

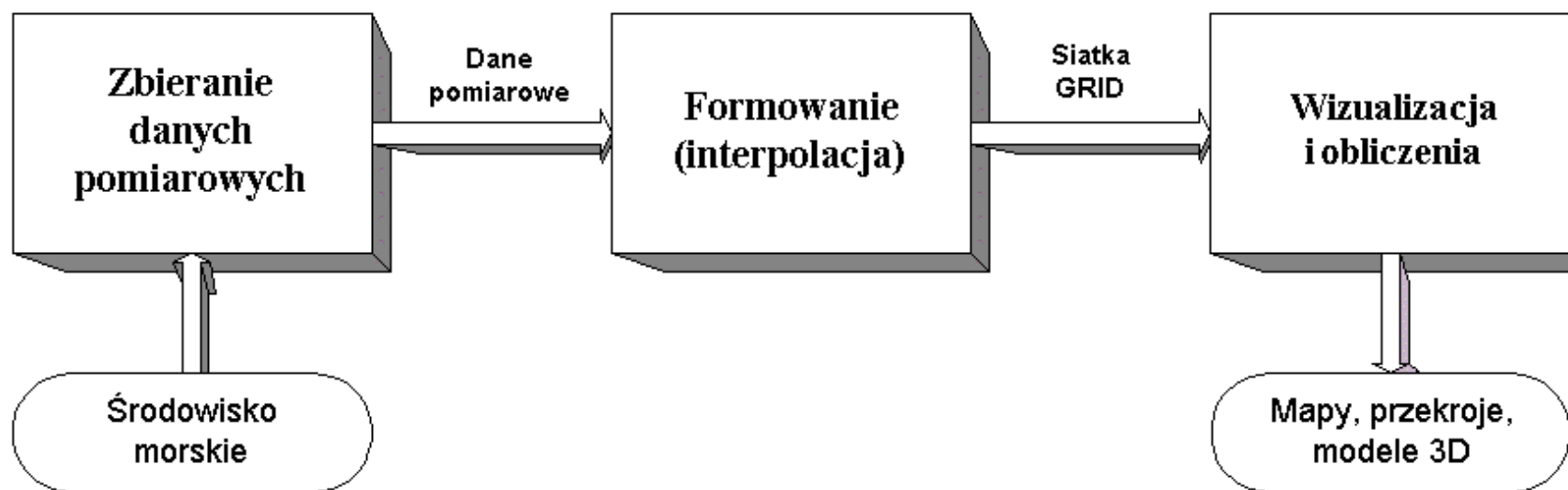
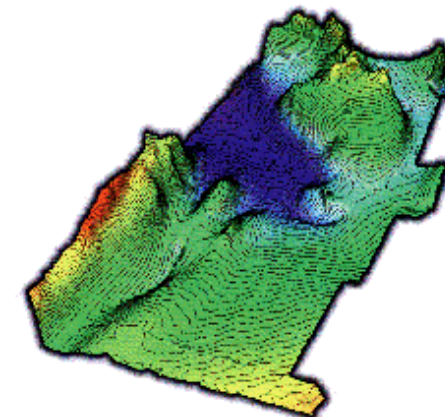
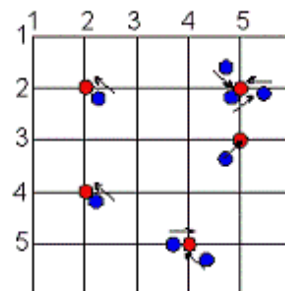
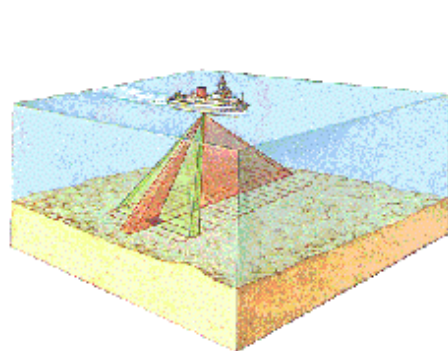
Języki analizy danych

Wykład – Interpolacja danych hydrograficznych



dr hab. inż. Wojciech Maleika
dr inż. Andrzej Łysko
WI ZUT, Szczecin
wmaleika@wi.zut.edu.pl
alysko@zut.edu.pl

Etapy tworzenia modelu dna morza



Format danych batymetrycznych

XYZ ASCII (.txt)

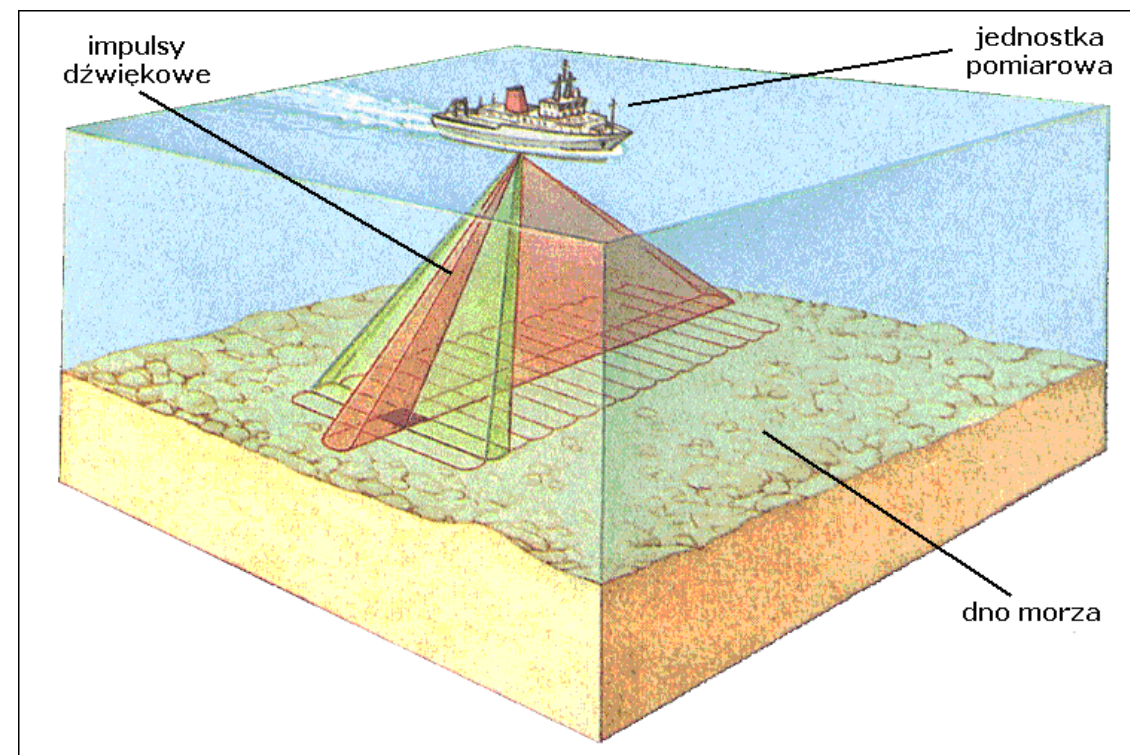
```
14.338404367 53.809770650 5.31
14.338406267 53.809771517 5.19
14.338409267 53.809772767 5.11
14.338410567 53.809773367 4.93
14.338396667 53.809771100 5.67
14.338398467 53.809771950 5.58
14.338399967 53.809772500 5.46
14.338401467 53.809773167 5.33
14.338403867 53.809774250 5.23
14.338405467 53.809774900 5.08
14.338407067 53.809775650 4.91
14.338388367 53.809770717 6.06
14.338389667 53.809771317 5.96
```

XYZ ASCII (.txt)

```
53 57 40.78546 14 14 46.05805 10.690
53 57 40.78240 14 14 45.95437 10.600
53 57 40.78998 14 14 46.56963 10.640
53 57 40.78926 14 14 46.53866 10.680
53 57 40.78817 14 14 46.50770 10.680
53 57 40.78728 14 14 46.47674 10.640
53 57 40.78313 14 14 46.35110 10.640
53 57 40.78133 14 14 46.29638 10.660
53 57 40.77917 14 14 46.22582 10.730
53 57 40.77395 14 14 46.06346 10.690
53 57 40.77034 14 14 45.94970 10.680
53 57 40.76619 14 14 45.81757 10.550
53 57 40.79359 14 14 47.09595 10.670
```

XYZ UTM (.txt)

```
4.5053863e+005 5.9794823e+006 -1.0690000e+001
4.5053674e+005 5.9794822e+006 -1.0600000e+001
4.5054796e+005 5.9794823e+006 -1.0640000e+001
4.5054740e+005 5.9794823e+006 -1.0680000e+001
4.5054683e+005 5.9794823e+006 -1.0680000e+001
4.5054627e+005 5.9794823e+006 -1.0640000e+001
4.5054397e+005 5.9794822e+006 -1.0640000e+001
4.5054298e+005 5.9794821e+006 -1.0660000e+001
4.5054169e+005 5.9794821e+006 -1.0730000e+001
4.5053873e+005 5.9794819e+006 -1.0690000e+001
4.5053666e+005 5.9794818e+006 -1.0680000e+001
4.5053425e+005 5.9794817e+006 -1.0550000e+001
```



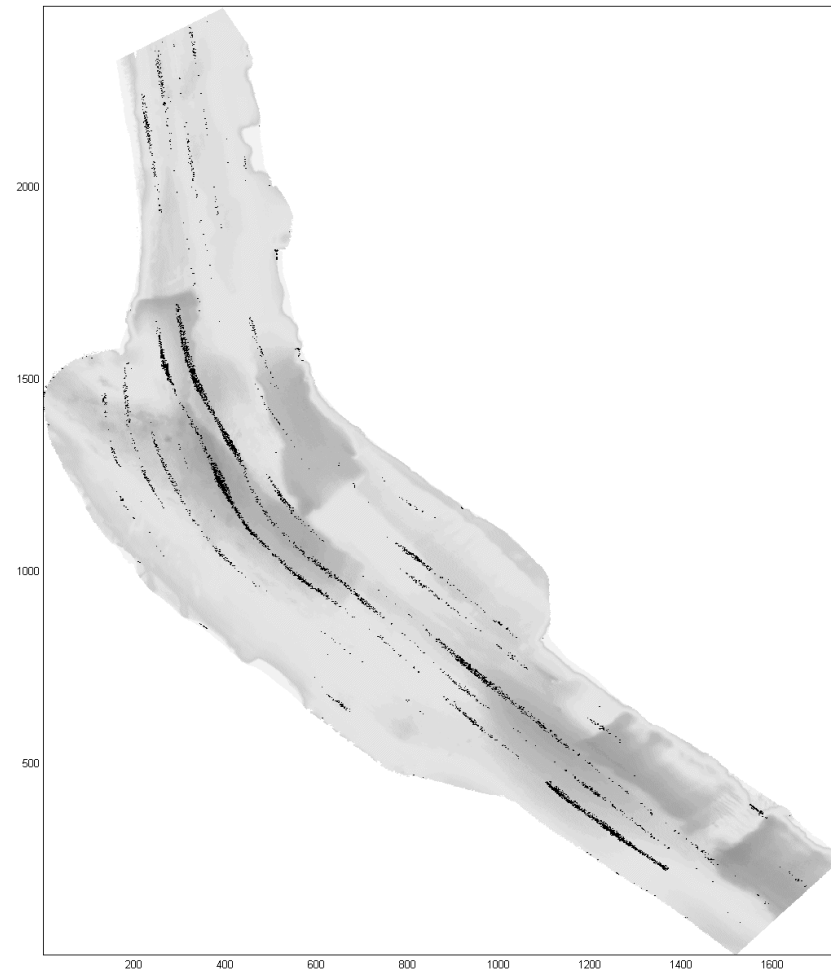
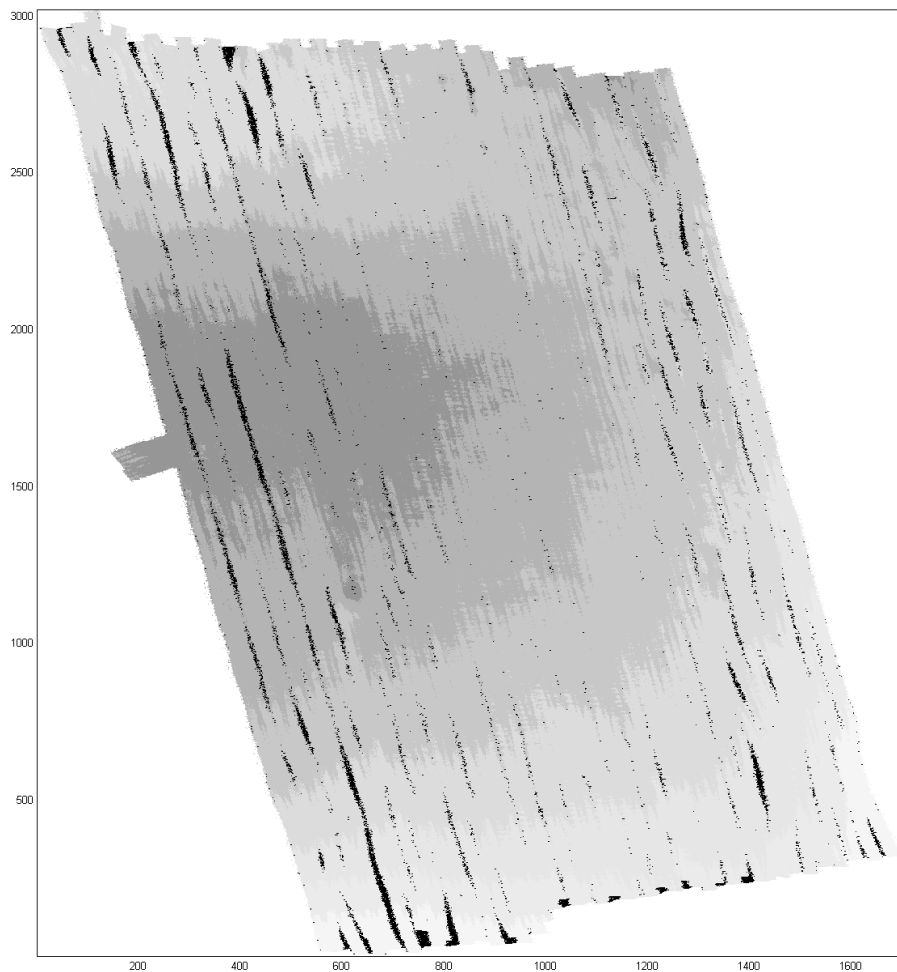
Formaty danych batymetrycznych

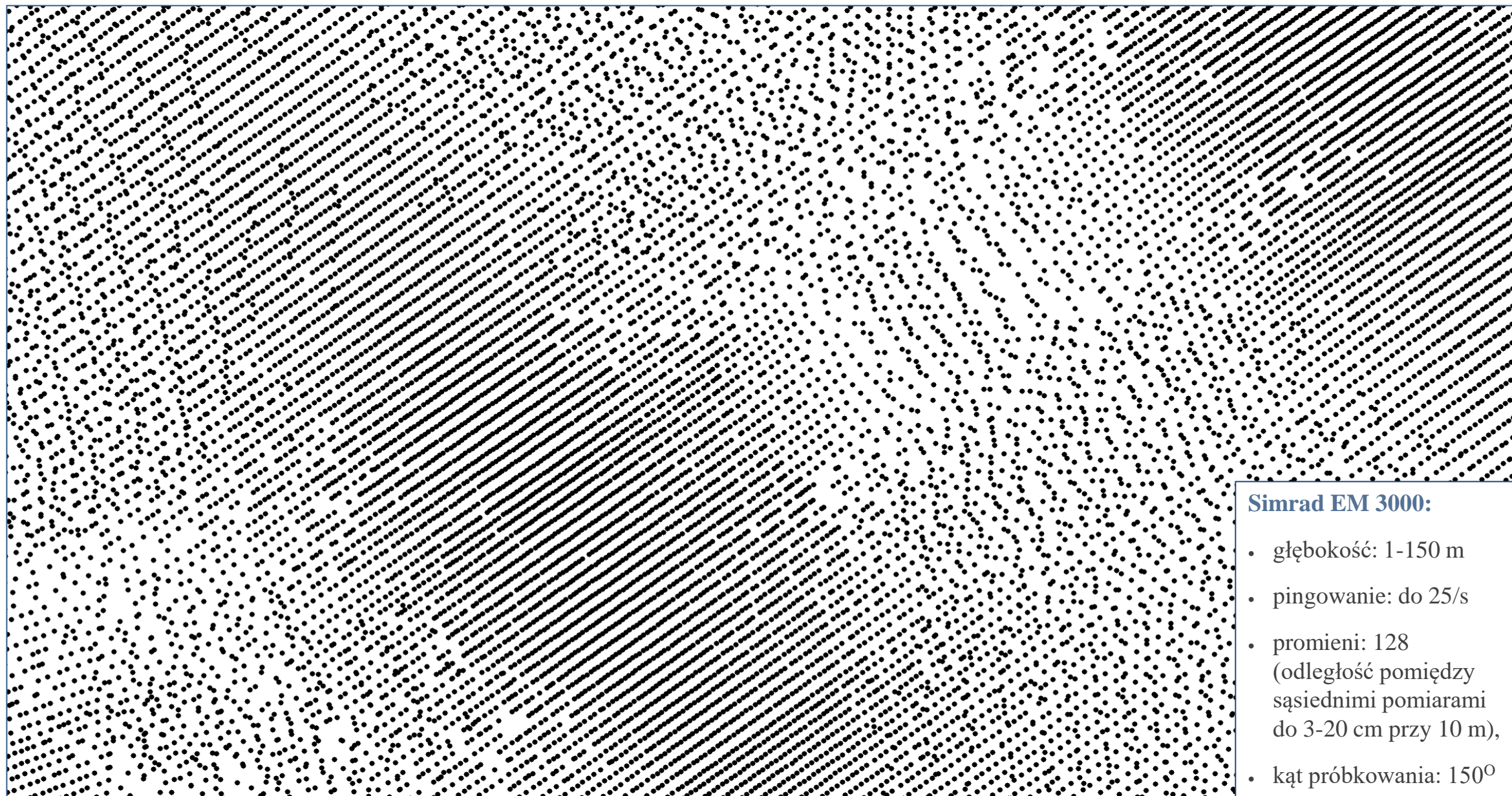
XYZ ASCII (.txt)

```
14.338404367 53.809770650 5.31
14.338406267 53.809771517 5.19
14.338409267 53.809772767 5.11
14.338410567 53.809773367 4.93
14.338396667 53.809771100 5.67
14.338398467 53.809771950 5.58
14.338399967 53.809772500 5.46
14.338401467 53.809773167 5.33
14.338403867 53.809774250 5.23
14.338405467 53.809774900 5.08
14.338407067 53.809775650 4.91
14.338388367 53.809770717 6.06
14.338389667 53.809771317 5.96
```



WGS 84





Simrad EM 3000:

- głębokość: 1-150 m
- pingowanie: do 25/s
- promieni: 128
(odległość pomiędzy sąsiednimi pomiarami do 3-20 cm przy 10 m),
- kąt próbkowania: 150°

Normy w pomiarach hydrograficznych

Tabela 1.2. Minimalne wymagania norm dla pomiarów hydrograficznych wg IHO [IHO98].

Parametr / Klasa	Specjalna	1	2	3
Dokładność horyzontalna (poziom ufności = 95%)	2 m	5 m+5% głębokości	20 m+5% głębokości	150 m + 5% głębokości
Dokładność głębokości dla głębokości skorygowanych (poziom ufności = 95%)	$a=0.25, b=0.0075$	$a=0.5, b=0.013$	$a=1, b=0.023$	$a=1, b=0.023$
100-procentowe przeszukiwanie dna	obowiązkowe	wymagane w wybranych obszarach	może być wymagane w wybranych obszarach	nie dotyczy
zdolność detekcyjna systemu	obiekty o objętości 1 m ³	obiekty o objętości 2 m ³ do 40 m; 10% głębokości powyżej 40 m	jak klasa 1	nie dotyczy
maksymalne odstępny pomiędzy liniami pomiaru	nie dotyczy (100 % przeszukiwanie)	3 średnie głębokości lub 25 m (zależnie od tego, co jest większe)	3-4 średnie głębokości lub 200 m (zależnie od tego, co jest większe)	4 średnie głębokości

Aby obliczyć wielkość błędu dla pomiaru głębokości, odpowiednie wartości a i b podane w tabeli 1.2 należy wprowadzić do wzoru [IHO98]:

$$e = \pm \sqrt{[a^2 + (b \cdot d)^2]} \quad (1.1)$$

gdzie: a – stały błąd głębokości, b – wskaźnik błędu zależnego od głębokości, d – głębokość.

Normy w pomiarach hydrograficznych

Tabela 1.3. Dopuszczalny błąd głębokości dla powierzchni testowych wg norm IHO.

Powierzchnia	Dopuszczalny błąd głębokości wg IHO [m]
Brama torowa	0.258
Kotwiczowisko	0.259
Obrotnica	0.258
Wraki	0.251

Źródła błędów przy wyznaczaniu głębokości

Przy ustalaniu dokładności głębokości modelu (dokładności pionowej) należy ocenić źródła indywidualnych błędów, a następnie je połączyć, aby otrzymać całkowity błąd propagacyjny (TPE) [IHO98]. Składowe błędów obejmują:

- błędy systemu pomiarowego i prędkości dźwięku,
- błędy pomiaru pływów i modelowania,
- błędy przetwarzania danych.

W przypadku stosowania echosondy wielowiązkowej źródłami błędów systemu pomiarowego są:

- błędy niedokładności urządzenia,
- nieprecyzyjna kalibracja urządzenia (niedokładny pomiar prędkości rozchodzenia się dźwięku w wodzie),
- kołysanie boczne i wzdłużne jednostki pomiarowej,
- falowanie,
- geologiczna budowa dna (np. muły, piaski, skały).

Na błędy przetwarzania istotny wpływ mają:

- **zaokrąglenia i przybliżenia w procesie tworzenia grid (interpolacji),**
- przybliżenia w wyznaczaniu izobat w procesie wizualizacji danych.

Interpolacja

Interpolacja:

określanie wartości szacunkowych na podstawie nieciągłych danych pomiarowych

Interpolacja przestrzenna pozwala na podstawie danych punktowych obliczyć poszukiwane wartości na analizowanym obszarze.

Uzyskany wynik charakteryzują **ciągłe**, a nie skokowe zmiany wartości w przestrzeni, bardziej typowe i dostosowane dla opisywanej rzeczywistości.

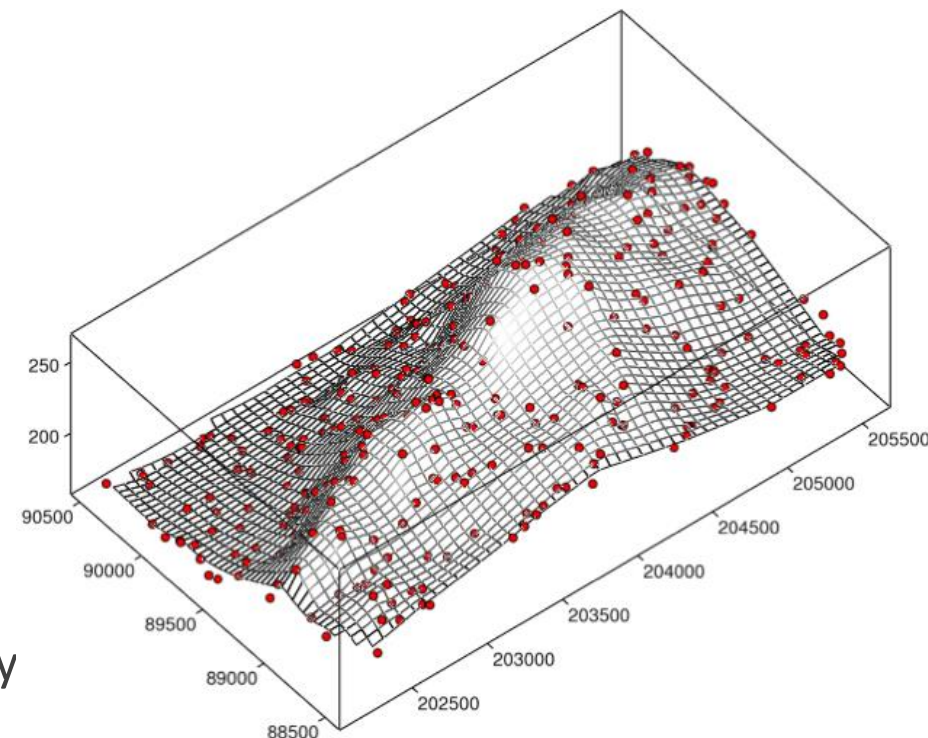
Interpolację stosuje się w przypadku posiadania ograniczonej ilości danych lub danych rozłożonych w przestrzeni w sposób **nieregularny**.

Pozwala na pewną **generalizację danych** np. rozkład zanieczyszczeń, ukształtowanie terenu, przestrzenny rozkład badanego parametru (także tworzymy nie tylko powierzchnie, ale i bryły 3D).

Interpolacja

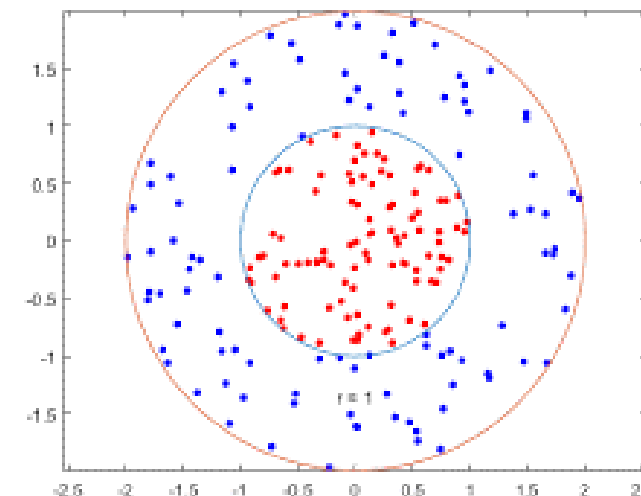
Lista potencjalnych zastosowań:

- Tworzenie regularnych izolinii na mapach.
- Szacowanie wartości w punktach gdzie ich nie ma.
- Zmiana rozdzielczości obrazu.
- Zmiana skali barwnej.
- Zmiana układu odniesienia lub współrzędnych geograficzny
- Obracanie mapy/obrazu.
- Kalibracja np. zdjęcia lotniczego i jego ortorektyfikacja.,
- Zmniejszanie ilości danych poprzez zmniejszanie gęstości danych w miejscach gdzie nie potrzeba dużej dokładności.



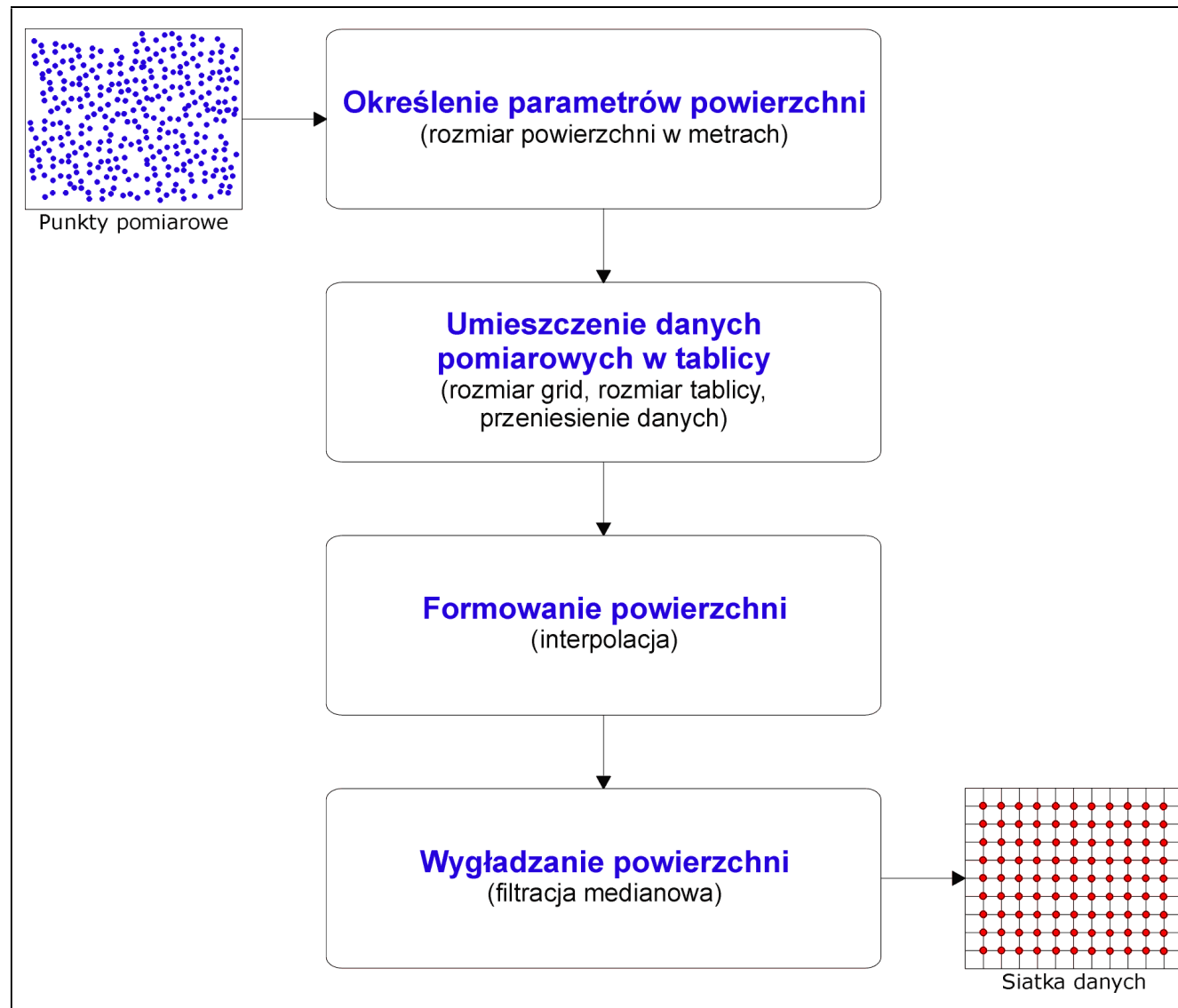
Interpolacja – założenia podstawowe

- Typy danych
 - Dane dyskretne (nieciągłe) obserwacje dla punktów, lub wzdłuż profili
 - przykłady: pomiary GPS, stacje meteorologiczne, zanieczyszczenie gleby, wilgotność gleby, itp.
- Potrzeba konwersji danych dyskretnych do ciągłych powierzchni wynika z konieczności ich stosowania w modelowaniu za pomocą GIS



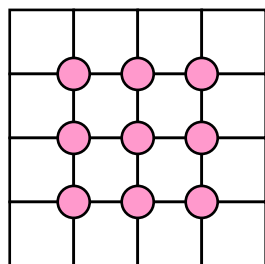
Rozwiązanie problemu – interpolacja

Ogólny schemat procesu modelowania / interpolacji

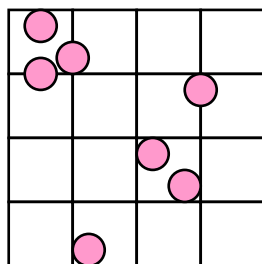


Interpolacja - typ próbkowania

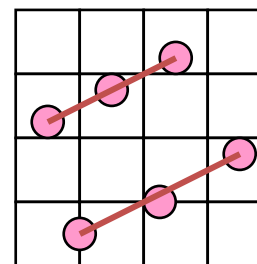
Zastosowana metoda poboru próbek ma podstawowe znaczenie dla wyboru metody i jakości interpolacji



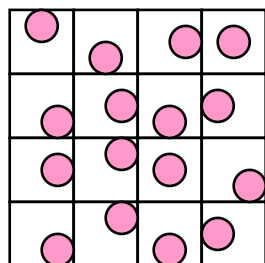
Regularna



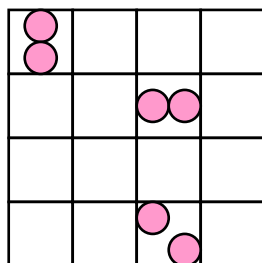
Losowa



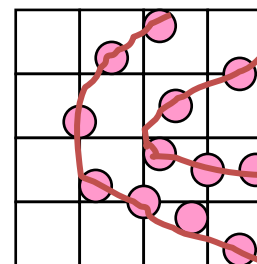
Profilowa



**Losowa
stratyfikowana**



**Preferencyjna
(skupiona)**



Izoliniowa

Interpolacja - metody

Metody małoskalowe (globalne):

Zastosowanie jednakowego algorytmu do danych ze wszystkich punktów pomiarowych, niezależnie od ich rozmieszczenia czy wagi.

Zalety: szybka, łatwa w zastosowaniu

Wady: zbyt ogólna, wyniki podlegają silnej generalizacji w stosunku do rzeczywistości

Metody wielkoskalowe (lokalne):

Dobór algorytmu w zależności od typu pomiarów, ich gęstości, oczekiwanych wyników.

Stosowana często do wielu zbiorów o dużej szczegółowości a następnie „sklejana” w jeden globalny zbiór danych.

Zalety: dokładna, dostosowana indywidualnie do potrzeb

Wady: pracochłonna, wymaga większych zasobów, może być zbyt szczegółowa i zajmować większą ilość miejsca

Interpolacja - metody

Metody wierne:

- Ściśle uwzględniają wszystkie dane pomiarowe zarówno w obszarach gdzie wymagana jest duża dokładność jak i tam gdzie dokładność mogłaby być mniejsza.
- Wskazane do zastosowania kiedy nie mamy konieczności ograniczenia ilości danych, mocy komputera.

Metody wygładzające:

- Nie uwzględniają ściśle danych pomiarowych,
- Wskazane w sytuacji niepewności co do jakości danych,
- Wyniki mogą odbiegać od rzeczywistości.

Interpolacja - metody

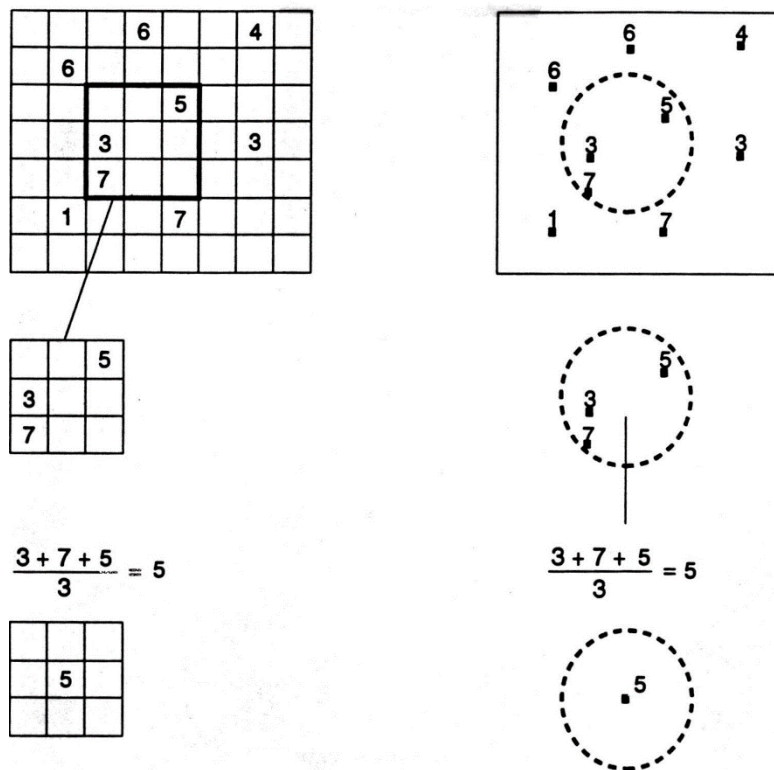
Większość oprogramowania GIS uwzględnia co najmniej kilka metod interpolacji

Najbardziej typowe z nich to:

- Poligony Thiessen'a
- Triangulacja (Triangulated Irregular Networks – TIN)
- Przestrzenne średnie ruchome (moving average)
- Powierzchnie trendu
- Metoda średniej ważonej odległością (inverse distance weighted)
- Metoda kriging

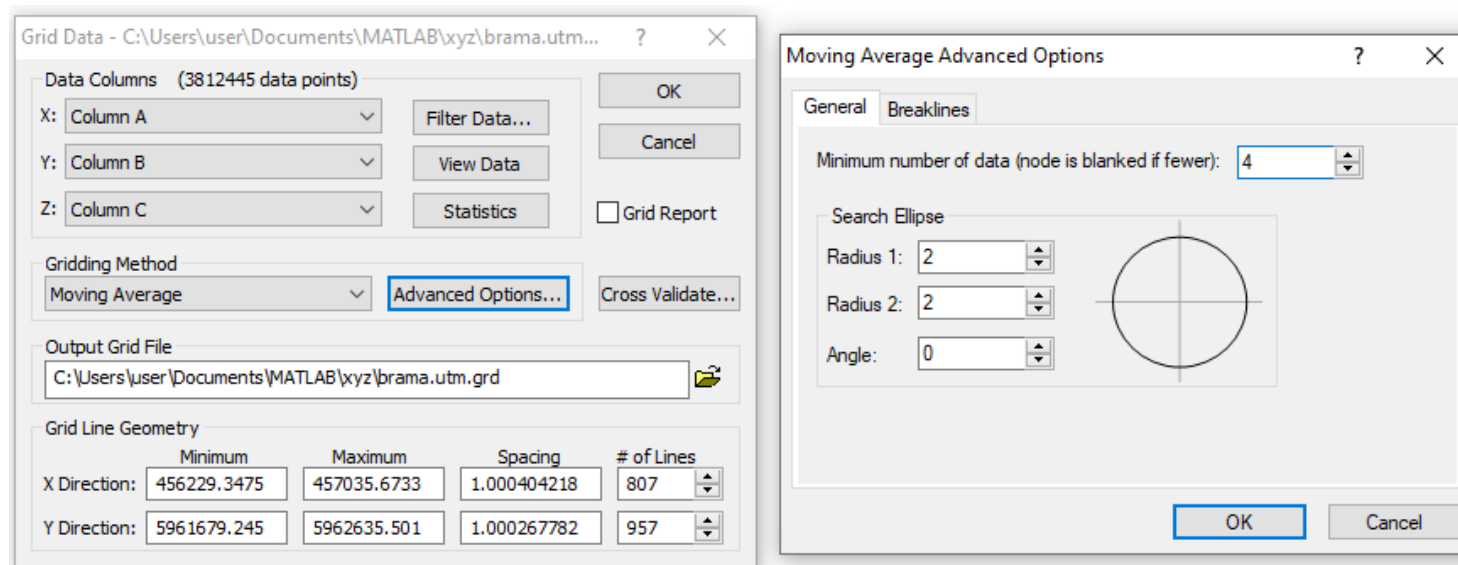
Interpolacja – metoda Moving Average - Zasada

Metoda moving average (MA) bazuje na założeniu, że wartość cechy badanej w danym punkcie jest zależna od średniej wartości tej cechy w punktach bazowych znajdujących się w określonym otoczeniu.

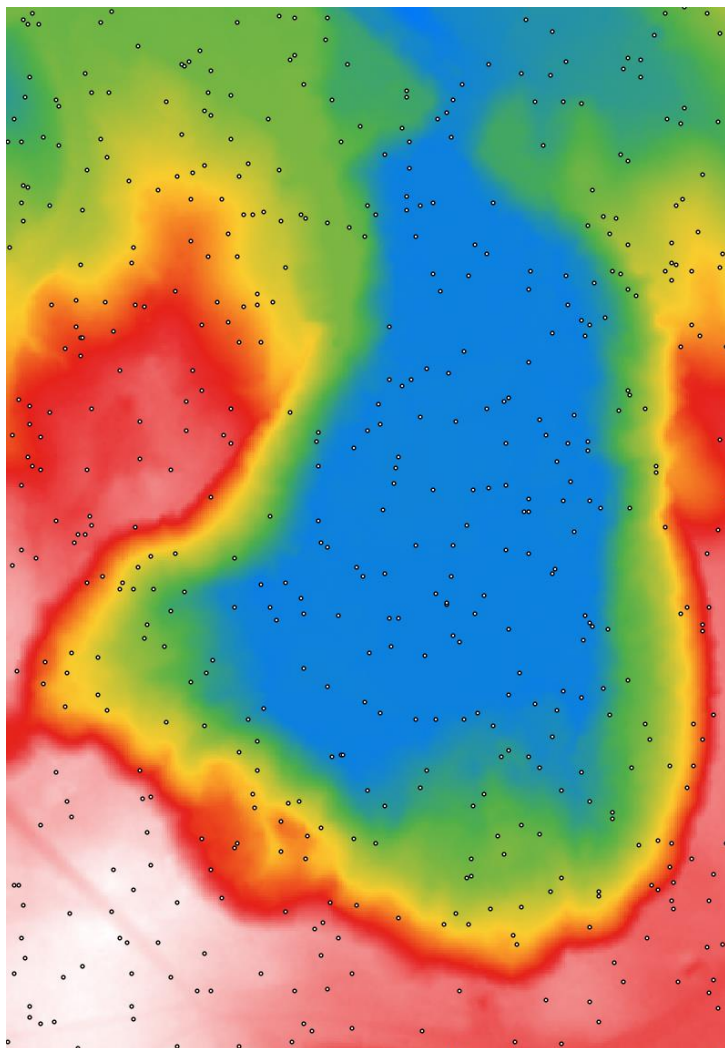


Interpolacja - metoda MA - Właściwości

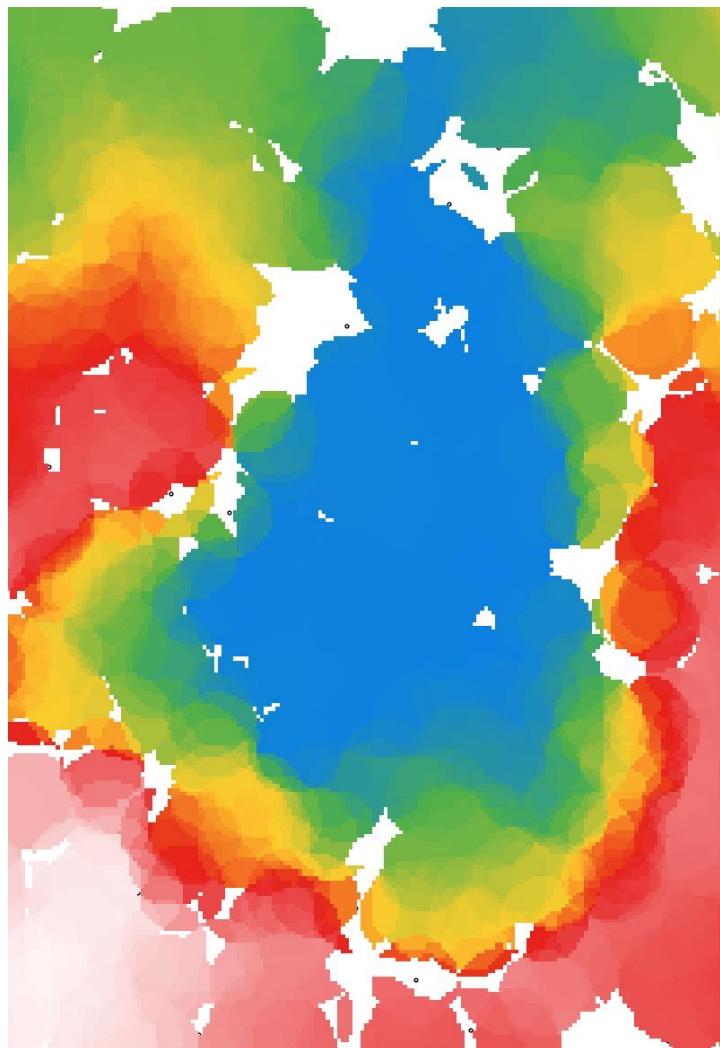
- Metoda mająca zastosowanie zarówno dla danych wektorowych, jak i rastrowych.
- Oblicza nieznaną wartość cechy dla określonej lokalizacji na podstawie zakresu wartości dla najbliższych leżących punktów pomiarowych.
- Kryteria „sąsiedztwa” do obliczeń są określane za pomocą reguły wprowadzanej przez operatora.
- Przy interpolacji określamy: wielkość i kształt sąsiedztwa, minimalną liczbę punktów pomiarowych.
- Bardzo popularna w GIS, bardzo szybka, niezbyt precyzyjna.



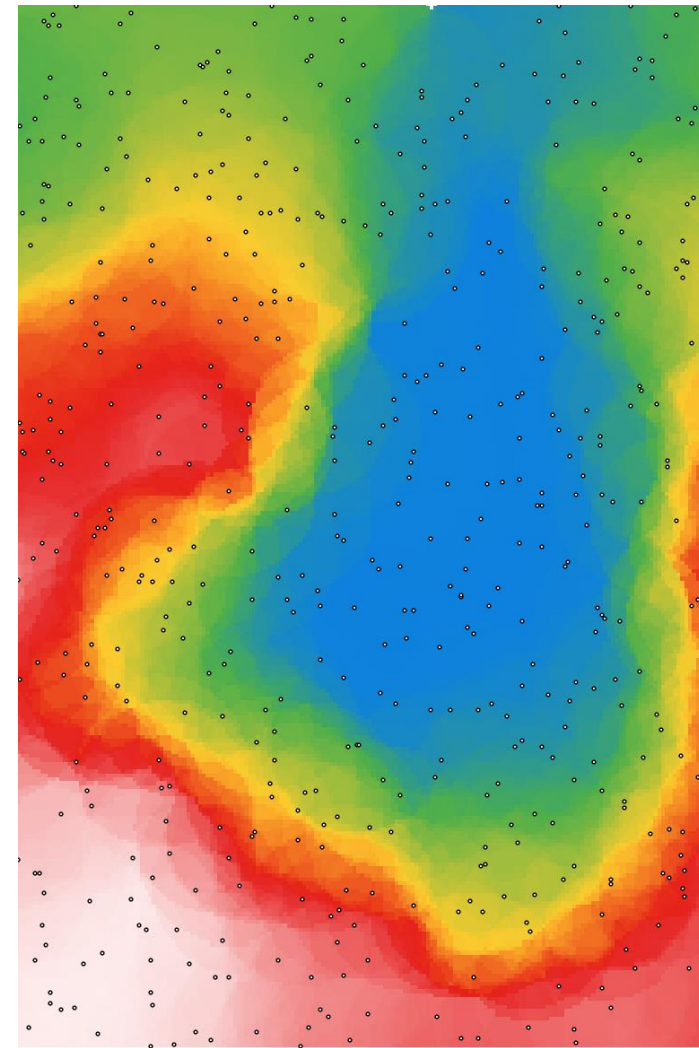
Interpolacja – metoda Moving Average - Przykłady



Oryginał



20m



40m

Interpolacja – metoda IDW

Metoda odwrotnej odległości (IDW) bazuje na założeniu, że wartość cechy badanej w danym punkcie jest zależna od wartości tej cechy w najbliższych punktach bazowych. Wartości mierzone w odległych punktach bazowych mają znacznie mniejsze znaczenie lub w ogóle nie mają znaczenia.

$$\hat{Z}_j = \frac{\sum_{i=1}^n \frac{Z_i}{h_{ij}^\beta}}{\sum_{i=1}^n \frac{1}{h_{ij}^\beta}}$$

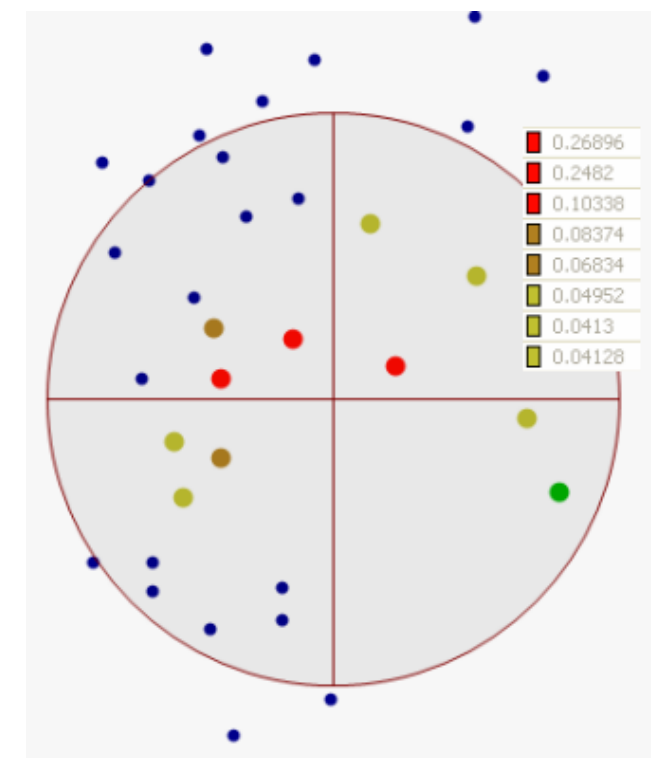
Z_j – wartość cechy Z estymowanej w punkcie j
 Z_i – wartość cechy Z zmierzona w punkcie i
(jednym z n punktów danych w otoczeniu)
 h_{ij} – efektywna odległość między punktami i i j
 β – wykładnik potęgowy – waga odległości

Planarna: $z(x,y) = A + Bx + Cy$

Bi-liniowa: $z(x,y) = A + Bx + Cy + Dxy$

Kwadratowa: $z(x,y) = A + Bx + Cy + Dx^2 + Exy + Fy^2$

Sześcienne: $z(x,y) = A + Bx + Cy + Dx^2 + Exy + Fy^2 + Gx^3 + Hx^2y + Ixy^2 + Jy^3$



Interpolacja – metoda Inverse Distance Weighted (IDW)

- Oblicza nieznaną wartość cechy dla określonej lokalizacji na podstawie zakresu wartości dla najbliższych leżących punktów pomiarowych, jednocześnie zależnych od ich odległości od lokalizacji.
- Kryteria „sąsiedztwa” do obliczeń są określane za pomocą reguły wprowadzanej przez operatora.
- Przy interpolacji określamy: wielkość i kształt sąsiedztwa, minimalną liczbę punktów pomiarowych.
- Bardzo popularna w GIS, dość szybka, precyzyjna dla dużej ilości danych, średnio precyzyjna dla niewielkiej ilości danych.

Grid Data - C:\Users\user\Documents\MATLAB\xyz\brama.utm... ? X

Data Columns (3812445 data points)

X: Column A Filter Data... OK

Y: Column B View Data Cancel

Z: Column C Statistics

☐ Grid Report

Gridding Method

Inverse Distance to a Power Advanced Options... Cross Validate...

Output Grid File

C:\Users\user\Documents\MATLAB\xyz\brama.utm.grd

Grid Line Geometry

	Minimum	Maximum	Spacing	# of Lines
X Direction:	456229.3475	457035.6733	1.000404218	807
Y Direction:	5961679.245	5962635.501	1.000267782	957

Inverse Distance Advanced Options ? X

General Search Breaklines and Faults

Power: 2

Smoothing: 0

Anisotropy

Ratio: 1

Angle: 0

OK Cancel

Inverse Distance Advanced Options ? X

General Search Breaklines and Faults

☐ No Search (use all of the data)

Search Ellipse

Radius 1: 625

Radius 2: 625

Angle: 0

Number of sectors to search: 4

Maximum number of data to use from ALL sectors: 64

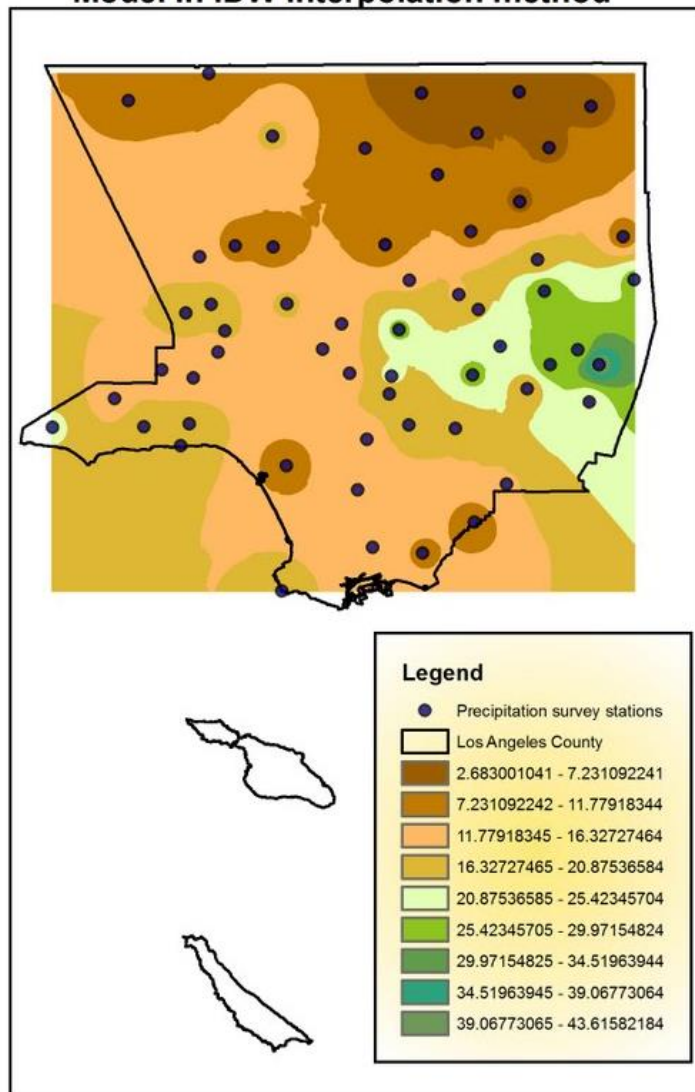
Maximum number of data to use from EACH sector: 16

Minimum number of data in all sectors (node is blanked if fewer): 8

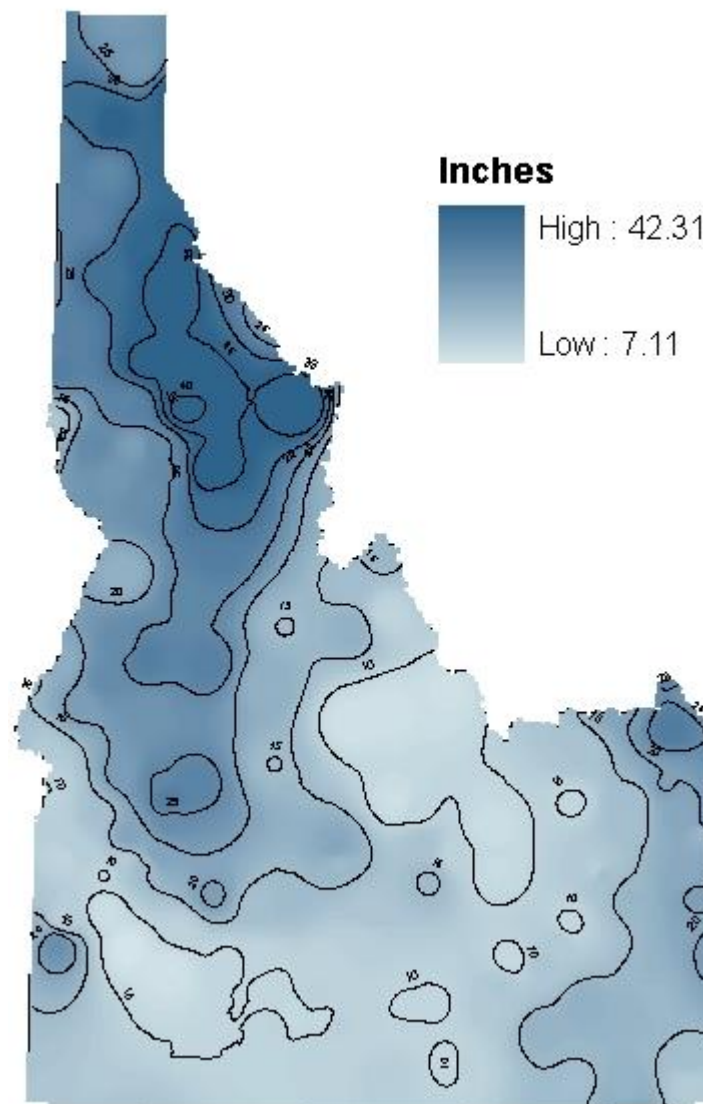
Blank node if more than this many sectors are empty: 3

OK Cancel

Model in IDW interpolation method



Inverse Distance



Interpolacja – metoda kriging

Kriging (interpolacja geostatystyczna) to grupa metod estymacji w których główna zasada mówi, że prognoza w danej lokalizacji jest kombinacją sąsiednich obserwacji. Waga nadawana każdej z obserwacji jest zależna od stopnia (przestrzennej) korelacji - stąd też bierze się istotna rola semiwariogramów.

Lokalna średnia może być uzyskana za pomocą wyliczenia regresji liniowej pomiędzy zmienną badaną a zmienną dodatkową.

For instance, if one assumes, based on the homogeneity of samples in area A where the variable is distributed, the hypothesis that the **first moment** is stationary (i.e. all random variables have the same mean), then one is assuming that the mean can be estimated by the arithmetic mean of sampled values.

The hypothesis of stationarity related to the **second moment** is defined in the following way: the correlation between two random variables solely depends on the spatial distance between them, and is independent of their location. Thus if $\mathbf{h} = x_2 - x_1$ and $|\mathbf{h}| = h$ then:

$$C(Z(x_1), Z(x_2)) = C(Z(x_i), Z(x_i + \mathbf{h})) = C(h)$$

$$\gamma(Z(x_1), Z(x_2)) = \gamma(Z(x_i), Z(x_i + \mathbf{h})) = \gamma(h)$$

and, for simplicity, we define $C(x_i, x_j) = C(Z(x_i), Z(x_j))$ and $\gamma(x_i, x_j) = \gamma(Z(x_i), Z(x_j))$.

This hypothesis allows one to infer those two measures – the **variogram** and the **covariogram**:

$$\gamma(h) = \frac{1}{2|N(h)|} \sum_{(i,j) \in N(h)} (Z(x_i) - Z(x_j))^2$$

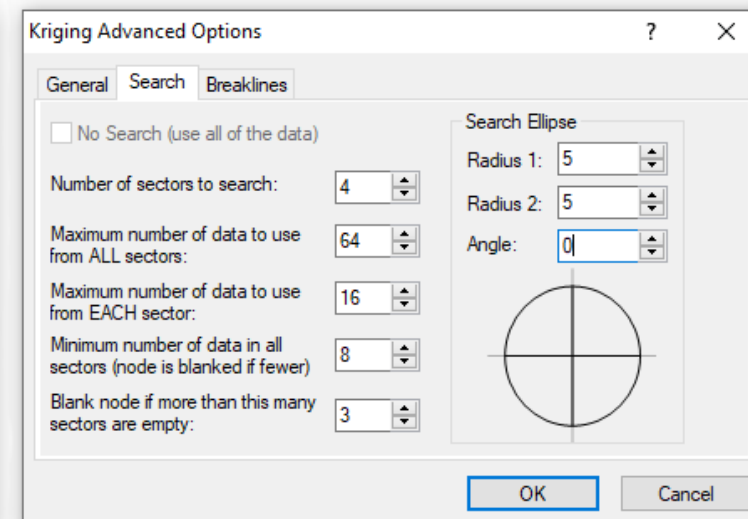
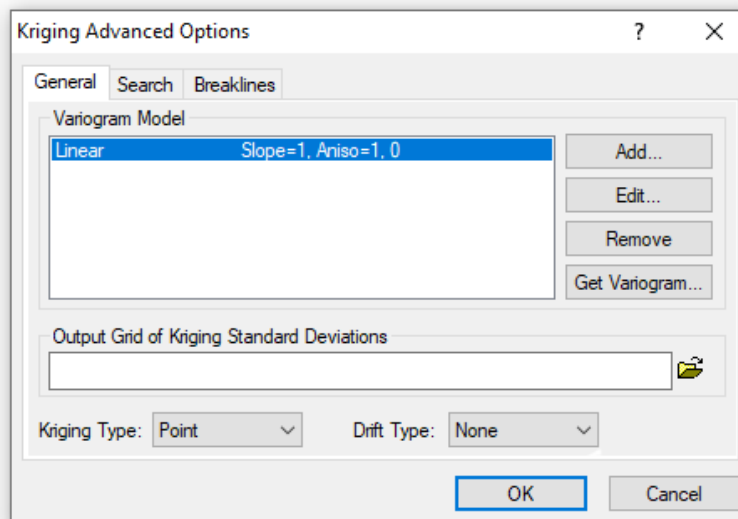
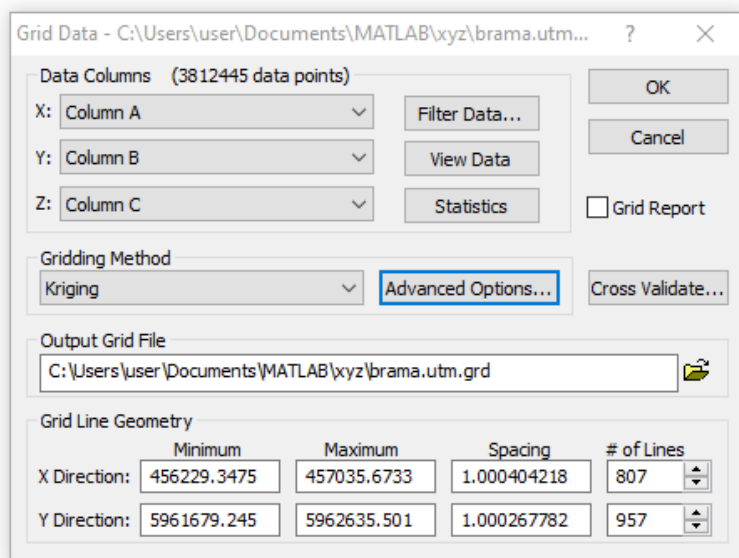
$$C(h) = \frac{1}{|N(h)|} \sum_{(i,j) \in N(h)} (Z(x_i) - m(h)) (Z(x_j) - m(h))$$

where:

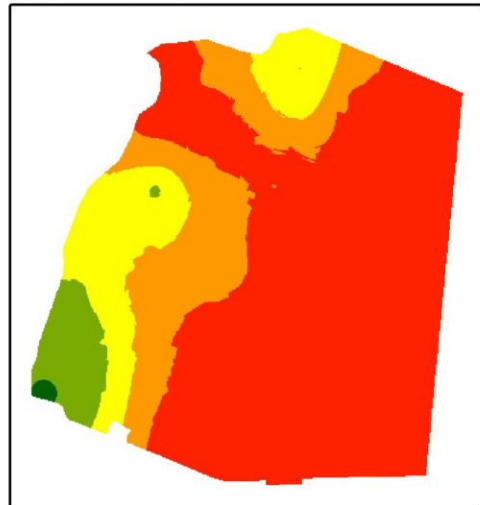
- $m(h) = \frac{1}{2|N(h)|} \sum_{(i,j) \in N(h)} Z(x_i) + Z(x_j)$;
- $N(h)$ denotes the set of pairs of observations i, j such that $|x_i - x_j| = h$, and $|N(h)|$ is the number of pairs in the set. In this set, (i, j) and (j, i) denote the same element. Generally an "approximate distance" h is used, implemented using a certain tolerance.

Interpolacja – metoda kriging

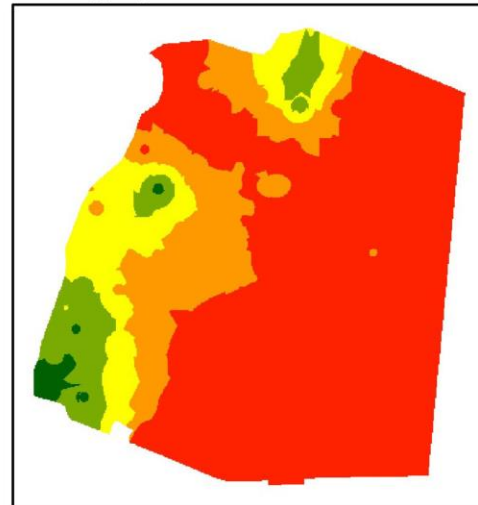
- Oblicza nieznaną wartość cechy dla określonej lokalizacji na podstawie korelacji wartości pobliskich punktów pomiarowych.
- Kryteria „sąsiedztwa” do obliczeń są określane za pomocą reguły wprowadzanej przez operatora.
- Przy interpolacji określamy: wielkość i kształt sąsiedztwa, minimalną liczbę punktów pomiarowych, stosowany semiwariogram.
- Dość popularna w GIS, najlepsze wyniki dla małej ilości danych, dość dobre wyniki dla dużej ilości danych. **Bardzo powolna, szczególnie gdy mamy wiele punktów pomiarowych.**



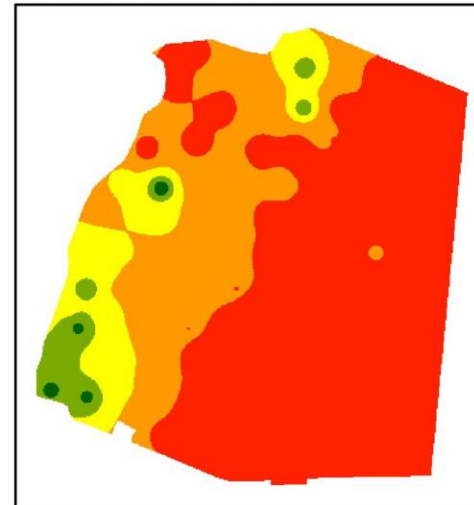
KRIG - Dense



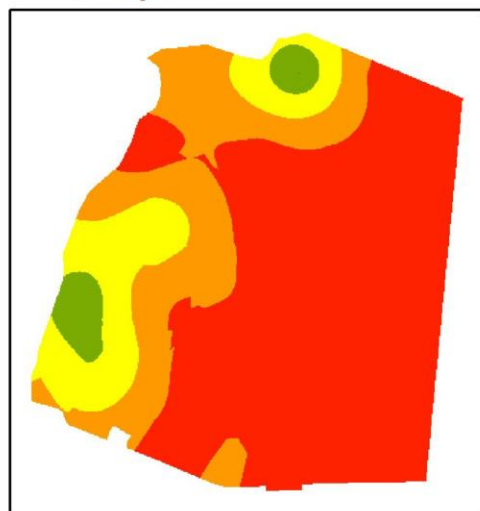
IDW [p1] - Dense



IDW [null] - Dense



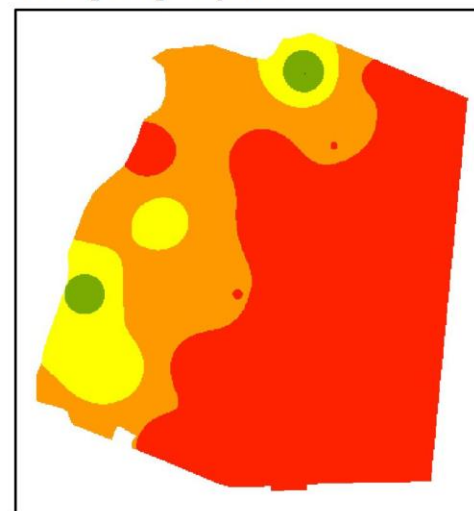
KRIG - Sparse



IDW [p1] - Sparse

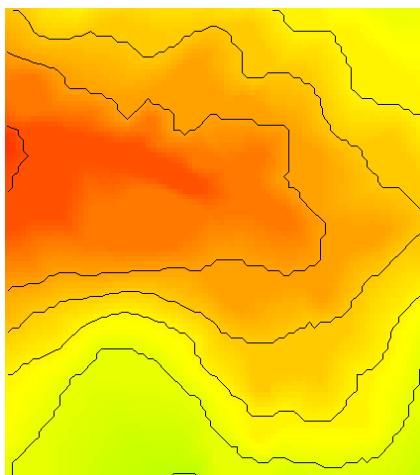


IDW [null] - Sparse

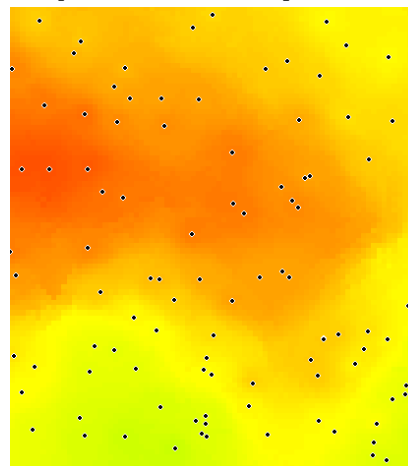


K Rate [kg/ha]  60 - 77.2  77.3 - 94.4  94.5 - 111.6  111.7 - 128.8  128.9 - 146

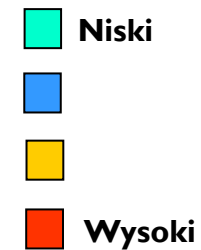
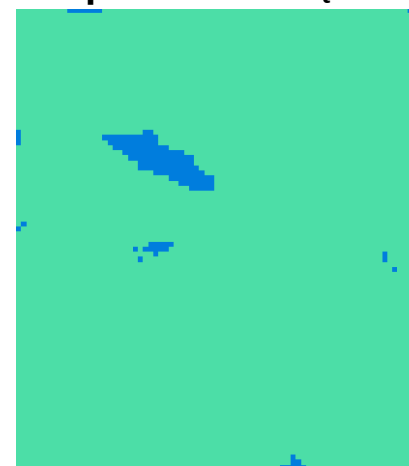
Rzeczywista powierzchnia



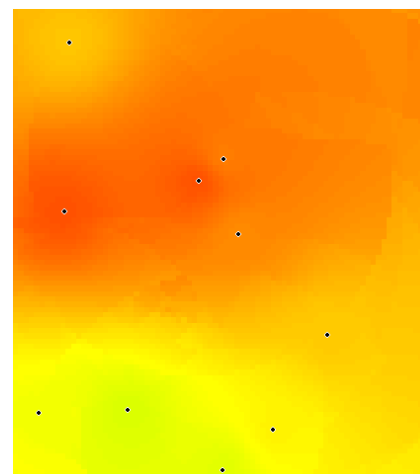
Interpolacja w oparciu o 100 punktów danych



Mapa rozkładu błędów



Interpolacja w oparciu o 10 punktów danych



Mapa rozkładu błędów



THE END

