right] = 7.55,$$

natomiast między punktami o numerach 1 i 9, odległych o $h = 2\sqrt{2}$, odpowiednio

$$\gamma(x_1, x_0) = 7.2 + 2.1 \left[\frac{3}{2} \frac{2\sqrt{2}}{9} - \frac{1}{2} \left(\frac{2\sqrt{2}}{9} \right)^3 \right] = 8.18.$$

Wyznaczone semiwariogramy $\gamma(x_i, x_j)$ i $\gamma(x_i, A)$ oraz dodatkowy komponent λ tworzą układ $n + 1 = 13$ równań:

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	λ	$\gamma(x_i, A)$
x_1		$7.6w_2+7.9w_3+7.6w_4+7.7w_5+8.0w_6+7.9w_7+8.0w_8+8.2w_9+8.2w_{10}+8.3w_{11}+8.5w_{12}+1$												$= 7.9$
x_2	$7.6w_1+$		$7.6w_3+7.7w_4+7.6w_5+7.7w_6+8.0w_7+7.9w_8+8.0w_9+8.3w_{10}+8.2w_{11}+8.3w_{12}+1$											$= 7.8$
x_3	$7.9w_1+7.6w_2+$			$7.6w_4+7.7w_5+7.6w_6+8.2w_7+8.0w_8+7.9w_9+8.5w_{10}+8.2w_{11}+8.2w_{12}+1$										$= 7.9$
x_4	$7.6w_1+7.7w_2+7.6w_3+$				$7.6w_5+7.9w_6+7.6w_7+7.7w_8+8.0w_9+7.9w_{10}+8.0w_{11}+8.2w_{12}+1$									$= 7.7$
x_5	$7.7w_1+7.6w_2+7.7w_3+7.6w_4+$					$7.6w_6+7.7w_7+7.6w_8+7.7w_9+8.0w_{10}+7.9w_{11}+8.0w_{12}+1$								$= 7.6$
x_6	$8.0w_1+7.7w_2+7.6w_3+7.9w_4+7.6w_5+$						$7.6w_7+7.7w_8+7.6w_9+8.2w_{10}+8.0w_{11}+7.9w_{12}+1$							$= 7.7$
x_7	$7.9w_1+8.0w_2+8.2w_3+7.6w_4+7.7w_5+7.6w_6+$							$7.6w_8+7.9w_9+7.6w_{10}+7.7w_{11}+8.0w_{12}+1$						$= 7.7$
x_8	$8.0w_1+7.9w_2+8.0w_3+7.7w_4+7.6w_5+7.7w_6+7.6w_7+$								$7.6w_9+7.7w_{10}+7.6w_{11}+7.7w_{12}+1$					$= 7.6$
x_9	$8.2w_1+8.0w_2+7.9w_3+8.0w_4+7.7w_5+7.6w_6+7.9w_7+7.6w_8+$									$7.6w_{10}+7.7w_{11}+7.6w_{12}+1$				$= 7.7$
x_{10}	$8.2w_1+8.3w_2+8.5w_3+7.9w_4+8.0w_5+8.2w_6+7.6w_7+7.7w_8+7.6w_9+$										$7.6w_{11}+7.9w_{12}+1$			$= 7.9$
x_{11}	$8.3w_1+8.2w_2+8.2w_3+8.0w_4+7.9w_5+8.0w_6+7.7w_7+7.6w_8+7.7w_9+7.6w_{10}+$											$7.6w_{12}+1$		$= 7.8$
x_{12}	$8.5w_1+8.3w_2+8.2w_3+8.2w_4+8.0w_5+7.9w_6+8.0w_7+7.7w_8+7.6w_9+7.9w_{10}+7.6w_{11}+$												1	$= 7.9$
	w_1+	w_2+	w_3+	w_4+	w_5+	w_6+	w_7+	w_8+	w_9+	$w_{10}+$	$w_{11}+$	w_{12}		$= 1$

Rozwiązanie układu równań prowadzi do następujących oszacowań wag w_i :

$$w_1 = 0.0786, w_2 = 0.0821, w_3 = 0.0786, w_4 = 0.0849, w_5 = 0.0908, w_6 = 0.0849, \\ w_7 = 0.0849, w_8 = 0.0908, w_9 = 0.0849, w_{10} = 0.0786, w_{11} = 0.0821, w_{12} = 0.0786.$$

oraz $\lambda = 0.547$.

Stąd wartość pierwszego interpolowanego punktu (wzór 22), zogniskowanego w centrum bloku A, wynosi

$$\hat{z}_A = 12 \cdot 0.0786 + 18 \cdot 0.0821 + 16 \cdot 0.0786 + 15 \cdot 0.0849 + \dots + 13 \cdot 0.0786 = 12.9.$$

Do wyznaczenia wariancji błędu predykcji potrzebne jest ustalenie wielkości jednostek (bloków), dla których dokonujemy szacowania wartości. Determinuje to także końcową siatkę obserwacji po kringingu. W naszym przykładzie założyliśmy, że predykcja na podstawie 12 sąsiednich obserwacji będzie dotyczyła centralnego bloku o powierzchni 1 m^2 ($1 \times 1 \text{ m}$). Przy takim założeniu, do wyznaczenia kolejnego średniego semiwariogramu $\gamma(A, A)$, odpowiadającemu wielkości bloku $A - 1 \text{ m}^2$, wykorzystamy funkcję pomocniczą $F(d, b)$ (Ryc. 7a):

$$F(d, b) = F(1/9, 1/9) = H(0.11, 0.11) = 0.0867;$$

po podstawieniu wartości oszacowań parametrów semiwariogramu z modelu sferycznego otrzymujemy

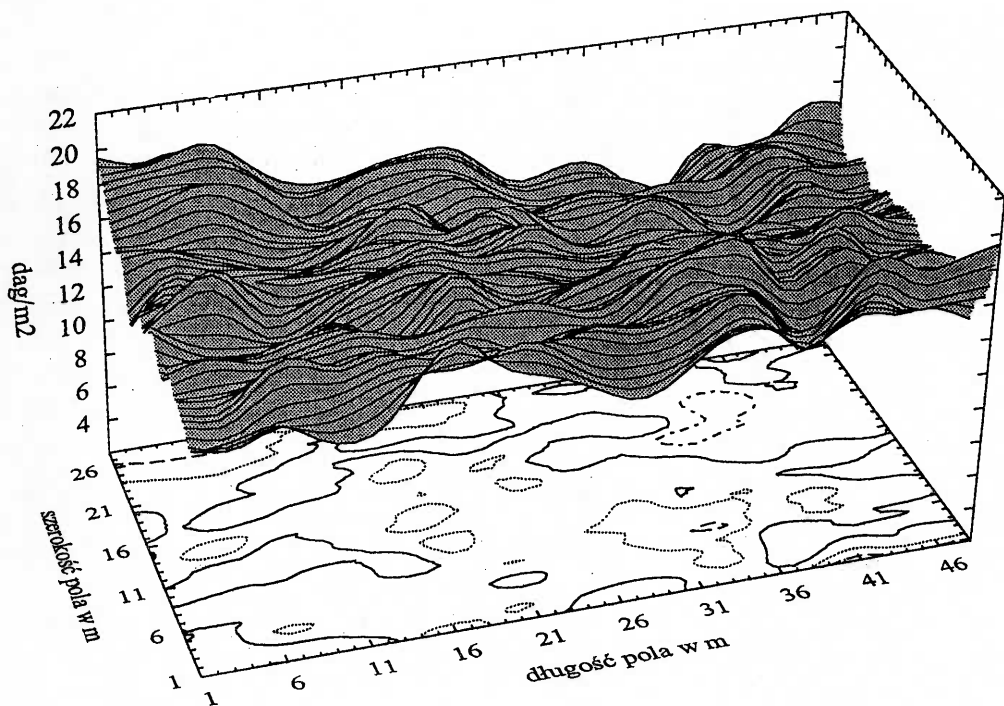
$$\gamma(A, A) = F(d, b) = F(d, b)C + C_0 = (0.0867)(2.1) + 7.2 = 7.38.$$

Stąd estymatory wariancji i odchylenia standardowego po krigingu:

$$\hat{s}^2 = 7.86 \cdot 0.0786 + 7.77 \cdot 0.0821 + \dots + 7.86 \cdot 0.0786 + 0.547 - 7.38 = 0.90$$

$$\hat{s} = 0.95$$

Wyznaczona wariancja oraz odchylenie standardowe przyjmują niższe wartości aniżeli analogiczne wyznaczone z danych empirycznych (Ryc. 4). Kriging wykonany dla całego pola prowadzi do nowego zestawu danych w układzie 47×28 , a ich graficzna prezentacja (Ryc. 8) potwierdza tezę o ciągłości zmiennej regionalizowanej. Warto zaznaczyć, że średni plon ze wszystkich jednostek podstawowych i po kringingu pozostaje niezmienny, natomiast odchylenie standardowe po kringingu zmniejszyło się trzykrotnie.



Ryc. 8. Plony nasion lubinu żółtego – dane po kringingu

8. Podsumowanie

Jednym z elementów udanego eksperymentu polowego jest właściwe jego zaplanowanie i przeprowadzenie, a następnie zastosowanie odpowiedniego aparatu statystycznego do analizy uzyskanych wyników. Metodyczne założenia, związane z określeniem wielkości poletka, jego kształtu, pojemności bloku, liczby bloków oraz układu doświadczalnego, powinny uwzględniać faktyczną zmienność glebową pola doświadczalnego. Jeżeli zmienność glebowa ma charakter losowy, a sąsiednie obserwacje nie wykazują przestrzennego skorelowania, wtedy klasyczne układy doświadczalne (zwykle blokowe) oraz metody statystycznego opracowania danych (np. analiza wariancji i regresji) zapewniają wiarygodne porównania międzyobiektywne. Z drugiej strony zaś, wysoka zmienność glebowa może "maskować" prawdziwe efekty obiektywne, szczególnie wtedy, gdy są one niewielkie. Wówczas tradycyjne techniki analityczne mogą okazać się niewystarczające.

Powszechny w praktycznym doświadczalnictwie układ losowanych bloków traci swoją skuteczność w warunkach dużej zmienności glebowej wewnątrzblokowej. Jeśli ponadto liczba analizowanych obiektów jest duża, to nawet w warunkach umiarkowanej zmienności glebowej, trzeba liczyć się z dalszym zmniejszeniem dokładności doświadczenia, rozumianej jako efekt większego błędu doświadczalnego. Skuteczniejsze w takich przypadkach mogą okazać się układy o blokach niekompletnych lub też – bardziej elastyczne w stosowaniu, lecz o potencjalnie mniejszej efektywności – układy α -rozkładalne (Patterson i in., 1978). Dalszych korzyści związanych z redukcją błędu doświadczalnego można oczekiwać po zastosowaniu bardziej zaawansowanych metod, uwzględniających przestrzenne usytuowanie jednostek eksperymentalnych i korektę zmiennej (np. plonu). Podstawą korekty może być: (i) pozycja na polu danego obiektu doświadczalnego względem obiektów sąsiadujących (metoda najbliższego sąsiada, Papadakis, 1937), (ii) ocena kierunkowych zmian żyzności gleby (metoda analizy trendu, np. Bowman, 1990), (iii) wykorzystanie zmiennej towarzyszącej (metoda kowariancji) oraz (iv) wykorzystanie parametrów semiwariogramu.

Celem wielu ostatnio opublikowanych prac była ocena możliwości zwiększenia efektywności doświadczeń polowych poprzez estymację efektów zmienności przestrzennej (np. Bhatti i in., 1991; Brownie i in., 1993; van Es i van Es, 1993). Jednakże, jak dotychczas, wykorzystanie metod geostatystycznych do oceny wyników polowych doświadczeń czynnikowych z powtórzeniami jest na etapie analiz porównawczych z innymi metodami. Brownie i in. (1993) porównali efektywność klasycznych modeli analizy wariancji: układu losowanych bloków i układu łańciskowego z modelami uwzględniającymi trend, wzajemne sąsiedztwo oraz skorelowanie błędów doświadczalnych. Okazało się, że najefektywniejszym wariantem był model ujmujący równocześnie obecność trendu i przestrzenne skorelowanie błędów doświadczalnych. Z kolei w bada-

niach Bhatti'ego i in. (1991) wykrycie subtelnych różnic między obiektami nawozowymi umożliwiły metody korygujące dane pierwotne na podstawie modelu ujmującego przestrzenne skorelowanie obserwacji.

Wydaje się, że przyszłe wykorzystanie metod geostatystycznych, dzięki możliwości kwantyfikacji stopnia skorelowania między sąsiednimi jednostkami eksperymentalnymi (semiwariogram) oraz modelowania struktury zmienności przestrzennej (kriging), będzie zmierzało w kierunku diagnostyki danych pod kątem zasadności stosowania statystyki klasycznej. Obiektywne metody wspomagające wybór najlepszej analizy danych nie zostały jeszcze opracowane, jednak można przypuszczać, że dalsze badania w kierunku doskonalenia metod analizy danych z doświadczeń polowych uwzględniające efekty skorelowania przestrzennego przyczynią się do lepszego rozumienia metod statystyki przestrzennej oraz pełniejszego określenia ich zalet i wad.

Inną płaszczyzną wykorzystania metod statystyki przestrzennej w doświadczalnictwie rolniczym mogą być badania z zakresu nauk o żyzności gleby i nawożeniu, gdzie podstawą są doświadczenia wieloletnie. Jak prognozują Fotyma i Filipiak (1996), badania tego typu powinny być prowadzone na wielkopowierzchniowych polach o znanych właściwościach glebowych, regularnie monitorowanych co do stanu ładu i plonowania, a jedną z technik analitycznych powinny stanowić metody geostatystyczne.

Reasumując rozważania związane ze stosowaniem metod statystyki przestrzennej w doświadczalnictwie polowym należy podkreślić także znaczenie tych metod w procesie optymalizowania metodyki ścisłego doświadczenia polowego. Parametry semiwariogramu wyznaczone dla danego pola doświadczalnego wespół z oszacowaniami wskaźnika zmienności glebowej Smitha mogą stanowić podstawę racjonalnie zaprojektowanego doświadczenia, uwzględniającego m.in.: optymalne jednostki eksperymentalne, ich kształt, pasy ochronne (brzeżne i czołowe) stanowiące bufor dla oddziaływań sąsiedzkich, pojemność bloku oraz wybór odpowiedniego układu doświadczalnego.

LITERATURA

- Bhatti A.U., Mulla D.J., Koecher F.E., Gurmani A.H. (1991). Identifying and removing spatial correlation from yield experiments. *Soil Sci. Soc. Am. J.* **55**, 1523-1528.
- Binns M.R. (1982). The choice of plot size in randomized block experiments. *J. Am. Soc. Hort. Sci.* **107**, 17-19.
- Bowman D.T. (1990). Trend analysis to improve efficiency of agronomic trials in flue cured tobacco. *Agron. J.* **82**, 499-501.
- Brownie C., Bowman D.T., Burton J.W. (1993). Estimating spatial variation in analysis of data from field trials: a comparison of methods. *Agron. J.* **85**, 1244-1253.
- Burges T.M., Webster R. (1980). Optimal interpolation and isarithmic mapping of soil properties. I. The semivariogram and punctual kriging. *J. Soil Sci.* **31**, 315-331.
- Clark I. (1979). *Practical Geostatistics*. Applied Science Publ., London.

- Fotyma M., Filipiak K. (1996). Eksperyment polowy jako element rozwiązywania zagadnień badawczych w zakresie nauki o żyzności gleby i nawożeniu. *Fragmenta Agronomica* **3**(51), 85-101.
- Gołaszewski J., Eaton G.W., Baumann T.E. (1995). Optimum plot size in field experiments with strawberries. *J. of Small Fruit & Viticult.* **3**(1), 39-48.
- Gołaszewski J. (1996). Optymalizacja metodyki eksperymentu polowego z roślinami strączkowymi w aspekcie zmienności przestrzennej pola doświadczalnego. *Acta Acad. Agricult. Tech. Olst., Supl.C. Agricultura* **62**, 1-92.
- Guibal D. (1973). Les fonctions auxiliaires a deux dimensions pour le schéma sphérique. Report N-347 Centre de Géostatistique et de Morphologie Mathématique, Fontainebleau.
- Henley S. (1981). Nonparametric geostatistics. *Appl. Sci. Publ.*, London.
- Journel A.G., Huijbregts C.J. (1978). *Mining Geostatistics*. Academic Press, London.
- Matheron G. (1962-1963). *Traité de Géostatistique Appliquée*. Vol.1 and 2, Technip, Paris.
- Matheron G. (1971). The theory of regionalized variables and its applications. *Cahiers du Centre de Morphologie Mathématique*, Fontainebleau, **5**.
- Koch E.J., Rigney J.A. (1951). A method of estimating optimum plot size from experimental data. *Agron. J.*, **43**, 17-21.
- Krajewski P., Molińska A., Moliński K. (1994). Analiza eksploratywna zależności przestrzennych dla danych pochodzących z monitorowania powietrza. *XXIV Colloquium Metodologiczne z Agrobiometrii*, PAN, 327-340.
- Krajewski P., Molińska A., Moliński K. (1995). Modelowanie struktury dyspersji danych przestrzennych z wykorzystaniem przekształcenia liniowego. *XXV Colloquium Biometryczne z Agrobiometrii*, PAN, 164-173.
- Krige D.G. (1966). Two-dimensional weighted moving average trend surfaces for ore-evaluation. *J. South African Institute of Mining and Metallurgy* **66**, 13-38.
- Kristensen K., Ersboll A.K. (1992). The use of geostatistical methods in planing variety trials. *Biul. Oceny Odm.* **24-25**, 139-157.
- Papadakis J.S. (1937). Methode statistique pour des experiences sur champs. *Bull. Inst. Amel. Plantes a Salonique* **23**.
- Patterson H.D., Williams E.R., E.A.Hunter. (1978). Block design for variety trials. *J. Agric. Sci.* **90**, 395-400.
- Perrier E.R., Wilding L.P. (1986). An evaluation of computational methods for field uniformity studies. *Advances in Agronomy* **39**, 265-312.
- Rejman S., Szczepański K. (1978). Wielkość i kształt poletka w polowych doświadczeniach z truskawkami. *Pr. Inst. Sad. i Kwiac.* **20**, 264-286.
- Smith H.F. (1938). An empirical law describing heterogeneity in agricultural crops. *J. Agric. Sci.* **3**, 1-23.
- Stroup W.W., Baenzinger R.S., Miltze D.K. (1994). Removing spatial variation from wheat yield trials: a comparison of methods. *Crop Sci.* **86**, 62-66.
- Trangmar B.B., Yost R.S., Uehara G. (1985). Application of geostatistics to spatial studies of soil properties. *Advances in Agronomy* **38**, 45-94.

- Trętowski J. (1975). *Wielkość jednostek doświadczalnych w różnych układach polowych eksperymentów ziemniaczanych*. PWRiL, Poznań.
- Usowicz B., Baranowski P., Kossowski J. (1995). Spatial distribution of some physical quantities characterizing soil structure state in cultivated fields. *Polish Journal of Soil Science* XXVIII(1), 19-27.
- van Es H.M., van Es C.L. (1993). Spatial nature of randomization and its effect on the outcome of field experiments. *Agron. J.* 85, 420-428.
- Vieira S.R., Hatfield J.L., Nielsen D.R., Biggar J.W. (1983). Geostatistical theory and application to variety of some agronomical properties. *Hilgardia* 51(3), 1-75.
- Webster R., Burgess T.M. (1984). Sampling and bulking strategies for estimating soil properties in small regions. *Journal of Soil Science* 35, 127-140.
- Webster R., Oliver M.A. (1990). *Statistical methods in soil and land resource surveys*. Oxford University Press.

*Praca wpłynęła 5 marca 1997;
w wersji poprawionej 14 listopada 1997*

The application of geostatistical methods in field experimentation

SUMMARY

Field experiments carried out in the conditions of distinct soil heterogeneity are usually biased with large experimental error what can lead to poor estimation of the treatment effects. Rational experimentation connected with planning, laying out and carrying of field experiments as well as the proper method of statistical approach depends largely on the amount of information about spatial variability of the experimental field. The geostatistical approach is relatively new among the methods of spatial variability analysis. It enables to assess the extent of random and structural variance. The latter is attributed to spatial correlation of the observations.

In the paper the theoretical considerations and algorithm of algebraic calculations for the two pivotal methods of spatial statistics: semivariogram and kriging have been presented. Possible ways of application of the methods in field experimentation are pointed out. It is stated that estimation of the field heterogeneity with geostatistical methods can facilitate the optimisation of the methodical aspects of field experiments, can serve as a diagnostic tool in choosing the proper statistical method of data evaluation and may be helpful in analysing long-term experiments concerning research problems with soil fertility and fertilization.

KEY WORDS: field experimentation, geostatistics, spatial correlation, semivariogram, kriging.