

# Wykorzystanie metod interpolacji do numerycznego kreślenia map powierzchni geologicznych na podstawie nieregularnie rozmieszczonych danych

Paweł Goldsztejn\*, Grzegorz Skrzypek\*\*

Applications of numeric interpolation methods to the geological surface mapping using irregularly spaced data. *Prz. Geol.*, 52: 233–236.

*Summary.* In this paper we report a test of usage of interpolation methods available in Surfer 7.0 for interpolating irregularly distributed data. The experiment involved data of the top surface of glacial tills in the center of Wrocław. This surface is represented by 136 not evenly distributed data. Some methods are not useful for working with data on geological structures, as they produce maps that are highly unclear or unrealistic. Two parameters were used for describing usefulness of selected gridding methods: a sum of residuals to a power ( $\Sigma\delta^2$ ) and an overall look of a map. Firstly, five methods were tested on specially limited testing area with relatively dense data. Analysis of above mentioned parameters the afore mentioned parameters led us to choose the three following methods as producing clear maps with small  $\Sigma\delta^2$ : minimum curvature, modified Shepard's method and kriging. Those methods, together with their special, previously optimized, parameters were applied for mapping the whole investigated area. The results showed that both minimum curvature and kriging produce maps of similar appearance and comparable  $\Sigma\delta^2$ . According to Davis (1986) and the software user's guide kriging is good for small data sets (like our 136 points) while minimum curvature is useful rather for bigger data sets with more than 1000 points.

**Key words:** gridding, kriging, irregularly spaced data set

Numeryczne kreślenie map z zastosowaniem siatki interpolacyjnej umożliwia przeprowadzenie wymiernej oceny poprawności działania poszczególnych metod interpolacji dla różnych zjawisk geologicznych. Taka analiza jest szczególnie cenna przy kreśleniu map dla obszarów słabo rozpoznanych o niehomogenicznym rozprzestrzenieniu danych.

Chcąc wykreślić mapę o przynajmniej ogólnym charakterze z zaznaczeniem nadrzędnego trendu zmienności, oczekujemy zwykle odpowiedzi na pytanie o stopień wiarygodności takiej mapy. W poniższej pracy przeprowadzamy dyskusję oceny stopnia wiarygodności mapy na podstawie eksperymentalnych porównań efektów różnych typów interpolacji.

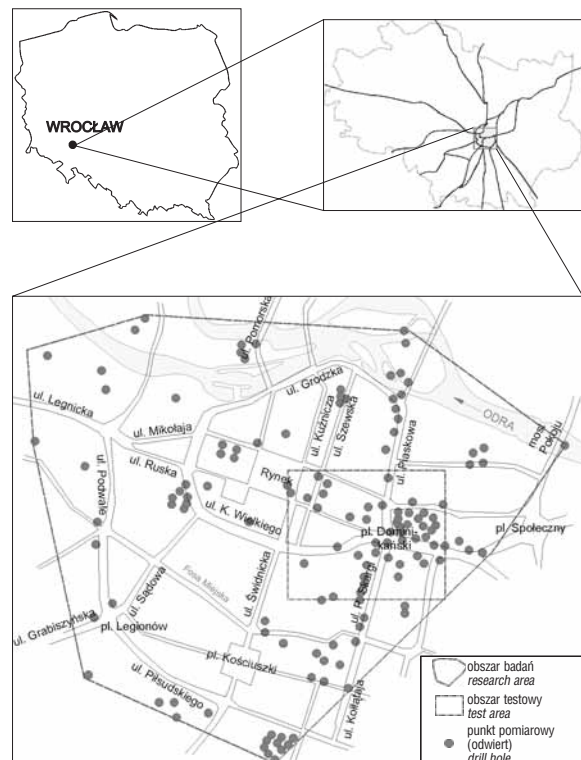
## Interpolacja danych

Program Surfer 7.0 umożliwia przeprowadzenie interpolacji kilkoma metodami: *kriging*, *minimum curvature* (metoda najmniejszej krzywizny), *nearest neighbor* (metoda najbliższego sąsiedztwa), *natural neighbor* (metoda naturalnego sąsiedztwa), *modified Shepard's method* (zmodyfikowana metoda Sheparda), *radial basis function* (metoda radialnych funkcji bazowych), *polynomial regression* (metoda regresji wielomianowej), *inverse distance to a power* (metoda odwrotnych odległości do potęgi), *triangulation with linear interpolation* (metoda triangulacji z liniową interpolacją). Jednak tylko część z nich teoretycznie nadaje się na potrzeby interpolacji w oparciu o nieregularnie rozmieszczone dane. Są to następujące metody: *kriging*, *minimum curvature* i *polynomial regression*.

W procesie interpolacji dochodzi do stworzenia siatki interpolacyjnej, która stanowi matrycę wykorzystywaną przez program Surfer 7.0 do kreślenia izolinii. Proces interpolacji umożliwia znaczne zagęszczenie siatki pomimo nieregularnie rozmieszczonych punktów pomiarowych, co pozwala na dokładne odzwierciedlenie analizowanej powierzchni. Metody interpolacji wykorzystują szereg zróżnicowanych algorytmów do ustalenia wielkości inter-

polowanych parametrów w punktach węzłowych. Zwykle dane pomiarowe, leżące najbliżej punktu węzłowego mają największy wpływ na ten punkt. Wielkość tego wpływu możemy zmieniać używając parametru tzw. wagi. Każda metoda wykorzystuje inny algorytm do wyliczania wartości w punktach węzłowych, a co się z tym wiąże, tworzy inny obraz wykreślonej powierzchni.

Dobór metody interpolacji w przypadku nierównomiernie rozmieszczonych danych pomiarowych powinien być podyktowany kilkoma cechami charakteryzującymi ten zbiór danych: stopniem homogeniczności rozprzestrzenienia danych, ilością punktów na jednostkę powierzchni,



**Ryc. 1.** Położenie obszaru badań wraz z lokalizacją punktów pomiarowych (odwiertów)

**Fig. 1.** Location of study area and sampling points (drill holes)

\*Instytut Górnictwa, Politechnika Wroclawska, pl. Teatralny 2, 50-051 Wrocław; pgold@ing.uni.wroc.pl

\*\*Instytut Nauk Geologicznych, Uniwersytet Wroclawski, ul. Cybulskiego 30, 50-205 Wrocław; buki@ing.uni.wroc.pl

wariancją populacji (stopniem zmienności danych) oraz typem powierzchni, którą odwziewiedlają dane.

### Obszar badań

Jako poligon badawczy, sprawdzający przydatność metod interpolacji do wykonania mapy w oparciu o populację danych wiertniczych nierównomiernie rozmieszczonych w przestrzeni, wybrano obszar centrum Wrocławia. Wykonano analizę wyników interpolacji danych dotyczących zalegania stropu glin zwałowych zlodowacenia południowopolskiego (sanu 2). Na obszarze o powierzchni 2,5 km<sup>2</sup>, zagęszczenie punktów jest bardzo nierównomierne, co jest związane głównie z lokalizacją dużych inwestycji budowlanych, wymagających rozpoznania podłoża budowlanego do głębokości kilkunastu metrów — na takiej głębokości zalega strop wyższej wymiennych glin. Ogółem zgromadzonych zostało 136 danych z archiwalnych otworów wiertniczych, co odpowiada gęstości ok. 45 punktów/km<sup>2</sup>. Zagęszczenie danych w siatce o boku 250 m waha się od 1 do 18 punktów na oczko siatki (Goldsztein & Skrzypek, 2004). Położenie obszaru badań wraz z lokalizacją punktów pomiarowych przedstawiono na ryc. 1. Interpretację geologiczną opisywanej powierzchni zamieszczono w odrębnej pracy (Goldsztein & Skrzypek, 2004).

### Dobór metody interpolacji na obszarze testowym

Przy ocenie przydatności metod interpolacji należy się kierować przede wszystkim dwoma czynnikami: wielkością błędów interpolacyjnych i wyglądem mapy (jej czytelnością i realizacyjnością). Błąd interpolacyjny ( $\delta$ ) powstaje w wyniku procesu interpolowania danych na siat-

kę interpolacyjną i wyliczany jest z różnicy pomiędzy wyinterpolowaną wielkością parametru ( $Z_{int}$ ) w punkcie badawczym, a zmierzoną wielkością tego parametru w terenie ( $Z_{dane}$ ). Sposób obliczania błędu prezentuje wzór [1].

$$\delta = Z_{int} - Z_{dane} \quad [1]$$

Wielkość błędów interpolacyjnych wyliczano przy użyciu funkcji *Grid/Residuals* w programie Surfer 7.0. Analizowano sumy kwadratów błędów interpolacyjnych ( $\Sigma\delta^2$  [m<sup>2</sup>]), co pozwoliło na oszacowanie bezwzględnych wartości błędów.

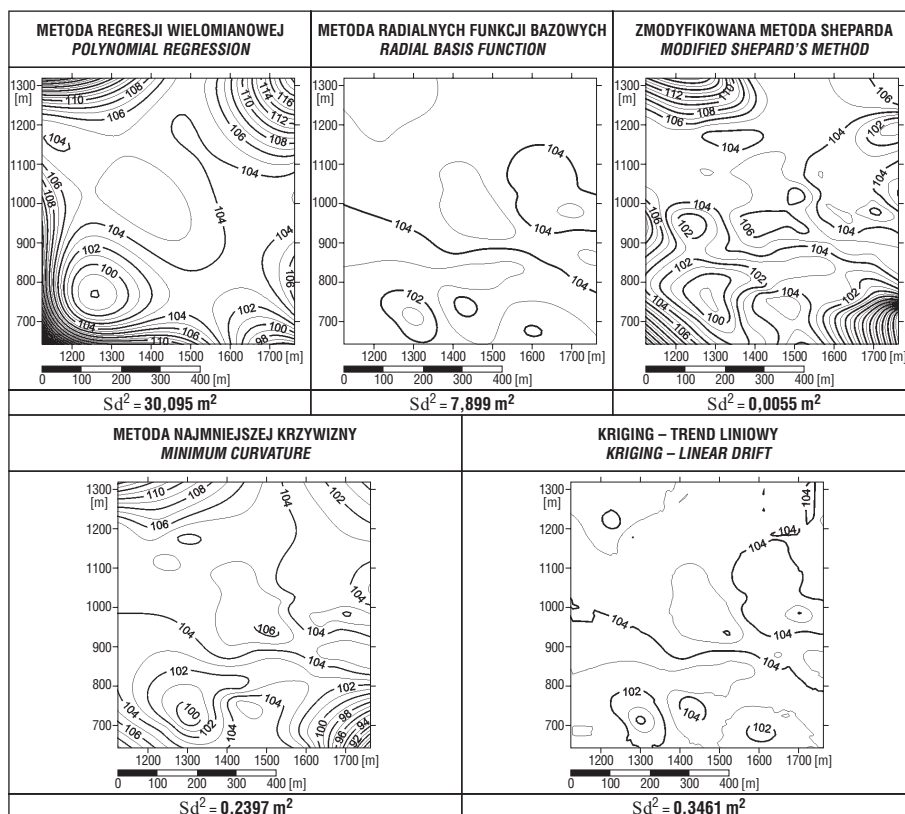
Charakter zmienności i rozprzestrzenienia danych o położeniu badanego stropu glin zwałowych wskazuje na możliwość dużych błędów interpolacyjnych. Wprowadzono testy, o większym zagęszczeniu i mniejszej zmienności danych, aby wstępnie sprawdzić wielkość błędów kreowanych przez wybrane metody interpolacji na stosunkowo dobrze rozpoznany teren.

Dla wymienionych powyżej metod interpolacyjnych sprawdzano wielkość błędów dla różnych gęstości siatki interpolacyjnej. Optymalna gęstość siatki ustalana jest w oparciu o algorytm metody (Instrukcja Surfer 7.0, 1999) oraz najmniejszą odległość między parą punktów. Na obszarze testowym odległość ta wynosi 6,85 m, co daje gęstość siatki interpolacyjnej 94 rzędy na 100 kolumn. Błędy interpolacji były sprawdzane także dla siatek o następującej geometrii (rzędy x kolumny): 73 x 75, 47 x 50 i 23 x 25.

Analizę błędów przeprowadzono w kilku wariantach doboru parametrów algorytmu interpolacji dla kilku wybranych metod. Wstępnie zrezygnowano z analizowania błędów interpolacyjnych map, powstałych przy użyciu następujących metod:

- 1) *natural neighbor* — metoda ta nie ekstrapoluje danych poza granice obszaru występowania danych, co zawęża obszar mapy;
- 2) *nearest neighbor* — metoda ta tworzy w badanym przypadku nieczytelne mapy, dobre rezultaty daje w przypadku bardzo dobrze rozpoznanych obszarów z dużą ilością równomiernie rozłożonych danych;
- 3) *triangulation with linear interpolation* — metoda tworzy w badanym przypadku mapy nieczytelne, złożone z siatki trójkątów, choć pozbawione błędów interpolacji; nadaje się również do dobrze rozpoznanych obszarów z równomiernie zagęszczonymi danymi;
- 4) *inverse distance to a power* — metoda tworzy struktury typu „bull's eye”, nierzeczywiste nagromadzenia izolinii wokół ekstremalnych wartości, tym samym generuje mało prawdopodobne powierzchnie; zwykle nie nadaje się do interpolowania powierzchni geologicznych.

Na obszarze testowym (0,45 km<sup>2</sup> — 55 punktów) przeprowadzono analizę map wykonanych metodami:



Ryc. 2. Mapy stropu glin glacialnych (m n.p.m.) wykonane wybranymi metodami interpolacji na obszarze testowym wraz z odpowiednimi wielkościami błędów interpolacji  $\Sigma\delta^2$  [m<sup>2</sup>]

Fig. 2. Maps of the glacial tills top surface (m a.s.l.) interpolated by selected methods within the test area, with respective values of residuals  $\Sigma\delta^2$  [m<sup>2</sup>]

□ *polynomial regression* — generalizuje ogólny trend danych, znacznie wygładzając powierzchnię nawet przy zastosowaniu wielomianów 4 i 5 stopnia. Suma kwadratów błędów interpolacyjnych ( $\Sigma\delta^2$ ) wyniosła 30,09 m<sup>2</sup> (ryc. 2);

□ *radial basis function* — analizowano wybór funkcji bazowej, tylko dwie z nich (*multilog* i *multiquadric*) pozwoliły na uruchomienie procedury interpolacji. Testowano wpływ zmienności *anisotropy ratio* (współczynnik anizotropii układu) i *anisotropy angle* (kąt anizotropii układu) na  $\Sigma\delta^2$ . Najmniejszą wartość  $\Sigma\delta^2$  (7,90 m<sup>2</sup>) uzyskano dla funkcji bazowej typu *multiquadric* (ryc. 2);

□ *modified Shepard's method* — dla różnych gęstości siatek zastosowano anizotropię układu o parametrach: *range 1* (zasięg południkowy) 1100, *range 2* (zasięg równoleżnikowy)

700 i kąt (*angle*) 15. Testowano wpływ zmian parametrów *quadratic neighbors* (wielkość lokalnego sąsiedztwa, dla którego obliczana jest wielkość parametru Z z danej ilości punktów) i *weighing neighbors* (ilość lokalnych sąsiedztw potrzebnych do wyznaczenia Z w węzłach siatki interpolacyjnej) na  $\Sigma\delta^2$ . Dla obszaru testowego metoda ta okazała się dokładna, a wykreślona powierzchnia ma realistyczny przebieg w obszarze reprezentowanym przez dane. Najmniejsza wartość  $\Sigma\delta^2$  uzyskana tą metodą wynosi 0,01 m<sup>2</sup>, jest to najniższa wartość dla obszaru testowego (ryc. 2);

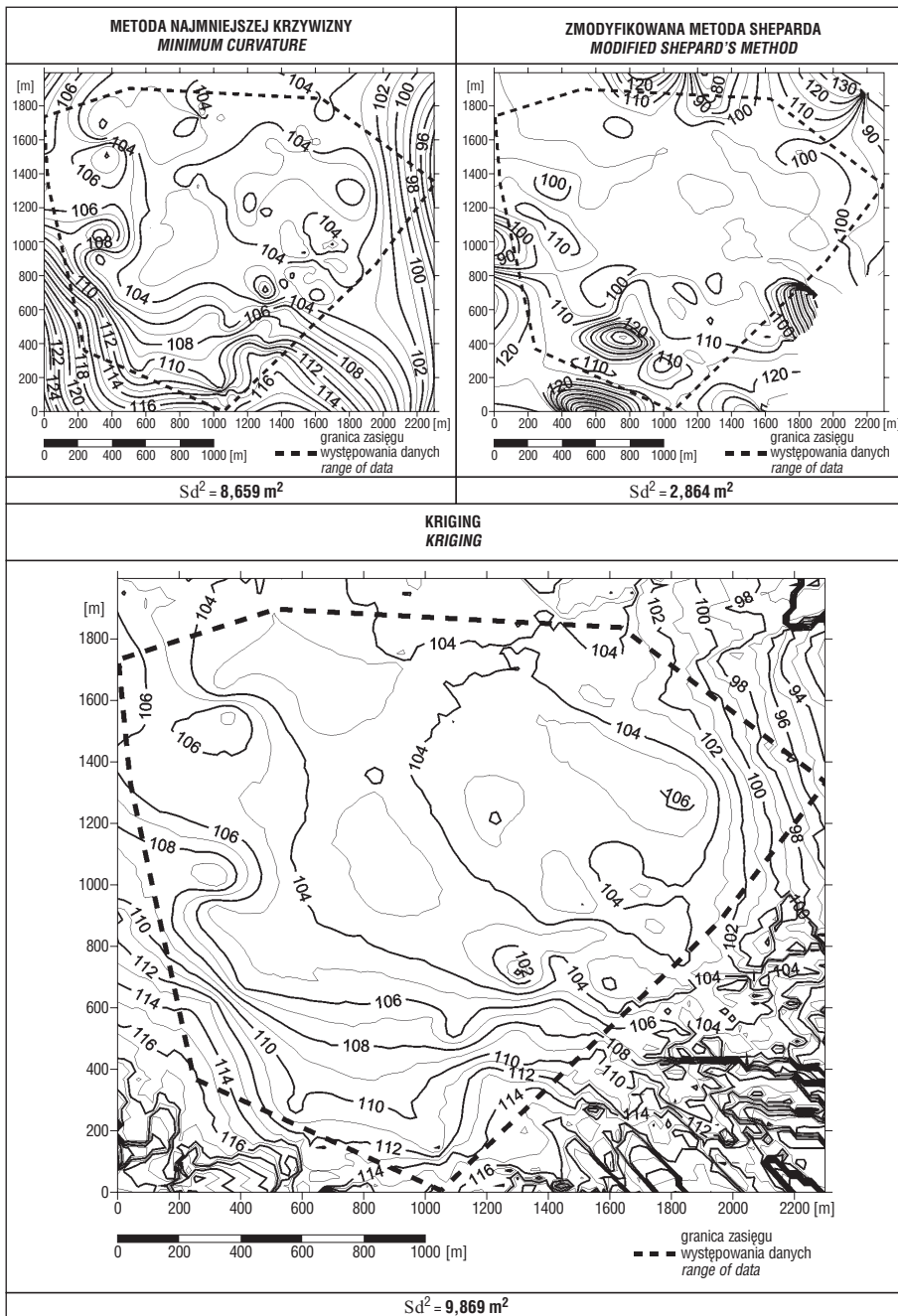
□ *minimum curvature* — analizowano sumy kwadratów błędów dla parametrów iteracji: *maximum residuals* (maksymalny błąd interpolacyjny) 0,0001 i *maximum iteration* (maksymalna liczba iteracji) 25 000. Najmniejsza

wartość  $\Sigma\delta^2$  wynosi 0,24 m<sup>2</sup>. Jest to niski błąd, a powierzchnia, podobnie jak w przypadku metody Sheparda, ma realistyczny przebieg (ryc. 2).

□ *kriging* — uznawany jest przez wielu specjalistów (Davis, 1986) za najlepszą metodę interpolacji niejednorodnych zjawisk przyrodniczych, uwzględnia ogólny trend danych i jest przydatny w sytuacji, gdy dane są rozmieszczone nierównomiernie. Przy analizie  $\Sigma\delta^2$  zwrócono uwagę na wpływ takich parametrów krigingu jak: typ i parametry funkcji odwzorowującej zmienność danych w przestrzeni oraz trend zmienności (*drift*). Testowano także wykorzystanie wariogramu.

**Analiza  $\Sigma\delta^2$  przy domyślnych (ustalonych przez producentów programu) parametrach funkcji:** *linear* (funkcja liniowa), *slope* (tangens kąta nachylenia funkcji) 1, *anisotropy ratio* 1, *angle* 0, bez uwzględnienia trendu zmienności (*no drift*), przy różnych typach wyszukiwania — najmniejsza wartość  $\Sigma\delta^2$  wynosi 0,36 m<sup>2</sup> (dla siatki 94 x 100 i wyszukiwaniu *all data*).

**Analiza  $\Sigma\delta^2$  przy użyciu funkcji wariogramu.** Wariogram analizuje zmienność i przestrzenne zróżnicowanie pomiędzy wszystkimi parami punktów w zbiorze danych. Wykonano wiele wariogramów, jednak żaden z nich nie był w stanie odpowiednio dokładnie przybliżyć zmienności przestrzennej wysokości stropu glin glacialnych za pomocą funkcji matematycznych. Stosowano także funkcję *AutoFit*, która automatycznie dopasowuje wariogram do danego typu zmienności danych. Najmniejsza wartość  $\Sigma\delta^2$  błędu wyniosła 11,3 m<sup>2</sup>. Testowano parametry wariogramu: *lag distance* (maksymalna brana pod uwagę odległość pomiędzy parami



**Ryc. 3.** Mapy stropu glin glacialnych [m n.p.m.] wykonane wyselekcjonowanymi metodami interpolacji na obszarze całego obszaru badań z zaznaczonym zasięgiem występowania danych i wielkością sumy kwadratów błędów interpolacji  $\Sigma\delta^2$  [m<sup>2</sup>]

**Fig. 3.** Maps of the glacial tills top surface [m a.s.l.] interpolated by selected methods within the whole research area, with data extent shown, as well as respective values of residuals  $\Sigma\delta^2$  [m<sup>2</sup>]

punktów) *angular* i *radial divisions* (ilość przedziałów w elipsoidzie wyszukiwania danych) oraz parametry funkcji przy wyszukiwaniu *all* (ryc. 2). Funkcja tworzenia wariogramów jest nowością dla programu Surfer, pojawiła się po raz pierwszy w wersji 7.0 (obecnie już na rynku jest dostępna wersja Surfer 8.0).

**Analiza  $\Sigma\delta^2$  przy zastosowaniu liniowego przybliżenia przestrzennej zmienności danych.** Testowano mapę pod kątem sumarycznego błędu  $\Sigma\delta^2$  przy założeniu równomiernej zmienności danych (*linear drift*) i różnych parametrach funkcji *linear* (*slope*, *anisotropy* i *angle*) i trybie wyszukiwania z uwzględnieniem różnej wagi wszystkich punktów mapy (*all*). Najmniejsze wartości sumy błędów  $\Sigma\delta^2$  uzyskano dla następujących parametrów funkcji *linear*: *slope* 1, *anisotropy* 1,3, *angle* -15° i gęstości siatki 94 x 100. Następnie dla tych parametrów testowano różne typy wyszukiwania. Najmniejszą wartość  $\Sigma\delta^2$  uzyskano ostatecznie dla siatki interpolacyjnej o gęstości 94 x 100, przy wyszukiwaniu danych typu *octant* i elipsoidzie wyszukiwania o parametrach: *radius1* 700, *radius2* 466, *angle* -15° i linearnym przybliżeniu zmienności o następujących parametrach funkcji: *slope* 1, *anisotropy* 1,3, *angle* -15°. W tym przypadku wartość  $\Sigma\delta^2$  wyniosła 0,35 m<sup>2</sup>. Taka mała wartość sumarycznej wielkości kwadratu błędu mapy ( $\Sigma\delta^2$ ), a także czytelność mapy i fakt, że na obszarach pozbawionych danych wartości są ekstrapolowane sprawia, że metodę tę można uznać za najlepiej spełniającą oczekiwania dla analizowanego typu i stopnia rozpoznania powierzchni.

Analiza wartości  $\Sigma\delta^2$  uzyskanych różnymi metodami dowodzi słuszności wyboru następujących metod interpolacji: *kriging*, *modified Shepard's method*, *minimum curvature* jako najbardziej przydatnych do wykreślenia mapy całego obszaru, reprezentującej strop powierzchni geologicznej.

#### Przydatność metod interpolacji do wykonania mapy powierzchni geologicznej z nierównomiernie rozmieszczonych danych

Przeprowadzono analizę wyselekcjonowanych metod interpolacji pod kątem sumarycznej wartości kwadratu błędów dla interpolacji dla całej mapy. Wybrane metody dały następujące efekty:

□ metoda najmniejszej krzywizny (*minimum curvature*) — analiza  $\Sigma\delta^2$  wykazała, że dla parametrów, które sprawdziły się dla obszaru testowego uzyskano  $\Sigma\delta^2$  dla całej mapy wynoszącą: 8,66 m<sup>2</sup>, przy siatce interpolacyjnej 100 x 82 punkty węzłowe. Mapa jest czytelna i realistyczna w obszarze zasięgu danych (ryc. 3);

□ zmodyfikowana metoda Sheparda — dla parametrów, które dały najmniejszą wartość  $\Sigma\delta^2$  na obszarze testowym, na całej mapie uzyskano wartość sumy kwadratów błędów  $\Sigma\delta^2$  wynoszącą 2,86 m<sup>2</sup>, przy siatce interpolacyjnej 100 x 82 punkty węzłowe.  $\Sigma\delta^2$  jest więc bardzo niska, ale na obszarach brzeżnych wykreślona powierzchnia jest nierealistyczna i nie nawiązuje do trendu zmienności danych (ryc. 3);

□ kriging — dla optymalnych parametrów, które uzyskano przy analizie wartości  $\Sigma\delta^2$  dla obszaru testowego, dla całej mapy uzyskano  $\Sigma\delta^2$  wynoszącą 9,87 m<sup>2</sup>. Przy czym największy jednostkowy błąd interpolacyjny ( $\delta$ ) wynosi 0,96 m. Mapa jest czytelna i realistyczna na całej powierzchni reprezentowanej przez dane (ryc. 3). 65% punktów ma  $\delta < 0,2$  m, a 86%  $< 0,4$  m. Tylko 3,6% punktów charakteryzuje się wartością  $\delta > 0,6$  m. Duże wartości  $\delta$  pojawiają się, gdy dane pomiarowe o znacznym różniących się wartościach występują w bliskim sąsiedztwie.

#### Wnioski

Analiza wartości  $\Sigma\delta^2$  uzyskanych za pomocą wyselekcjonowanych metod interpolacyjnych pozwala na uzyskanie następujących wniosków:

□ Metody interpolacji w różny sposób uwzględniają ogólny trend zmienności danych. Błędy interpolacyjne są największe w rejonach o relatywnie dużym zagęszczeniu punktów i względnie dużej zmienności danych w porównaniu do reszty mapy.

□ Pomimo że zmodyfikowana metoda Sheparda, daje dużo mniejsze wartości błędów interpolacji ( $\delta$ ) niż metoda krigingu oraz metoda najmniejszej krzywizny, to mapa uzyskana tą metodą jest mało realistyczna, dochodzi, bowiem do znacznego nagromadzenia wartości ekstremalnych, nie nawiązujących do ogólnego trendu zmienności danych w miejscach o słabym zagęszczeniu punktów badawczych.

□ Optymalnym zagęszczeniem siatki interpolacyjnej dla mapy w analizowanym przykładzie jest 100 x 82 punkty węzłowe, co odpowiada odległości 23–24 m między węzłami siatki w kierunku pionowym i poziomym. Zagęszczanie siatki powoduje znaczne zmniejszenie wartości sumy kwadratów błędów  $\Sigma\delta^2$ , ale zarazem negatywnie wpływa na czytelność mapy i realizm przebiegu interpolowanej powierzchni (izolinie są zbyt nieregularne).

□ Metody krigingu i metoda najmniejszej krzywizny dają w eksperymencie zbliżone wartości  $\Sigma\delta^2$  (odpowiednio: 9,87 i 8,66 m<sup>2</sup>). Różnica jest na tyle nieznaczna, że obydwie metody interpolacji niehomogenicznie rozproszonej populacji danych można uznać za wiarygodne, nie zmieniające w większym stopniu danych wejściowych.

□ Metoda najmniejszej krzywizny daje jednak nieco bardziej zgeneralizowany, przez co optycznie bardziej czytelny obraz wykreślonej powierzchni niż kriging. Cecha ta przejawia się w bardziej wygładzonym przebiegu izolinii. W przypadku krigingu złożoność przebiegu izolinii, można uznać za „nadinterpretację” danych, izolinie są bardziej nieregularne. Wynika to z charakteru algorytmu obliczeniowego, który w przypadku metody najmniejszej krzywizny dąży do wygładzania izolinii. Mankament ten, można usunąć w znacznym stopniu, stosując wtórnie algorytm wygładzania izolinii (*smoothing*), co spowoduje zgeneralizowanie przebiegu izolinii i zarazem ułatwi odbiór treści mapy, przy równocześnie wysokiej jakości odwzorowania rzeczywistego kształtu analizowanej powierzchni. Metoda krigingu została wybrana w trakcie eksperymentu porównawczego, jako metoda dająca najbardziej wiarygodne i realistyczne mapy w procesie interpolacji niehomogenicznej populacji danych odzwierciedlających przebieg powierzchni geologicznej.

□ Wynik eksperymentu potwierdza obecność w literaturze (Davis, 1986) opinie o przydatności krigingu do interpolacji stosunkowo niewielkich (w tym przypadku np. 136 rekordów) populacji danych charakteryzujących fizyczną powierzchnię morfologiczną.

Koszt badań pokryto z grantów: 1017/S/ING/03-IX, 2022/W/ING/03-34.

#### Literatura

- DAVIS J.C. 1986 — Statistics and Data Analysis in Geology. John Wiley & Sons, New York.  
 GOLDSZTEJN P. 2002 — Analiza numeryczna powierzchni stropowej glin glacialnych w centrum Wrocławia. Arch. Wydz. Nauk Przyr., UWrocł.  
 GOLDSZTEJN P. & SKRZYPEK G. 2004 — Morfologia powierzchni stropowej glin zwałowych zlodowacenia południowopolskiego (sanu 2) w centrum Wrocławia. Prz. Geol., 52: 160–162.  
 SURFER 7.0 – user's guide, 1999 — Golden Software Inc.