Geostatistika

Interpolacija

Milan Kilibarda
kili@grf.bg.ac.rs

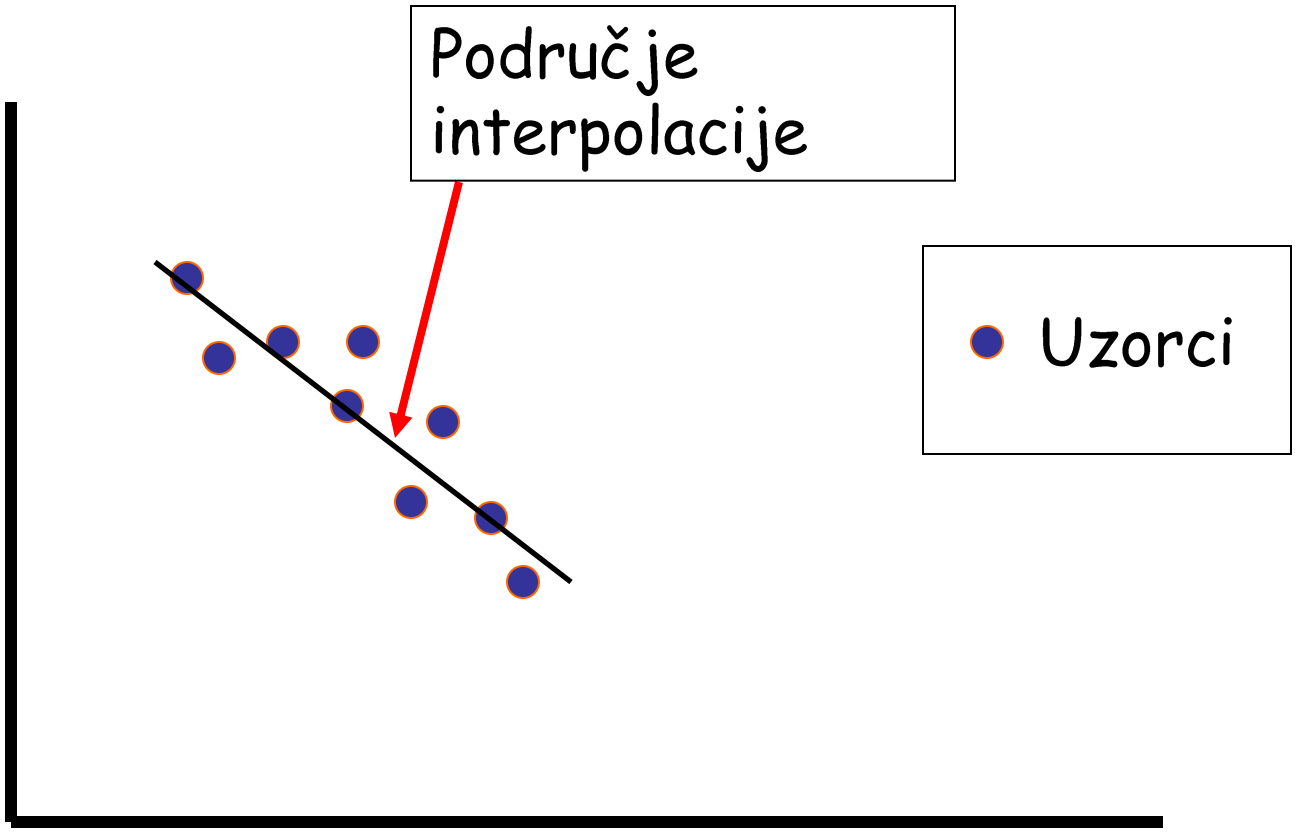
- Interpolacija

Procena vrednosti atributa unutar područja na kojem imamo merene uzorke.

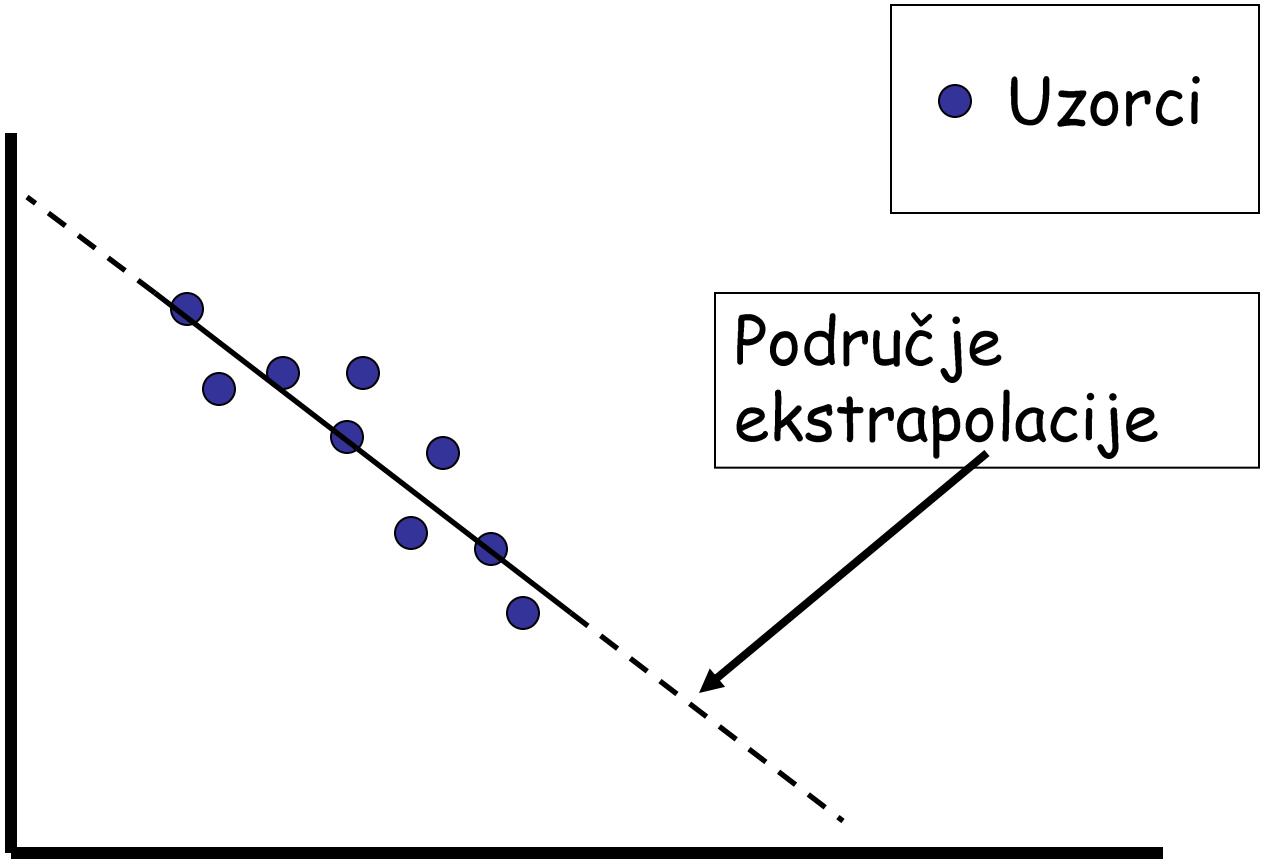
- Ekstrapolacija

Procena vrednosti atributa van područja na kojem imamo merene uzorke.

Interpolacija

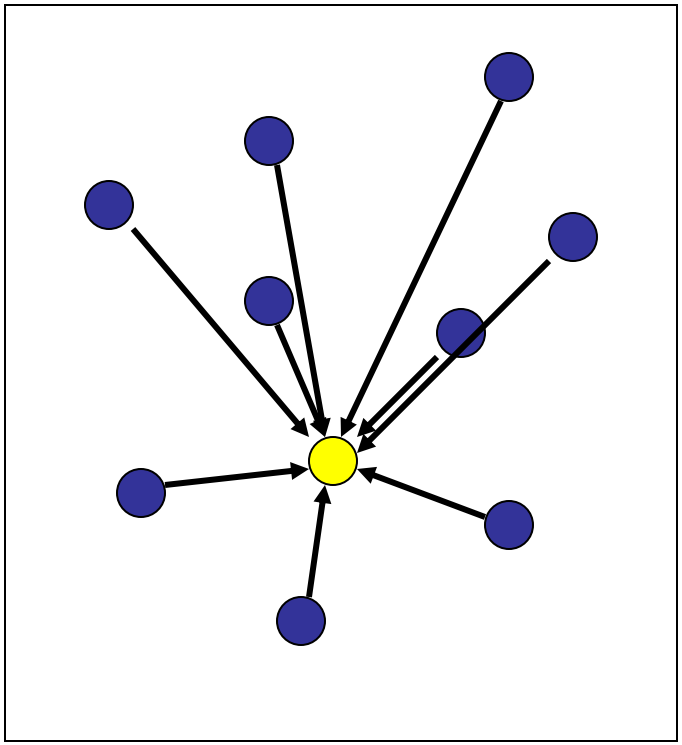


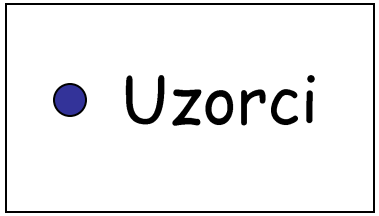
Ekstrapolacija



Globalna Interpolacija

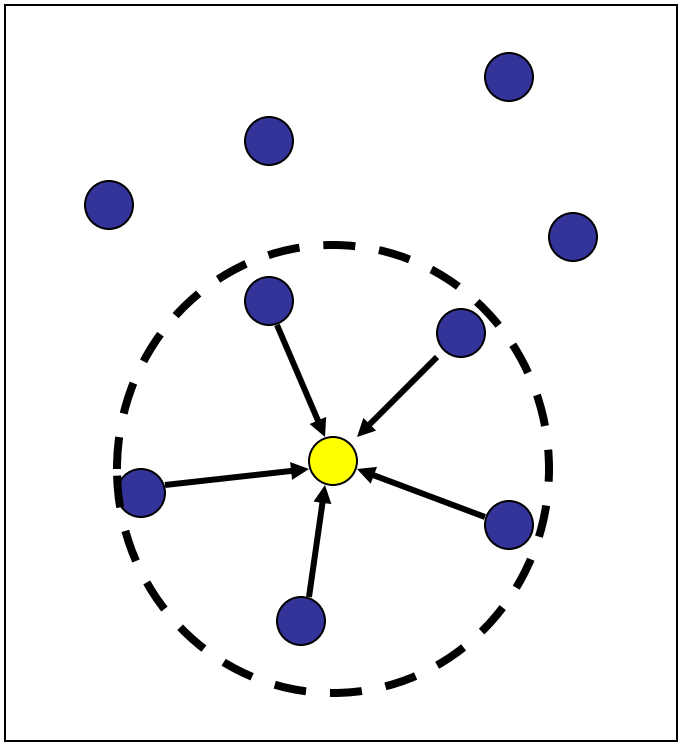
Koristi ***sve*** poznate uzorke na području kako bi se procenila vrednost na traženoj lokaciji





Lokana Interpolacija

Koristi samo susedne uzorke za procenu vrednosti na traženoj lokaciji



Uzorci

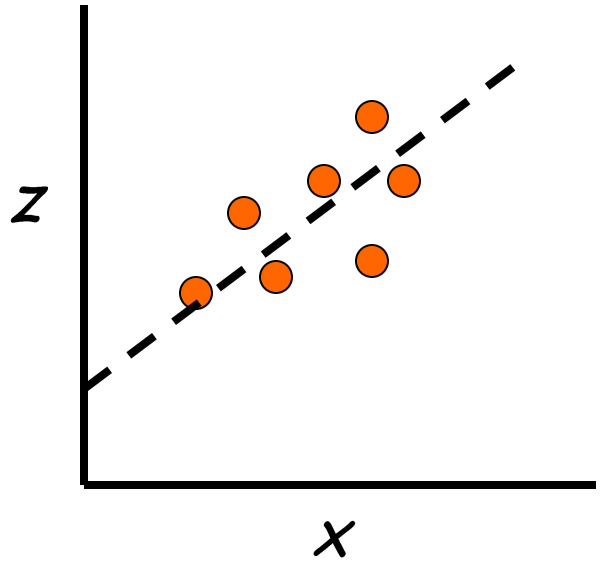
Susedni uzorci se biraju tako što se zada ***n*** broj najbližih uzoraka ili se zada radijus pretraživanja

Trend površi

Trend površi 1. reda

U jedno dimenzionalnom slučaju: **z** varira kao linerana funkcija oko **x**

$$z = b_0 + b_1 x + e$$

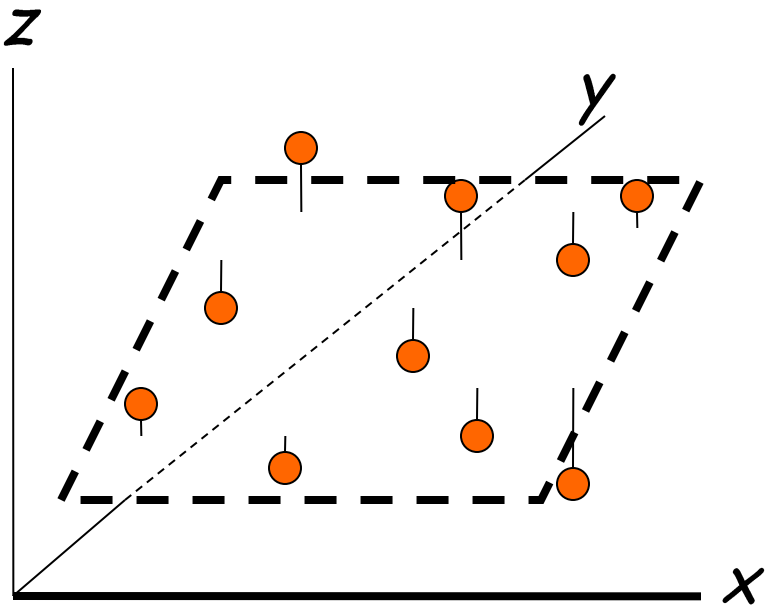


Trend površi

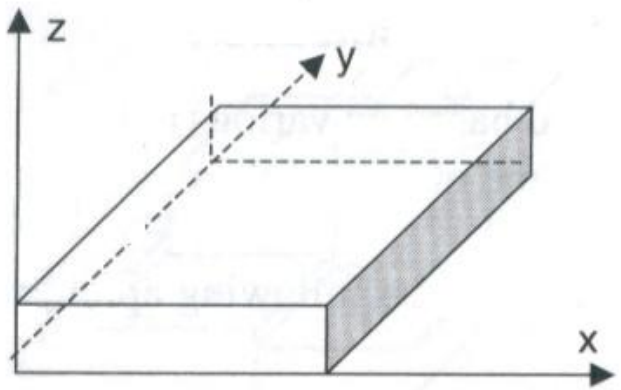
Trend površi 1. reda

U dvo dimenzionalnom slučaju: **z** varira kao linerana funkcija oko **x** i **y**

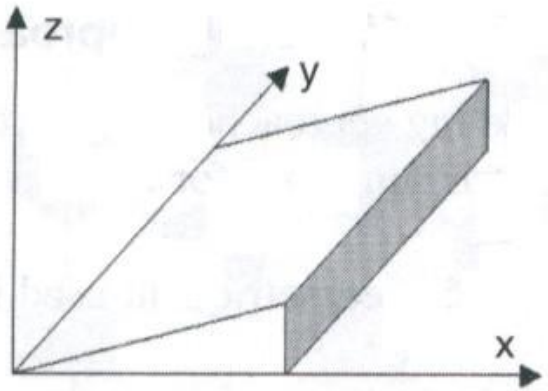
$$z = b_0 + b_1x + b_2y + e$$



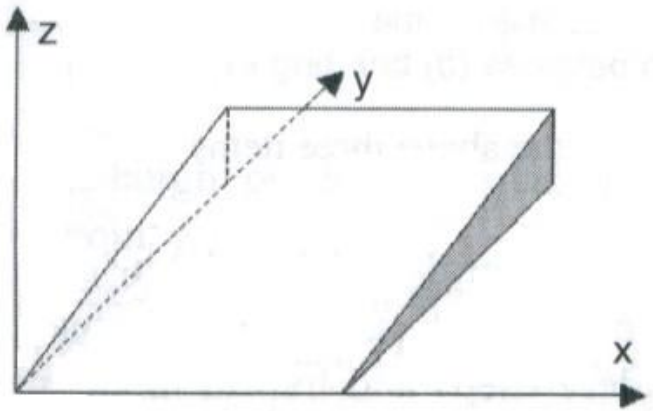
Trend površi



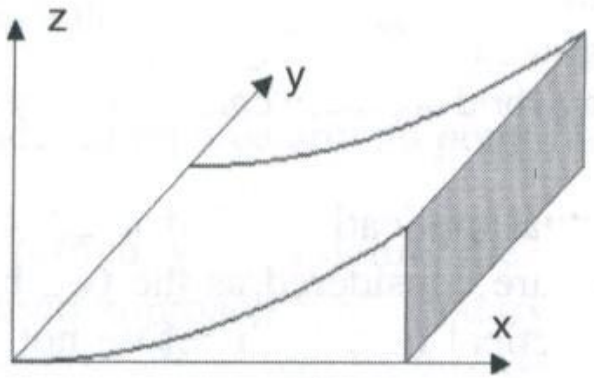
$Z = a_0$



$Z = a_1 x$



$Z = a_2 y$



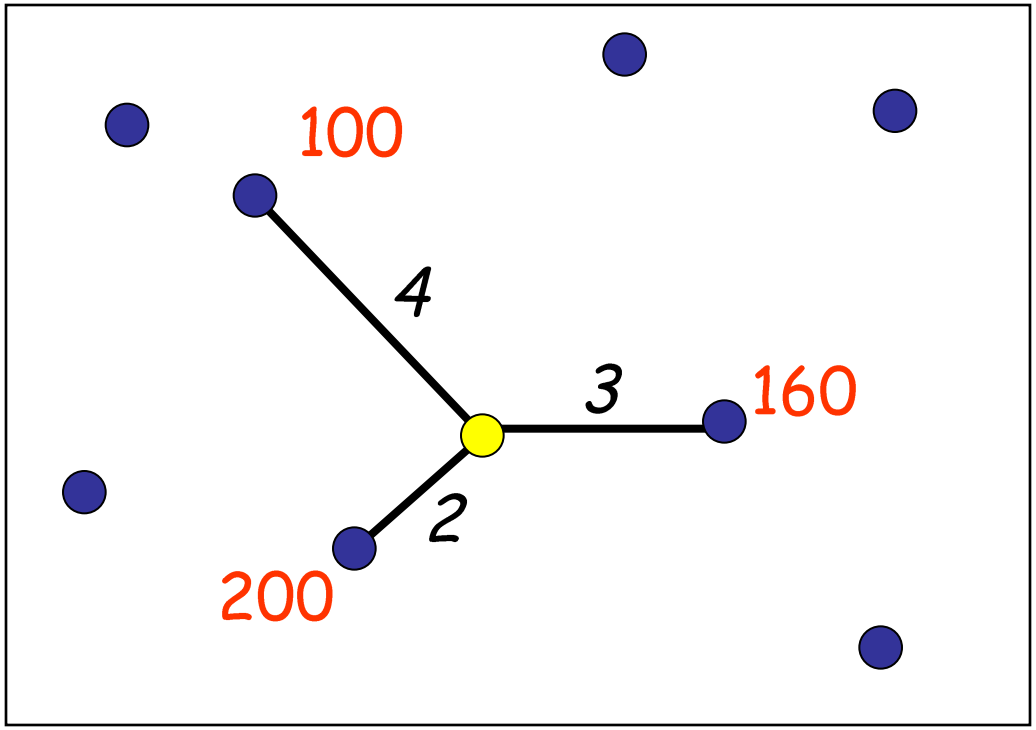
$Z = a_3 x^2$

Inverse Distance Weighted (IDW)

Metoda inverzne distance

Inverse Distance Weighted (primer)

$$z(x,y) = \frac{\sum_{i=1}^{n}\left[\frac{z_i}{d_i^p}\right]}{\sum_{i=1}^{n}\left[\frac{1}{d_i^p}\right]} \quad ili \quad z(x,y) = \sum \lambda_i \times z_i \xrightarrow{with} \sum \lambda_i = 1$$



IDW:
najbliža 3 suseda,

p = 2

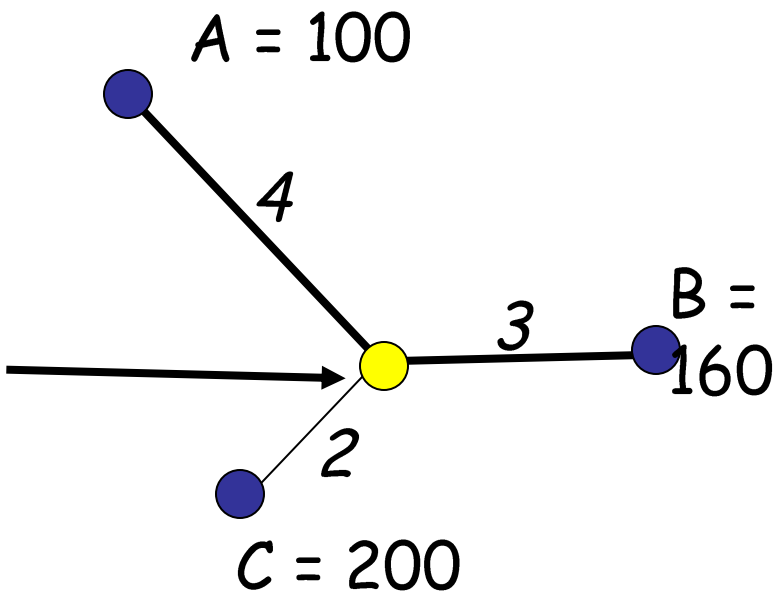
Inverse Distance Weighted (primer)

	Težine	Težine × vrednost
A	$1 / (4^2) = .0625$	$.0625 \times 100 = 6.25$
B	$1 / (3^2) = .1111$	$.1111 \times 160 = 17.76$
C	$/ (2^2) = .2500$	$.2500 \times 200 = 50.00$

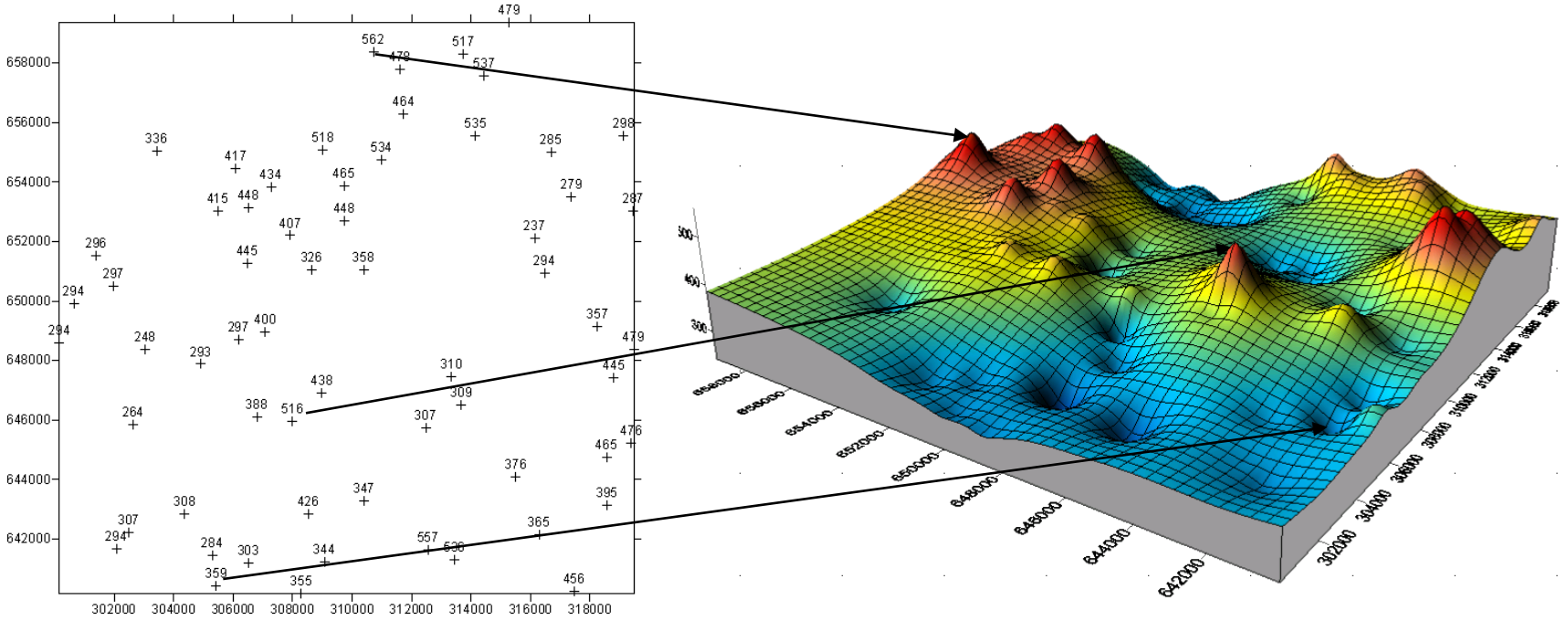
Suma = .4236

6.25 +17.76 + 50.00 = 74.01

74.01 / .4236 = **175**



Nedostaci IDW-a

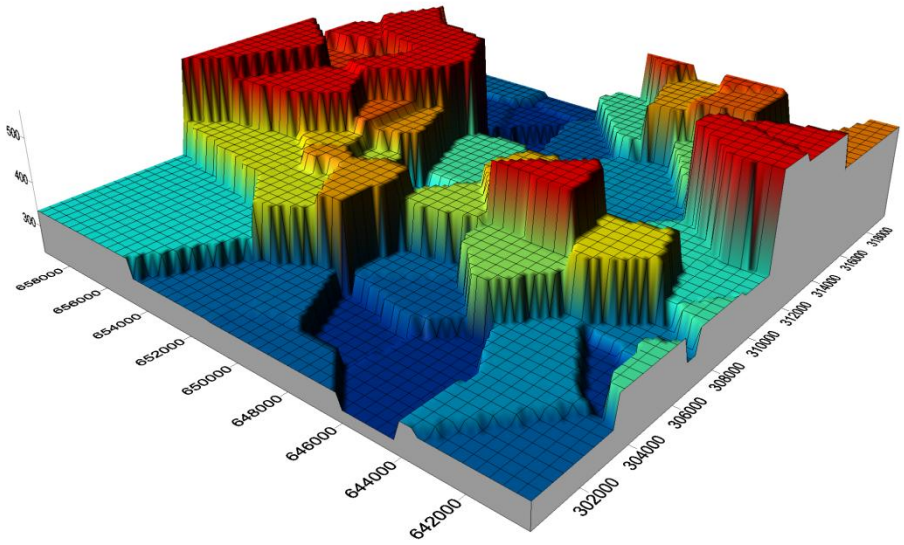


Uzorci

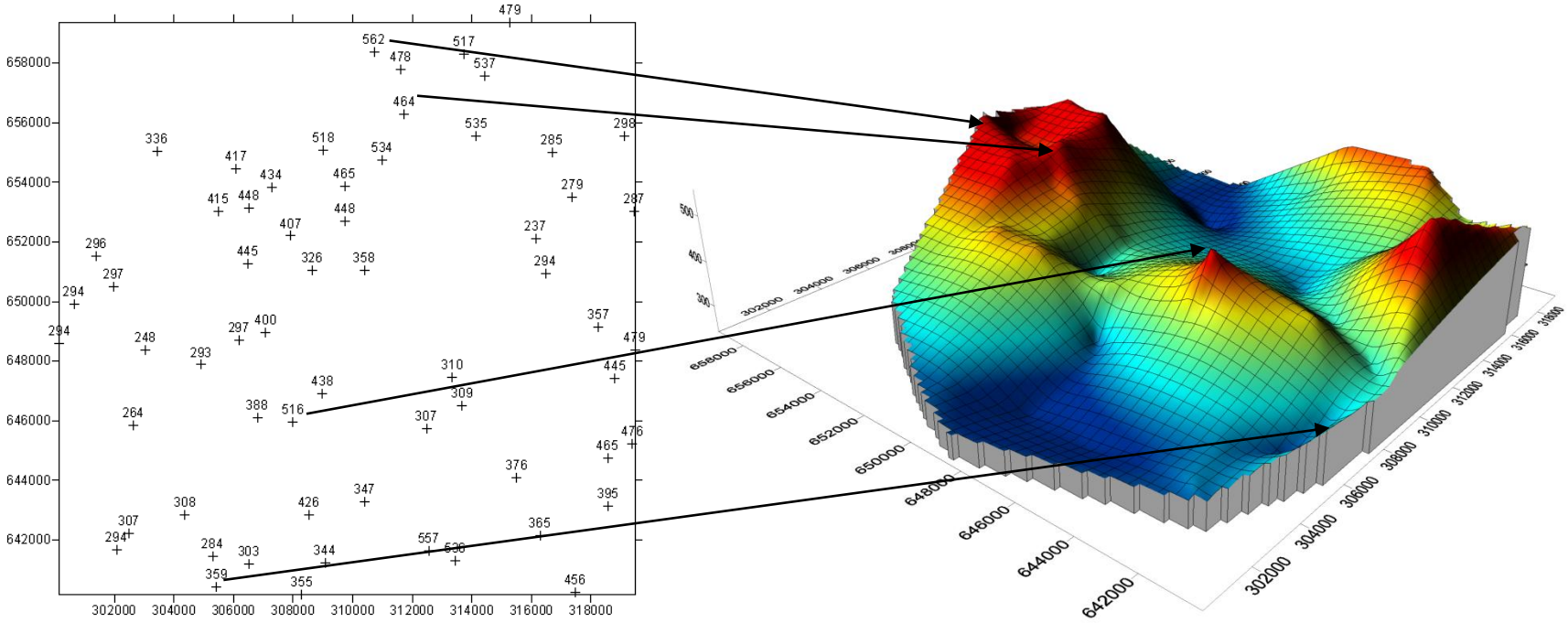
Interpolovana površ

Pregled ostalih metoda interpolacije

Nearest neighbour/najbliži sused

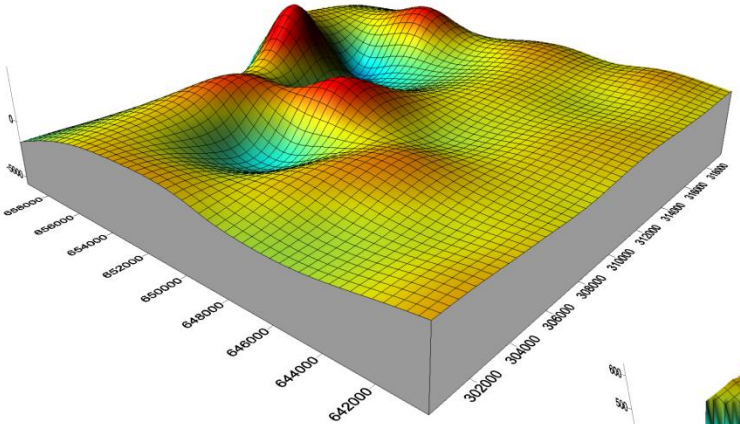


Natural neighbour/prirodni sused



Uzorci

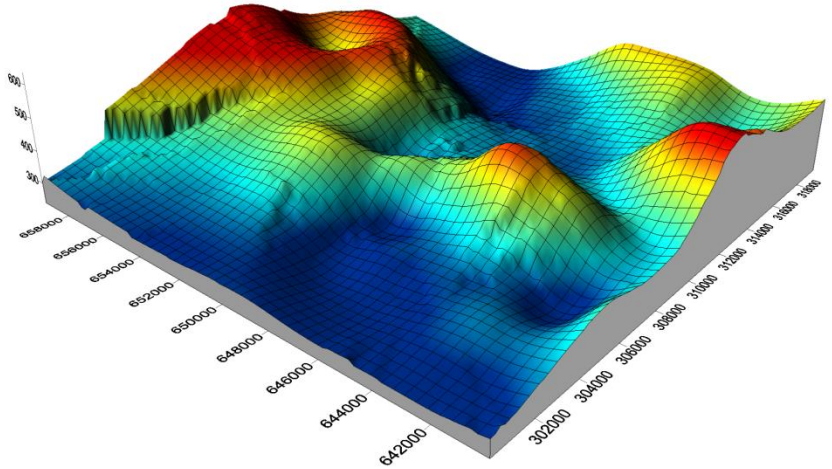
Plate spline/spline funkcija



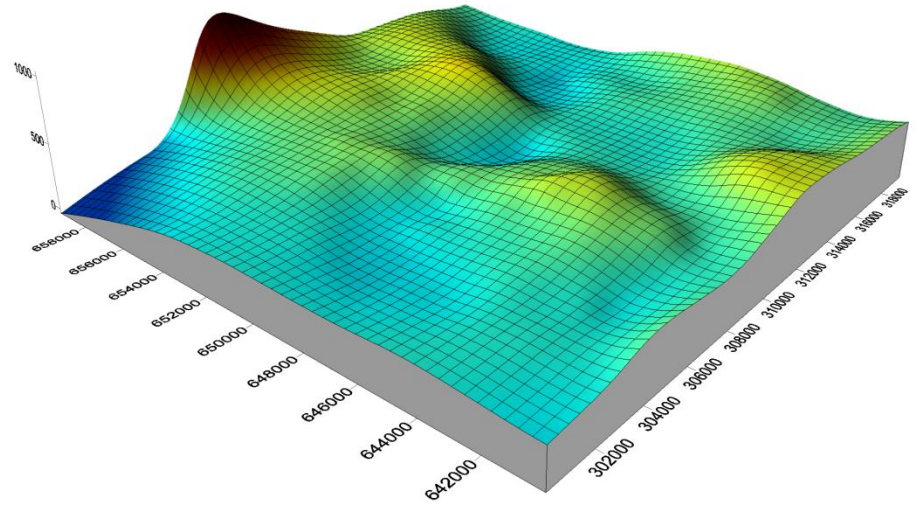
Korišćene sve 62 tačke

$$\phi(r) = (c^2 + r^2)\ln(c^2 + r^2)$$

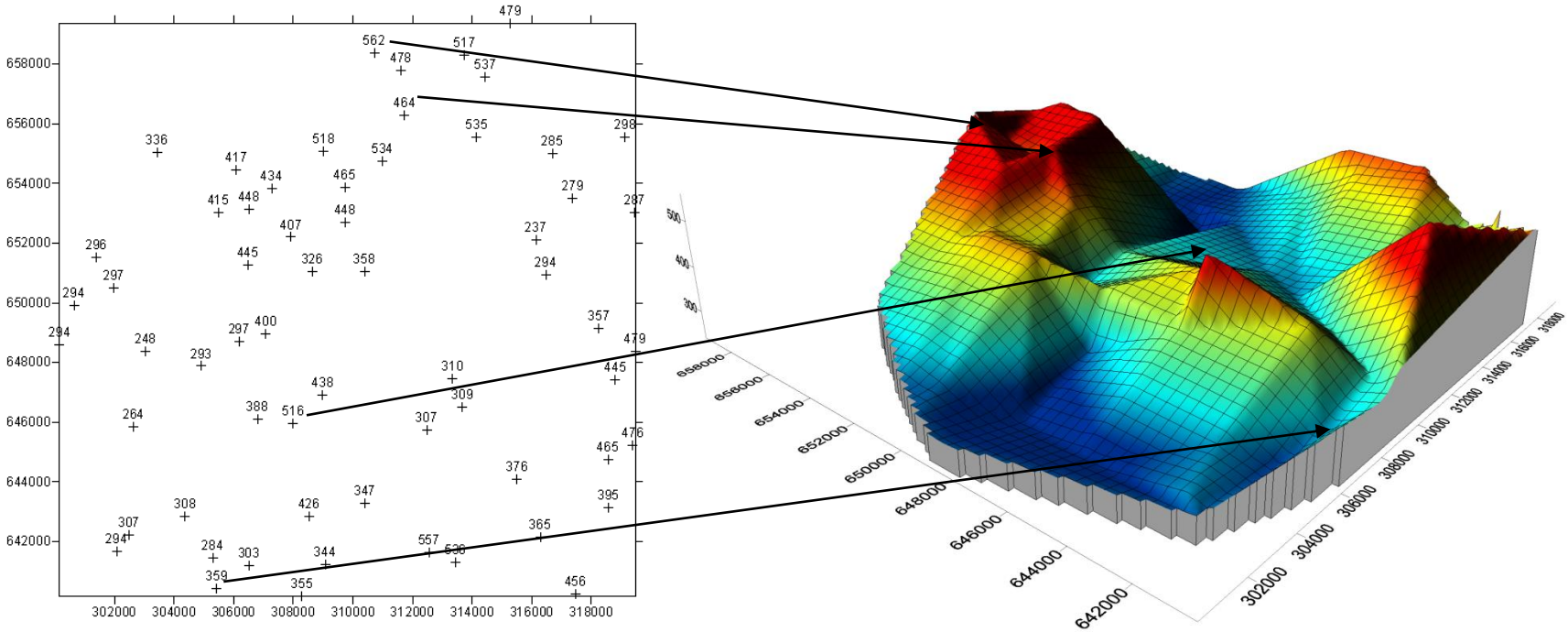
Korišćeno 12 susednih tačaka



Modified Shepard/ modifikovana Šepardova

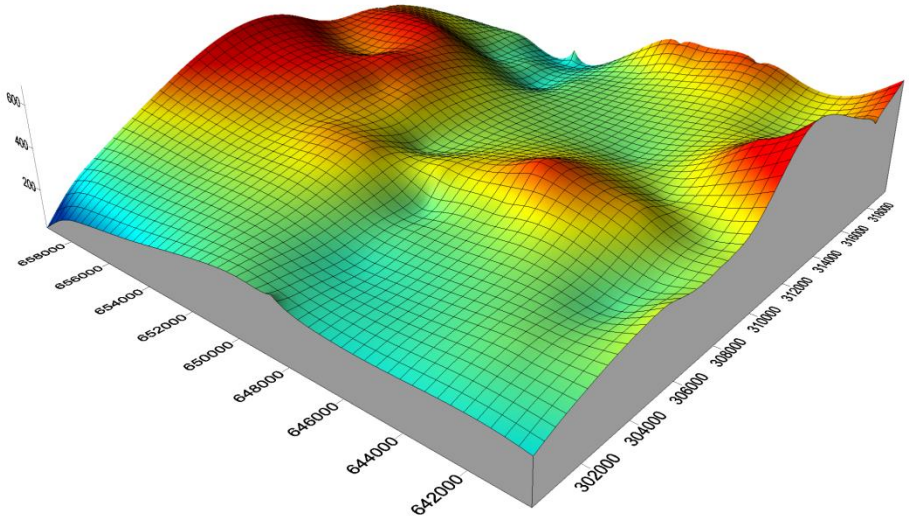


Triangulacija sa linearnom interpolacijom

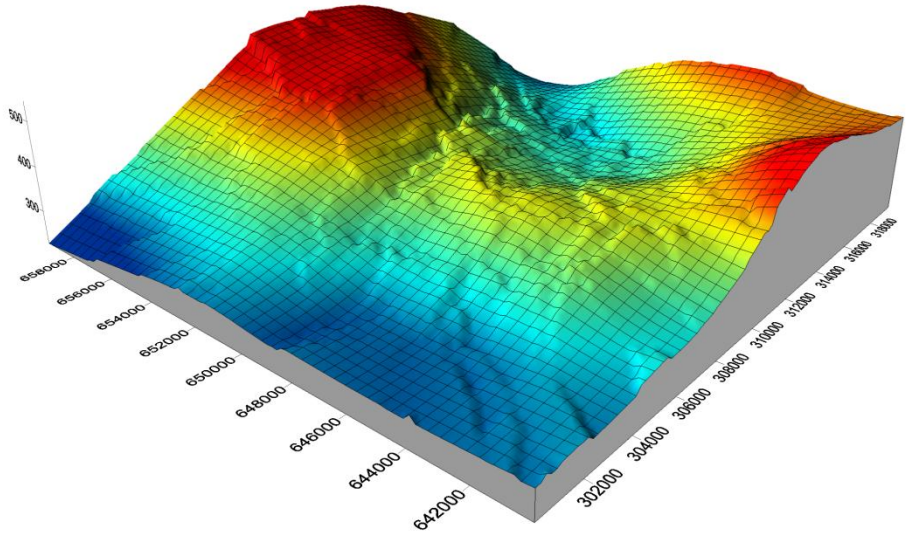


Uzorci

Minimum curvature/ minimalna zakrivljenost



Local polynomial / lokalni polinomi

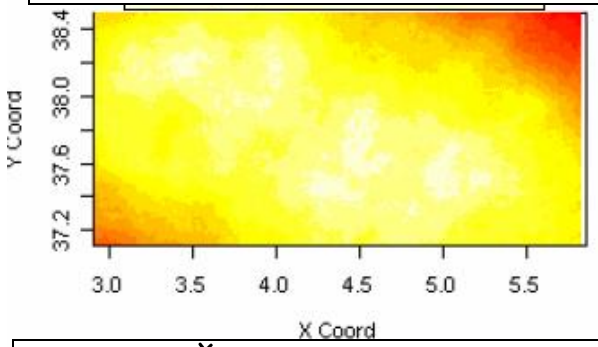


Geostatistika

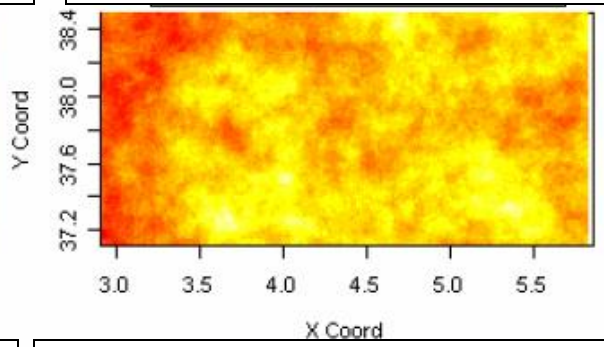
- ***Geostatistika***, nudi mogućnost analitičke obrade i opisa koncepta prostornog kontinuiteta, koji prisutan u svim geonaukama .
- ***Prvi zakon u geografiji***, glasi da je sve u nekom odnosu sa ostalim, ali da su vrednosti koje su prostorno bliže sličnije od onih udaljenijih.
- Geostatistika za razliku od klasične statistike koja obrađuje slučajne promenljive veličine, nije skoncentrisana samo na vrednosti promenljive, već i na položaj promenljive u prostoru, odakle i potiče termin ***regionalizovane promenljive***.

Primeri prostorno promenljivih veličina

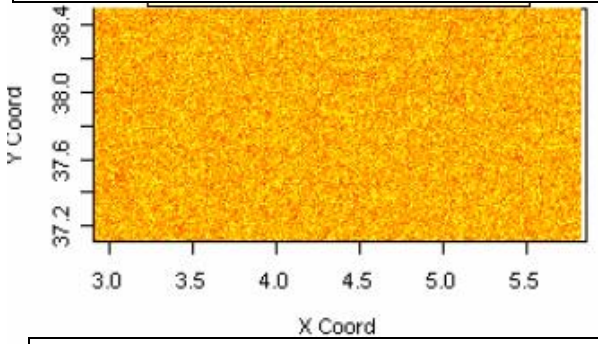
Prostorna karakteristika



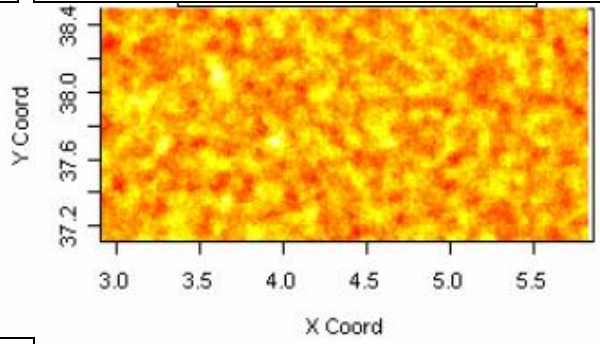
Srednja prostorna korelisanost



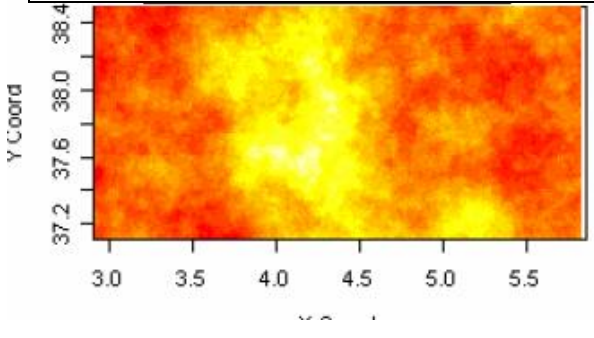
Šum merenja



Niska prostorna korelisanost



Visoka prostorna korelisanost



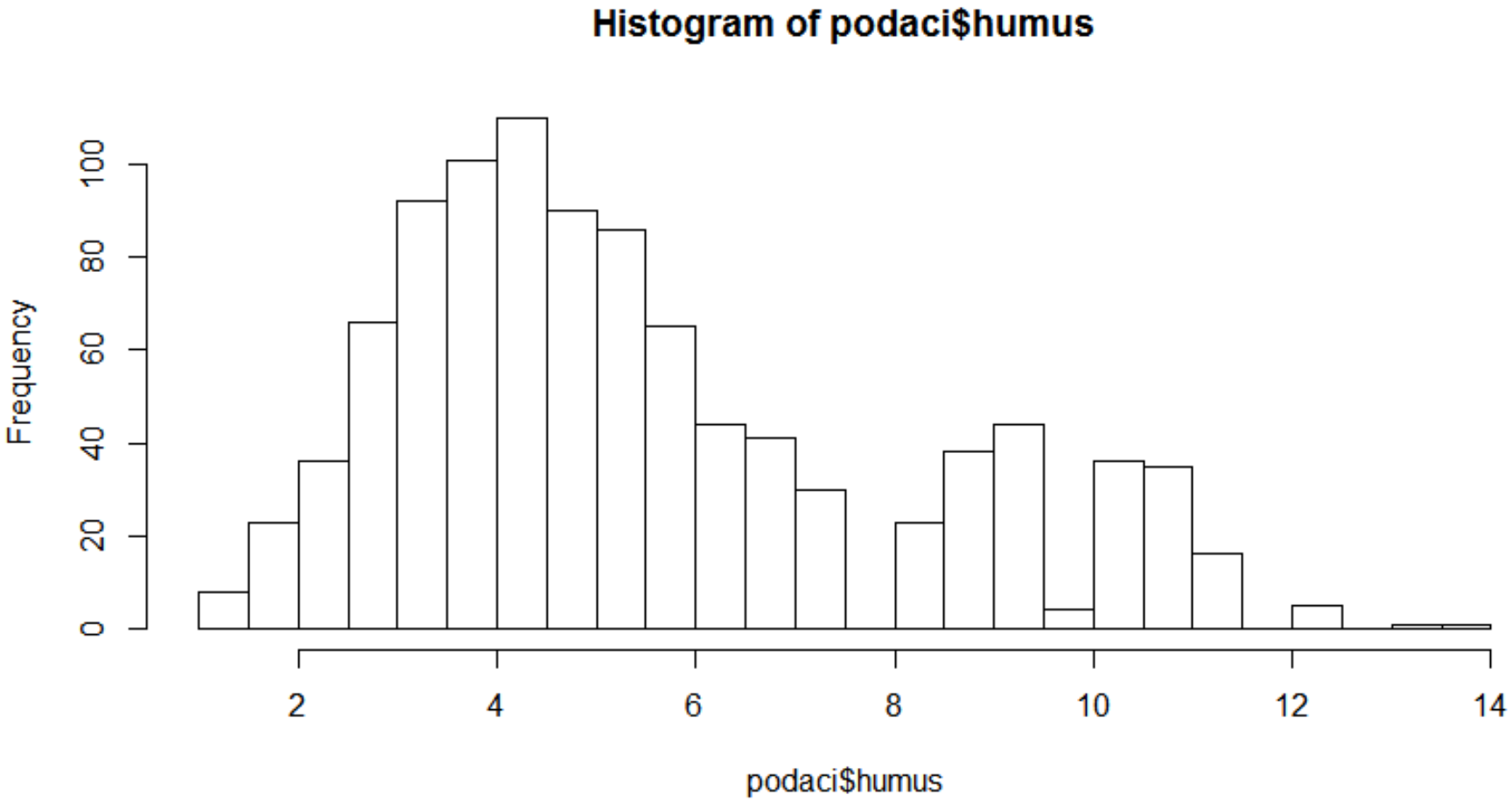
Primeri podataka

	humus	coords.x1	coords.x2
1	4.08	7361881	4915262
2	3.85	7382089	4921515
3	1.29	7375904	4912344
4	5.29	7364948	4922062
5	2.30	7375868	4912359
6	4.08	7375532	4912336
7	8.04	7376905	4913292
8	3.57	7365860	4917728
9	2.53	7378367	4911428
10	2.30	7365411	4921999

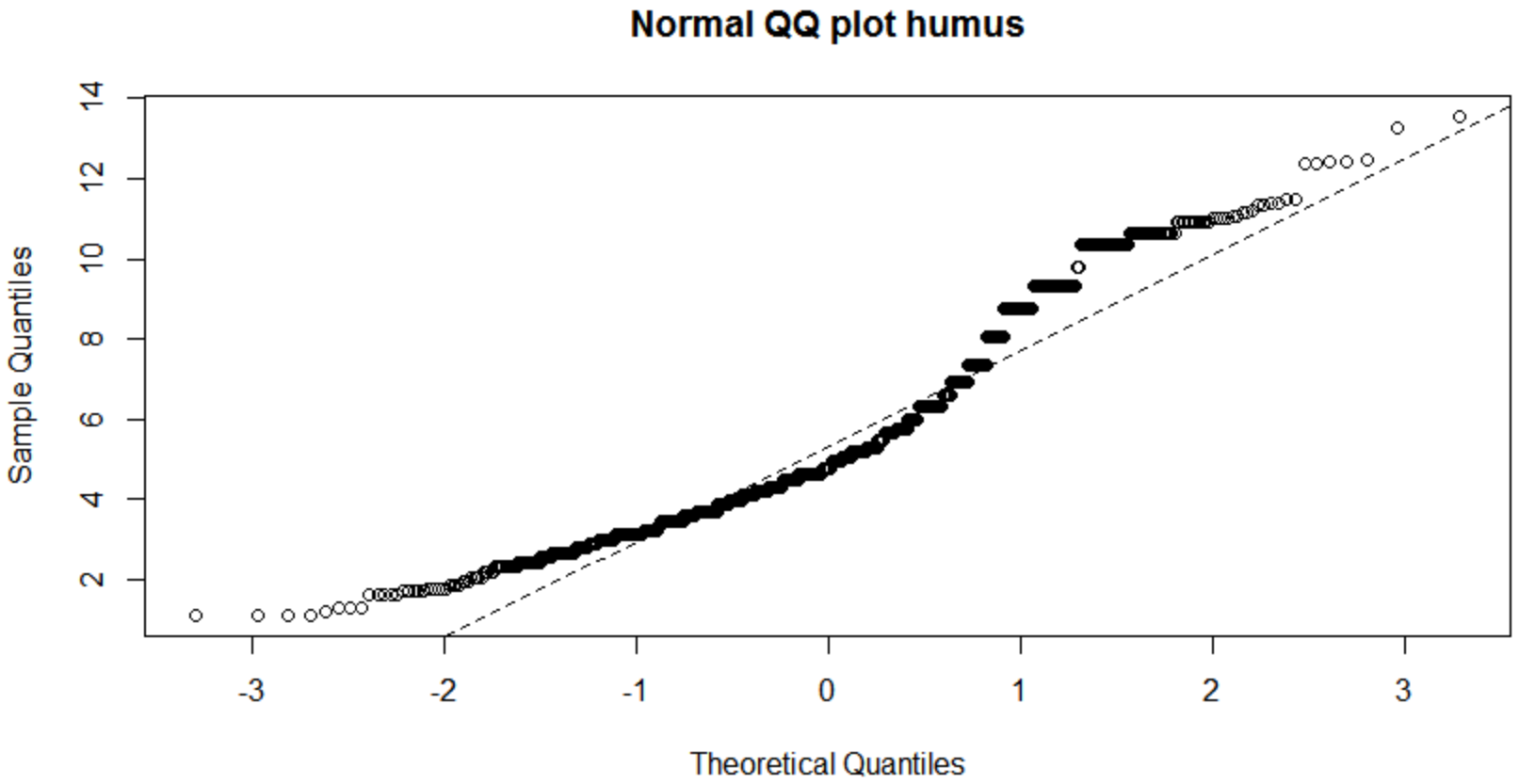
Deskriptivna statistika

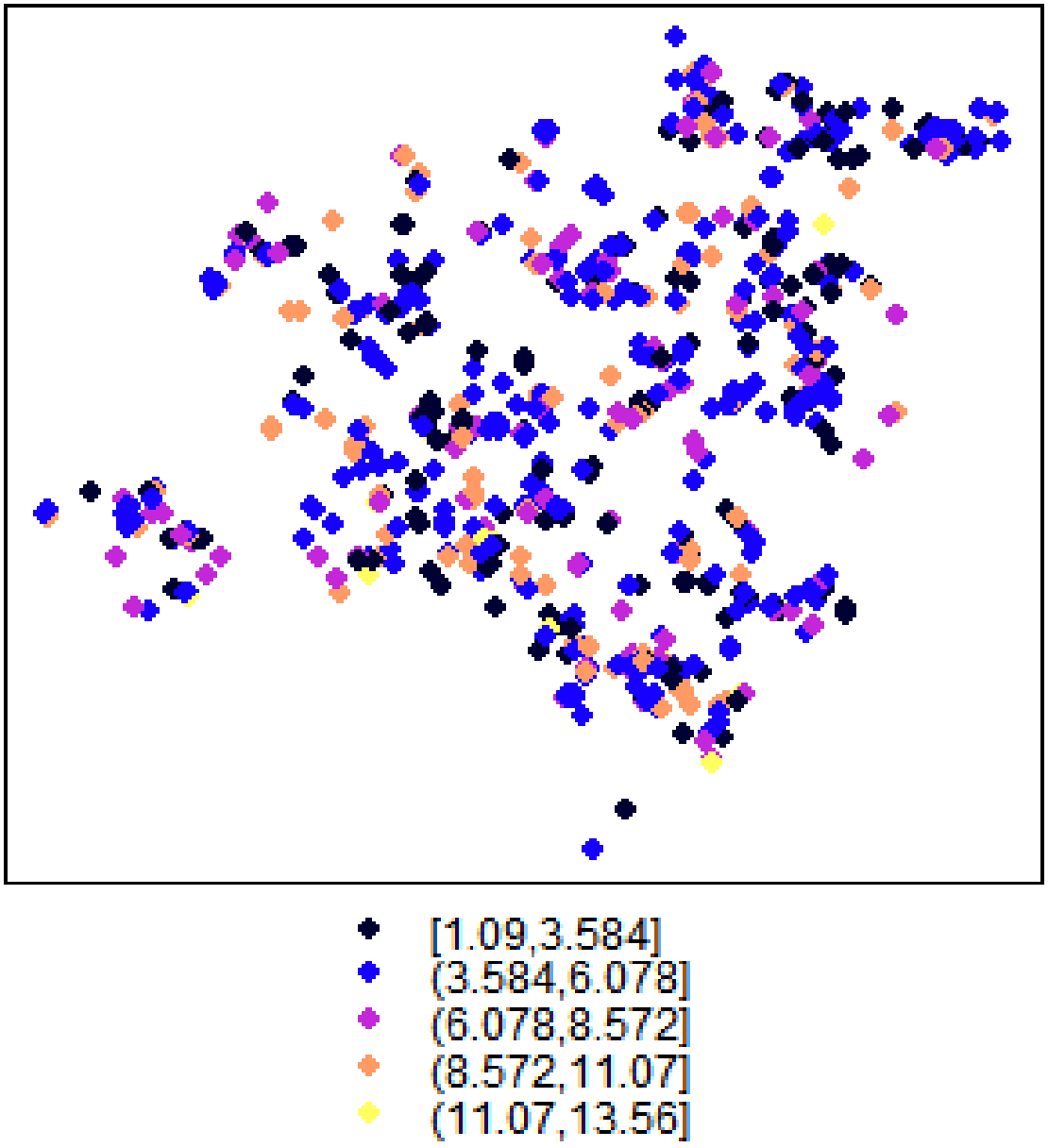
```
summary(podaci$humus)

 Min. 1st Qu. Median Mean 3rd Qu. Max.

 1.090 3.680 4.770 5.501 6.900 13.560
```

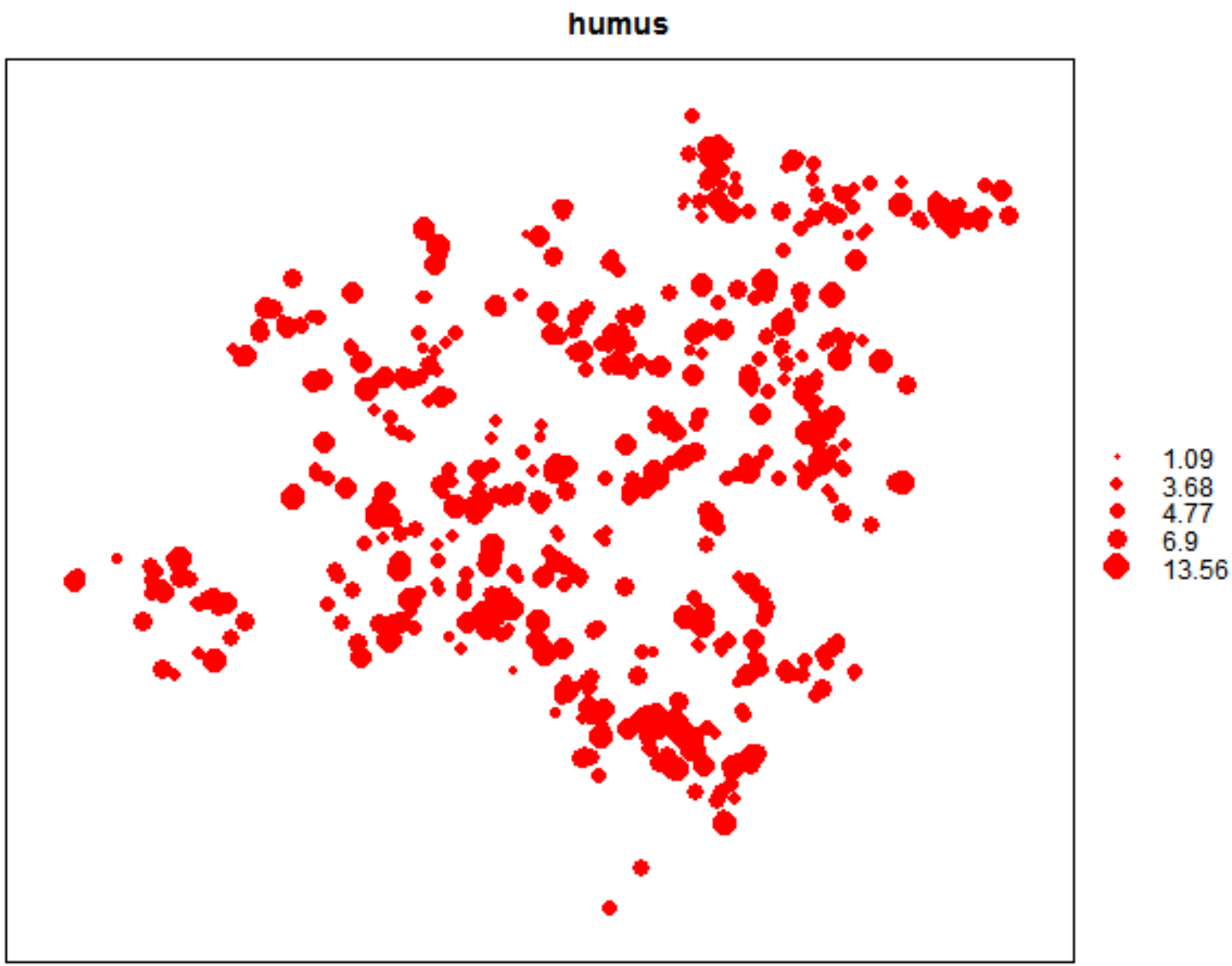
Deskriptivna statistika

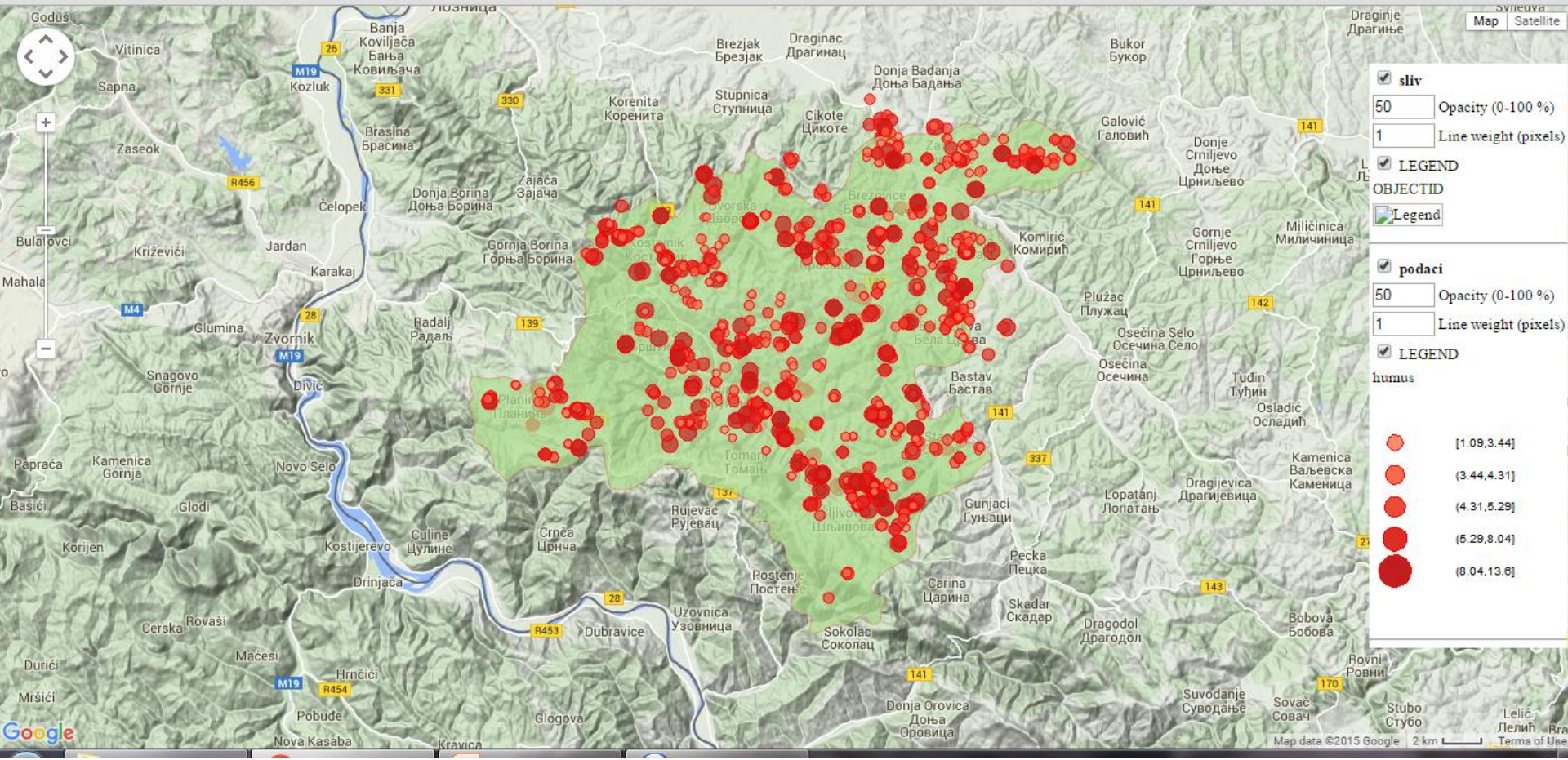


Deskriptivna statistika




[1.09,3.584]
(3.584,6.078]
(6.078,8.572]
(8.572,11.07]
(11.07,13.56]


humus
1.09
3.68
4.77
6.9
13.56


Map
Satellite
sliv
50
Opacity (0-100 %)
1
Line weight (pixels)
LEGEND
OBJECTID
Legend
podaci
50
Opacity (0-100 %)
1
Line weight (pixels)
LEGEND
humus
[1.09,3.44]
(3.44,4.31]
(4.31,5.29]
(5.29,8.04]
(8.04,13.6]
Map data ©2015 Google
2 km
Terms of Use

- Kriging je jedna od najpoznatijih intrepolacionih geostatističkih metoda.
- Naziv ove metode vezan je za ime D.G. Krige-a koji je prvi koristio prostornu korelaciju u valorizaciji rudarskih ležišta.
- Tehniku računanja ove metode postavio je francuski geomorfolog Georges Matheron, razvijajući teoriju regionalizovane promenljive.
- Matheron je ovu metodu razvio otprilike u isto vreme kada je prof. Moritz razvio kolokaciju metodom najmanjih kvadrata, krajem šezdesetih godina dvadesetog veka.
- Obe metode su bazirane na stohastičkim procesima i praktično daju iste razultate.

Stacionarnost

- Konstantna struktura na celom području od interesa.

For processes modelled using geostatistics, the stationary assumption is considered. The intrinsic stationary assumes that the observation Z can be decomposed into a mean and a residual part (Burrough, 1998; Hengl, 2009):

$$Z(\mathbf{s}) = m + \varepsilon'(\mathbf{s}) + \varepsilon' \tag{2.3}$$

where m is constant mean and $\varepsilon'(\mathbf{s})$ is the spatially correlated stochastic part of variation andε' uncorrelated stochastic component (pure noise) and:

$$\mathbb{E}(Z(\mathbf{s})) = m \tag{2.4}$$

and spatial dependency is defined by the variogram as:

$$\gamma(h) = \frac{1}{2}\mathbb{E}(Z(\mathbf{s}) - Z(\mathbf{s}+h))^2 \tag{2.5}$$

where$h$ is Euclidean distance $|\mathbf{h}|$, $\mathbb{E}$ denotes mathematical expectation and s is any location in the domain of interpolation.

Kriging je interpolaciona metoda uz koju se često vezuje skraćenica "***BLUE***", koja označava ***najbolju linearnu nepomerenu ocenu*** (best linear unbiased estimation).

- On je linearna ocena pošto u njegovim formulama figuriše linearna kombinacija merenih podataka.
- Kriging metoda je "nepomerena" ocena jer je očekivana srednja vrednost ocene grešaka jednaka 0.
- Najbolja jer nastoji da minimalizuje disperziju grešaka.

Semivariance

Poluvarijansa

Poluvarijansa

- Teorija regionalizovane promenljive koristi ***poluvarijansu*** da bi iskazala stepen uzajamne zavisnosti između između tačaka (uzoraka) na nekoj površi.

- **Poluvarijansa** je polovina vrednosti **varijanse** razlika vrednosti između svih uzoraka na konstantnom rastojanju.

Poluvarijansa je mera stepena prostorne zavisnosti između uzoraka

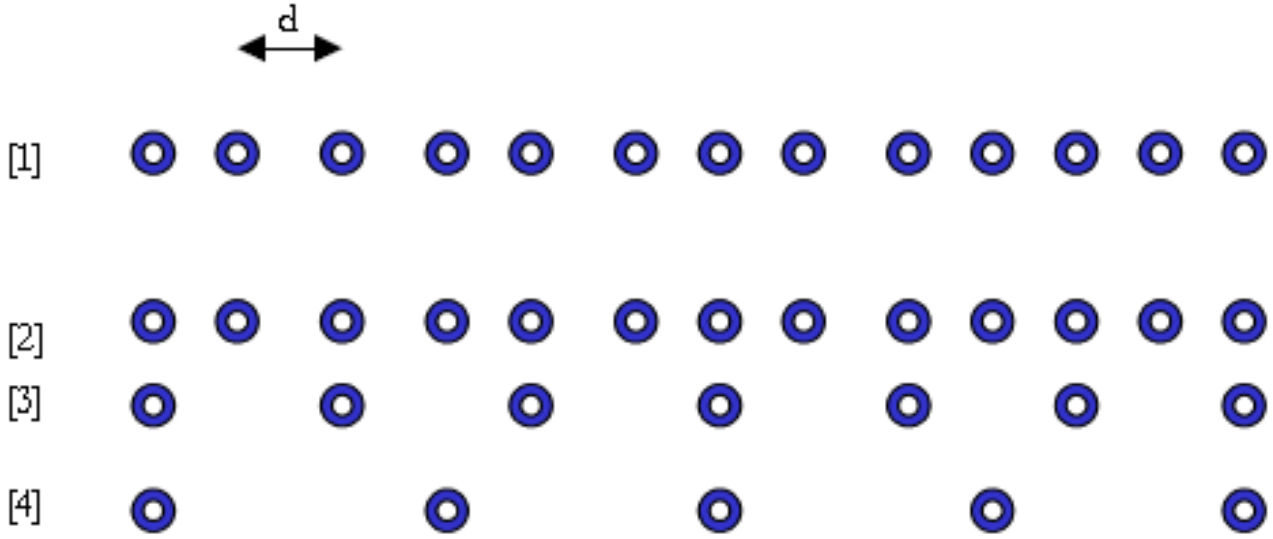
Poluvarijansa

- *Poluvarijansa:* Vrednost poluvarijanse između tačaka zavisi od rastojanja između njih. Manje rastojanje rezultira manjom vrednošću poluvarijanse, a veće rastojanje većom vrenošću.

Računanje poluvarijanse
(Između tačaka na pravilnom rastojanju)

- Razmotrimo uzorke pravilno udaljene (d), poluvarijansa se može odrediti za sva rastojanja koja su umnožak koraka (d):

$$\gamma(h) = \frac{1}{2N_h} \sum_{i=1}^{N_h} (z_i - z_{i+h})^2$$



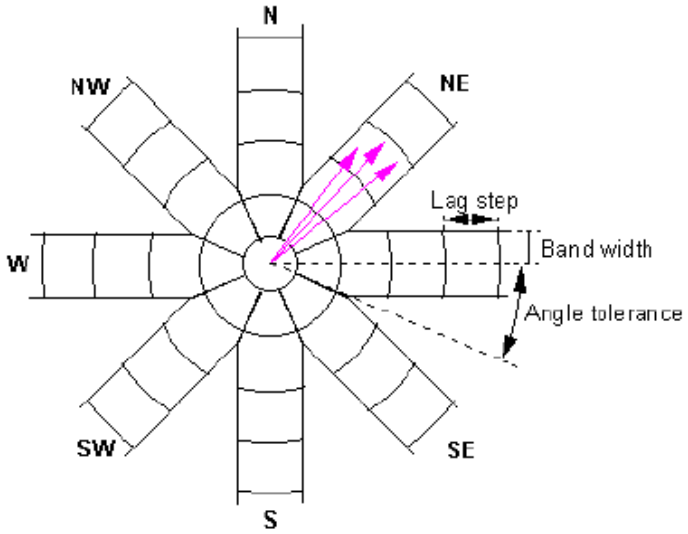
Poluvarijansa

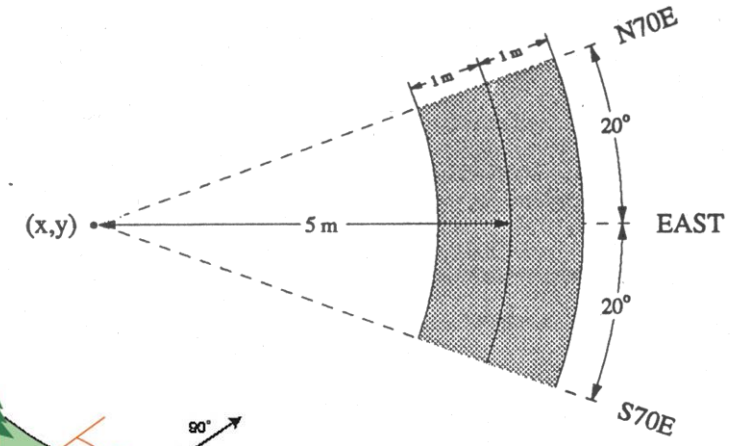
$$\gamma(h) = \frac{1}{2N_h} \sum_{i=1}^{N_h} (z_i - z_{i+h})^2$$

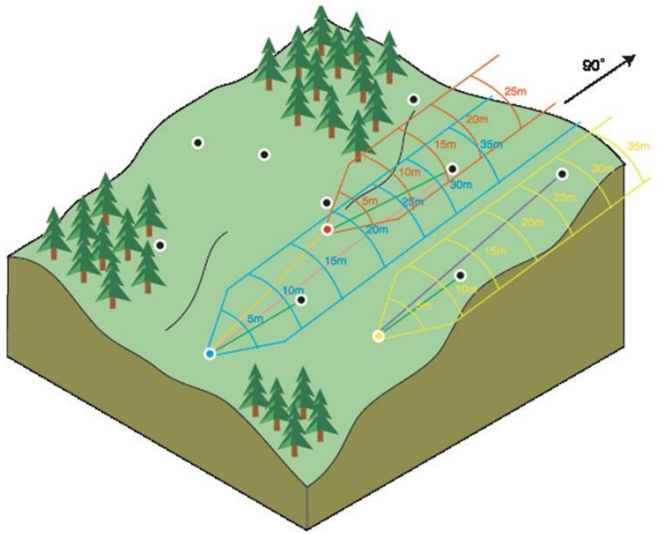
- Z_i izmerena vrednost regionalizovane promenljive na lokaciji ***i*** ,
- Z_{i+h} je druga vrednost izmerena na lokaciji i+***h*** (h=umnožak rastojanja ***d)***
- N_h broj parova unutar intervala ***h***

Računanje poluvarijanse
(Između tačaka na nepravilnom rastojanju)

- U praksi je veom teško naići na tačke koje se nalaze na striktnom rastojanju [d].
- Zato se rastojanja grupišu u klase.



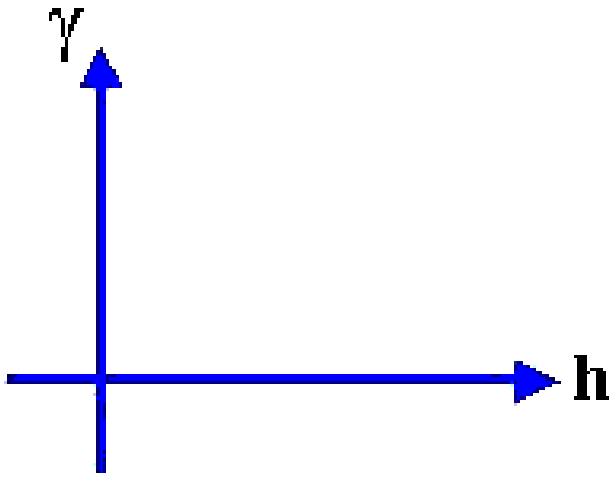




Variogram

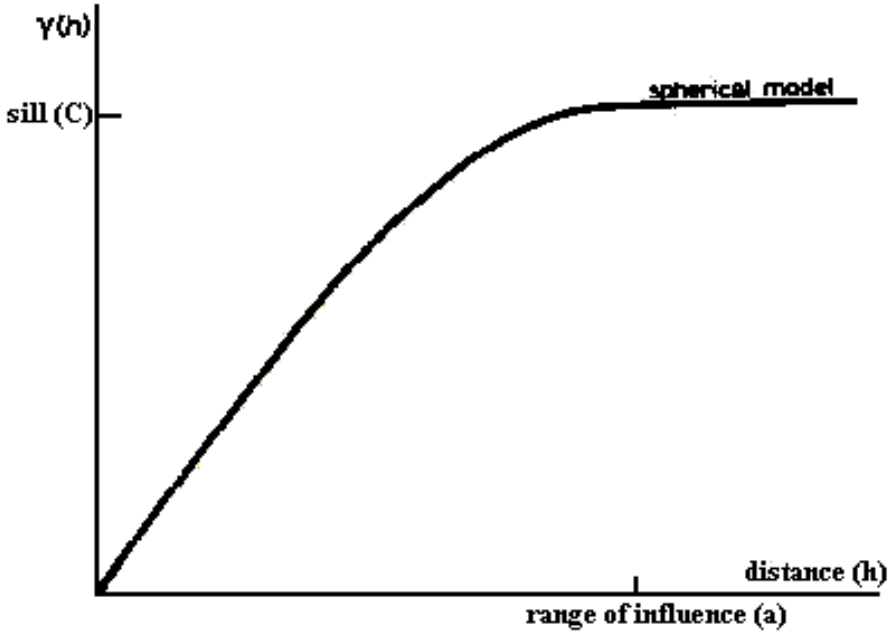
Variogram

- Predstavlja grafik poluvarijansi u funkciji od rastojanja između uzoraka.

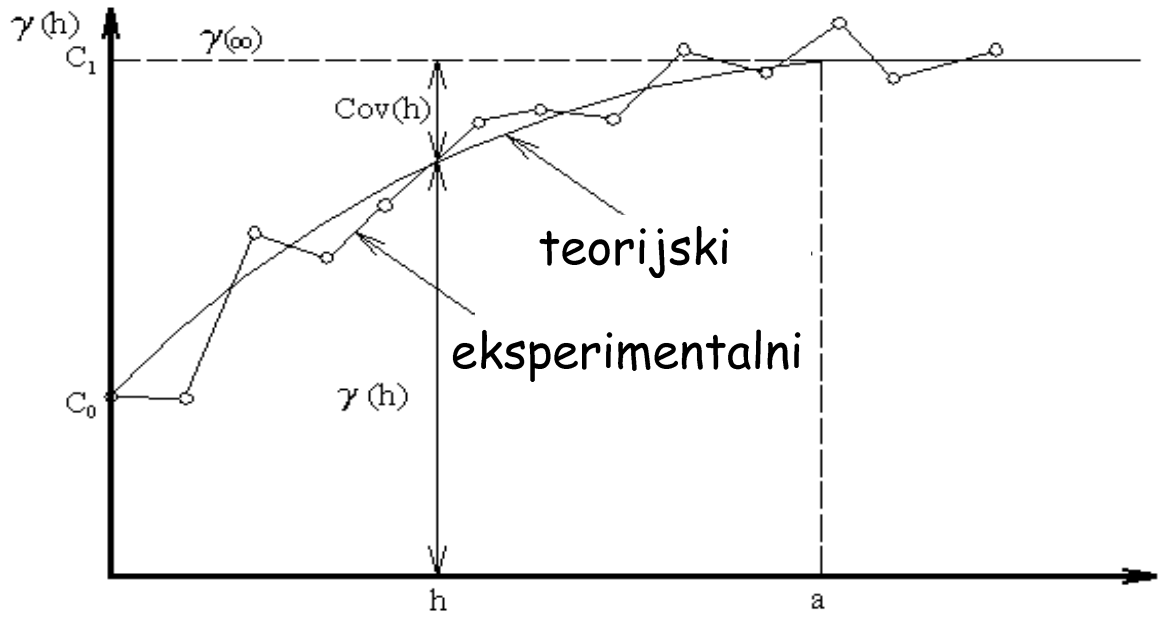


Variogram

- Poluvarijansa na rastojanju d = 0 trebala bi da bude jednaka nuli, pošto bi to teorijski bila razlika vrednosti promenljive u istoj tački. Odnosno jednaka varijansi merenja.
- Kako su tačke udaljenije jedna od druge, tako bi i vrednost poluvarijanse trebalo da raste.



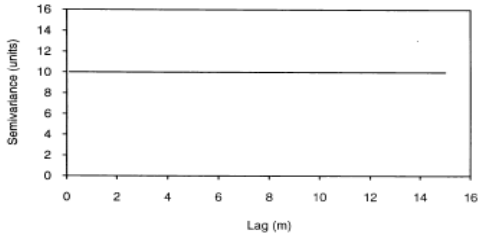
Modeliranje variograma



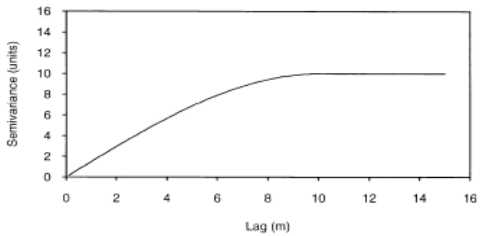
$$\gamma(h) = \frac{1}{2N_h}\sum_{i=1}^{N_h}(z_i - z_{i+h})^2$$

Modeli variograma

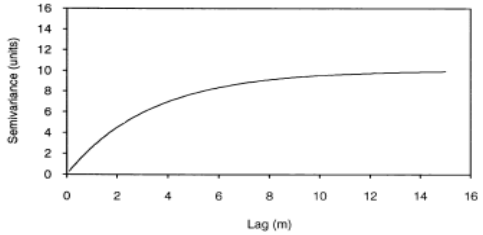
Nugget model



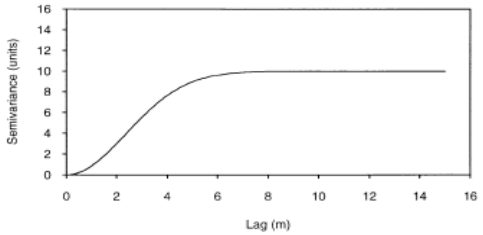
Spherical model



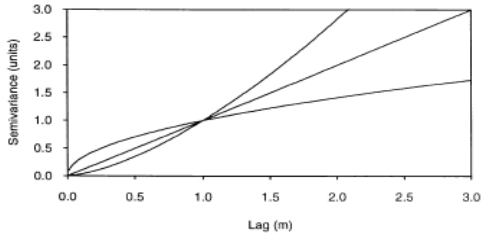
Exponential model



Gaussian model



Power model

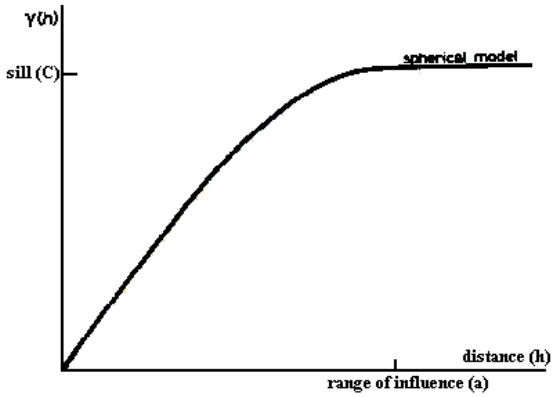


Modeli variogram

Model	Formula	Primedbe
Nugget effect	$\gamma(0) = C_0$	Variogram je konstantna veličina.
Linear	$\gamma(h) = C_1 h$	Nema praga. Obično se kombinuje sa drugim fukcijama. Često se koristi kao funkcija nagiba.
Exponential Exp()	$\gamma(h) = C_1 (1 - e^{-kh})$	k je konstanta, obično k=1 ili k=3. Koristi se kod variograma sa velikim grumenskim efektom, koji polako rastu do vrednosti praga.
Spherical Sph()	$\gamma(h) = C_1 \left(\frac{3h}{2} - \frac{1}{2} h^3 \right), h < 1$ $\gamma(h) = C_1, h \geq 1$	Koriste se u slučaju kada je grumenski efekat mali ali je bitan. Često se pojavljuje kao ***default*** model kod mnogih programskih paketa.

Variogrami (Modeli)

- **Sferni (spherical) model.**
- ***a*** domet variograma.
- ***C*** je prag.

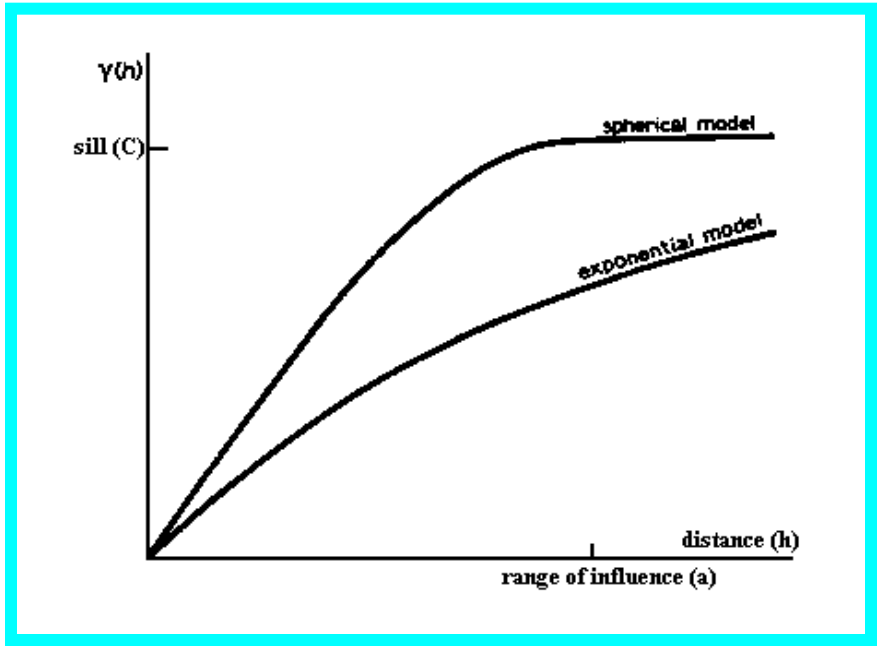


$$\gamma(h) = \begin{cases} C\left(\frac{3}{2}\frac{h}{a} - \frac{1}{2}\frac{h^3}{a^3}\right) & \text{gde je } h \le a \\ C & \text{gde je } h \ge a \end{cases}$$

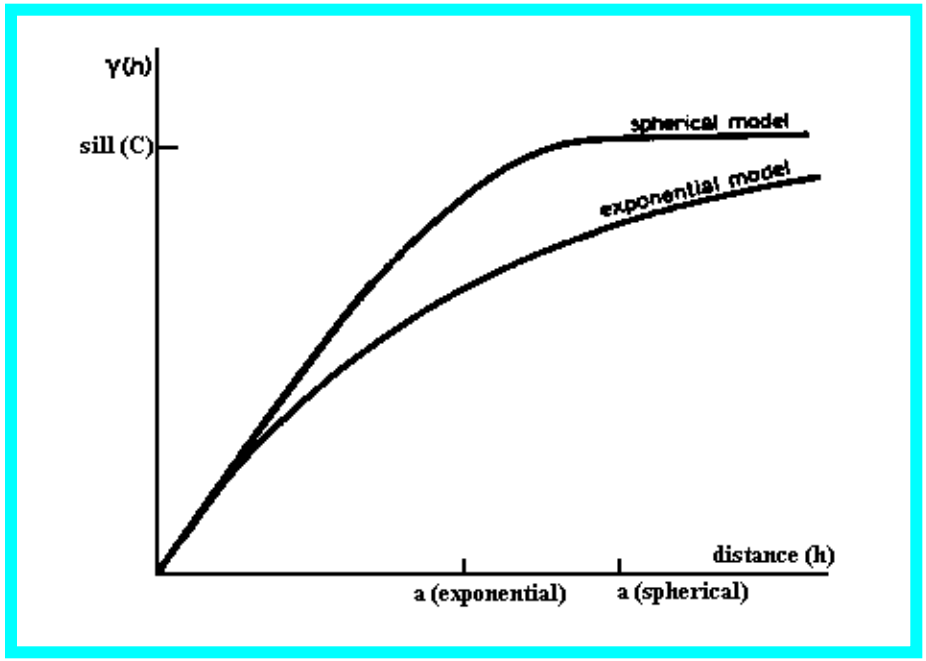
Variogrami (Modeli)

Eksponencijalni (Exponential) Model

$$\gamma(h) = C\left(1 - e^{-h/a}\right)$$



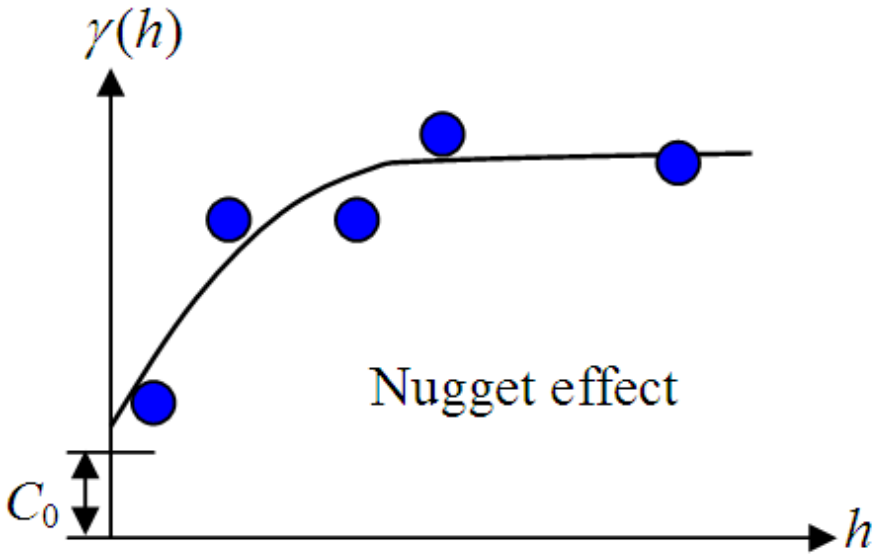
sferni i eksponencijalni sa istim pragom i dometom

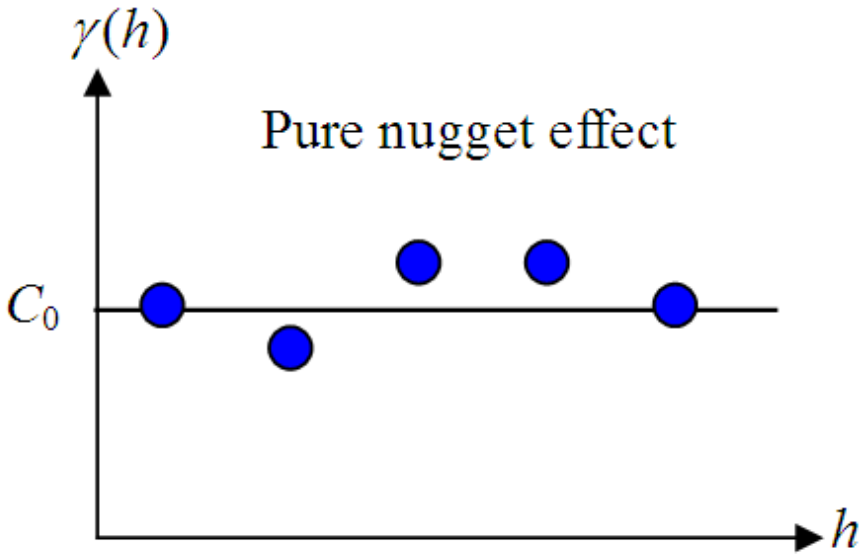


sferni i eksponencijalni sa istim pragom i istim početnim nagibom

Variogrami (Modeli)

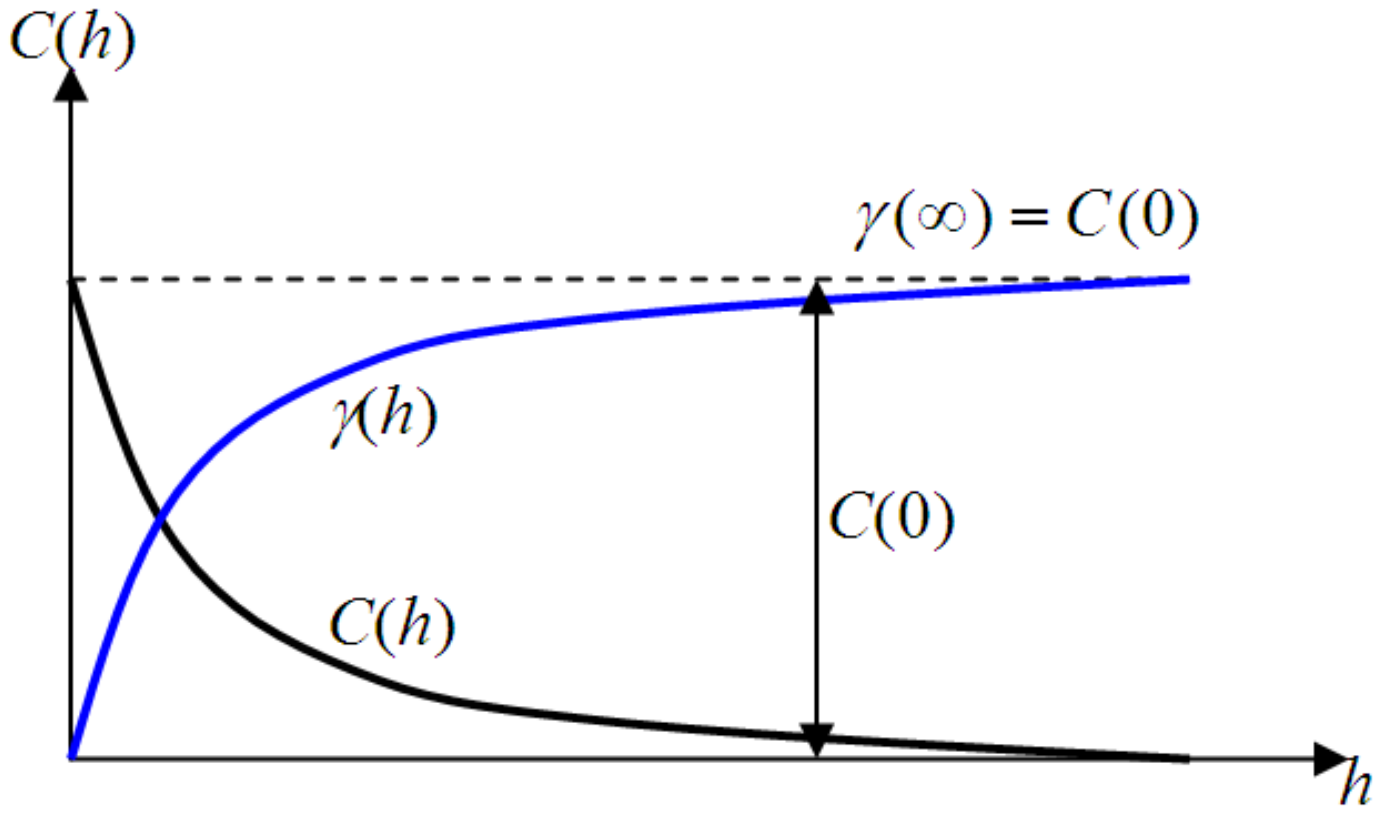
Grumenski (nugget) model





Odnos između kovarijanse i variograma u slučaju stacionarnih promenljivih

$$\gamma(h) = C(0) - C(h)$$



Kriging interpolacija

Kriging Interpolacija

- Ovaj postupak uključuje i grešku merenja i nesigurnost prilikom procene vrednosti.
- Kriging interpolacija u svakoj vrednosti koju procenjuje pridružuje i standardnu grešku ili varijansu.

- Bazira se određivanju optimalnih težinskih koeficijenata koji se pridružuju poznatim (uzorkovanim) vrednostima na osnovu poznavanja variograma. S obzirom da su variogrami u funkciji rastojanja, težine će zavisti od distribucije uzorkovanih tačaka.

Ordinary (obični) Kriging

Ordinary Kriging

- Ordinary kriging je najjednostavnija forma kriging-a.

- Kod Ordinary kriging-a, predpostavlja se da je regionalizovana promenljiva stacionarna.

Ordinary Kriging

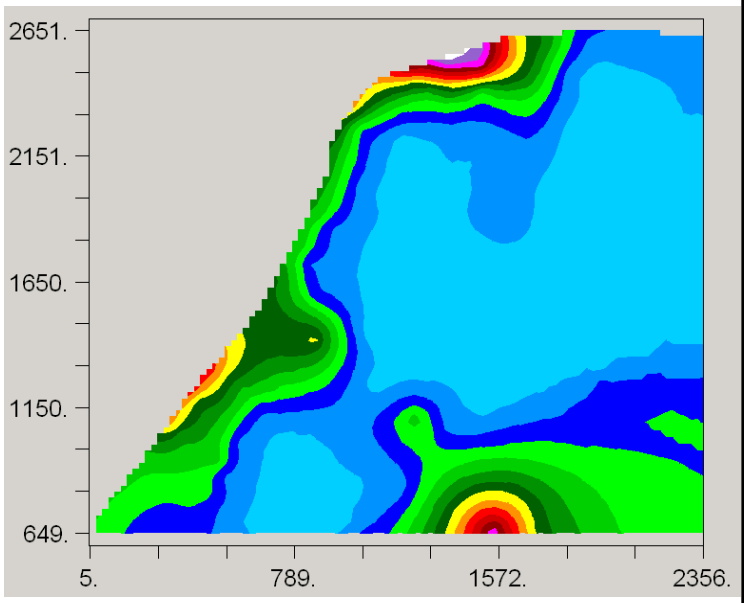
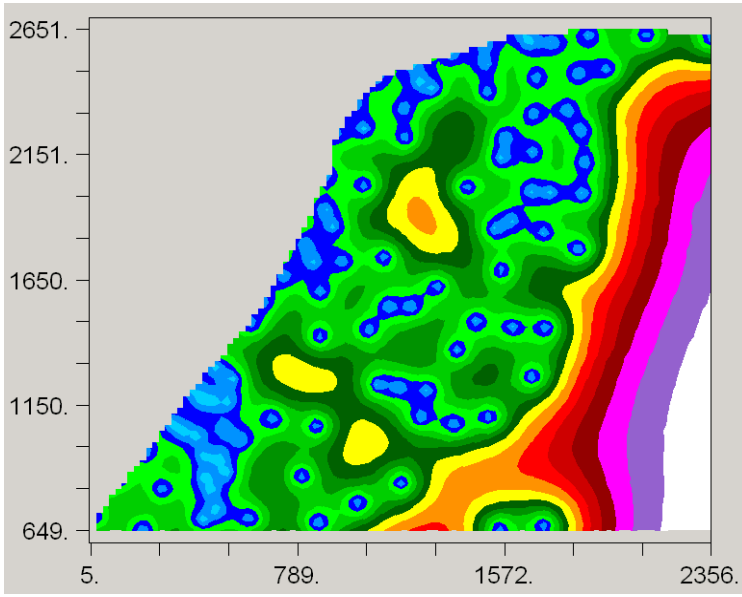
- U našem slučaju Ze (p) procenjena vrednost Z u tački p, računa se kao linearna kombinacija uzoraka kojima je pridružena težina:

$$z_e(p) = \sum w_i \cdot z(p_i)$$

- This estimated value will most likely differ from the actual value at point p, Za(p), and this difference is called the ***estimation error***:

$$\varepsilon_p = z_e(p) - z_a(p)$$

Ordinary Kriging

Zinc data: Predicted values	Zinc data: Estimated standard deviation
	

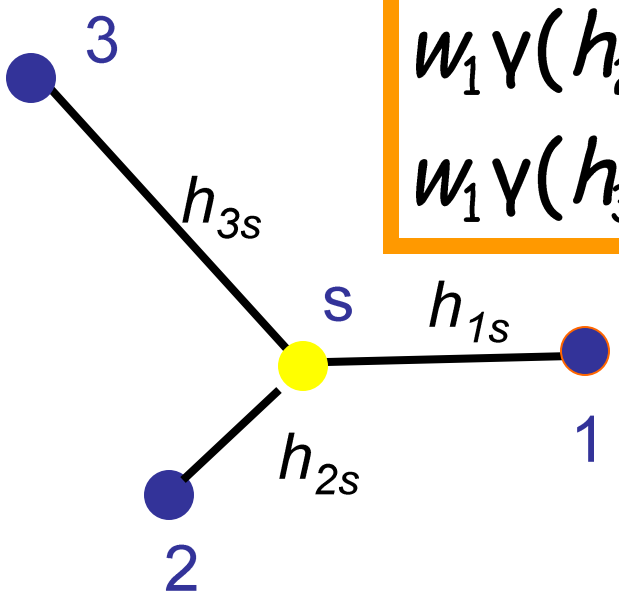
- Ukoliko ne postoji trend i suma težina je jednaka 1 (jedan), onda je ocena nepomerena **(unbiased)**.
- Rasipanje ocena oko tačne vrednosti se naziva greška ili varijansa ocene (**estimation variance**).

$$\sigma_z^2 = \frac{\sum_{i=1}^{n} [z_e(p_i) - z_a(p_i)]_i^2}{n}$$

Ordinary Kriging

- Kriging nastoji da izabere optimalne težine kako bi dobili minimalnu grešku procene.
- Optimalne težine, koje daju nepomerenu ocenu i minimalnu grešku procene, dobijaju se rešavanjem sledećeg sistema jednačina.
- Koliko je jednačina?.

$$
\begin{aligned}
w_1\gamma(h_{11}) + w_2\gamma(h_{12}) + w_3\gamma(h_{13}) &= \gamma(h_{1p}) \\
w_1\gamma(h_{21}) + w_2\gamma(h_{22}) + w_3\gamma(h_{23}) &= \gamma(h_{2p}) \\
w_1\gamma(h_{31}) + w_2\gamma(h_{32}) + w_3\gamma(h_{33}) &= \gamma(h_{3p})
\end{aligned}
$$



$$w_1 + w_2 + w_3 = 1$$

Ordinary Kriging

- Uvodi se pomoćna promenljiva poznata kao **Lagrange multiplier**, tako da je broj nepoznatih jednak broju jednačina

$$w_1\gamma(h_{11})+w_2\gamma(h_{12})+w_3\gamma(h_{13})+\lambda=\gamma(h_{1p})$$
$$w_1\gamma(h_{21})+w_2\gamma(h_{22})+w_3\gamma(h_{23})+\lambda=\gamma(h_{2p})$$
$$w_1\gamma(h_{31})+w_2\gamma(h_{32})+w_3\gamma(h_{33})+\lambda=\gamma(h_{3p})$$
$$w_1+w_2+w_3=1$$

$$\begin{bmatrix} \gamma(h_{11}) & \gamma(h_{12}) & \gamma(h_{13}) & 1 \\ \gamma(h_{21}) & \gamma(h_{21}) & \gamma(h_{23}) & 1 \\ \gamma(h_{31}) & \gamma(h_{32}) & \gamma(h_{33}) & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \lambda \end{bmatrix} = \begin{bmatrix} \gamma(h_{1p}) \\ \gamma(h_{2p}) \\ \gamma(h_{3p}) \\ 1 \end{bmatrix}$$

Ordinary Kriging

- Kada se odrede težine, ocenjena vrednost je:

$$z_e(p) = w_1 z_1 + w_2 z_2 + w_3 z_3$$

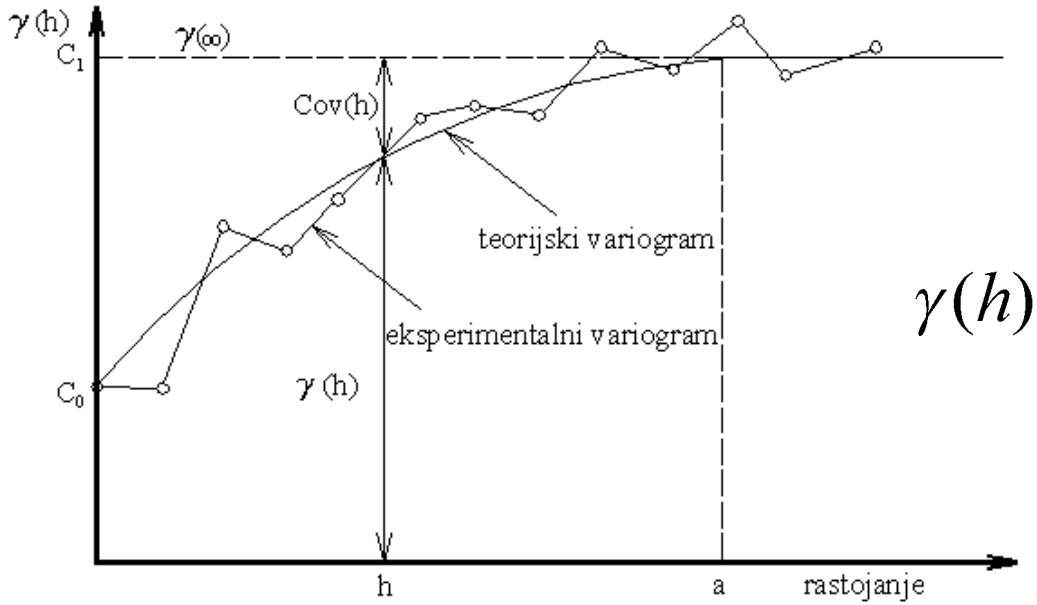
- Ocena greške se računa kao:

$$\sigma_z^2 = w_1 \gamma(h_{1p}) + w_2 \gamma(h_{21p}) + w_{13} \gamma(h_{3p}) + \lambda$$

- Uslovi koji zadovoljavaju hipotezu o prirođenosti (intrisic hypotesis) su:

$$E\left[V\ \ x+h\ -V\ \ x\ \right]=m\ \ h$$

$$Var\left[V\ \ x+h\ -V\ \ x\ \right]=2\gamma\ \ h$$



$$\gamma(h)=\frac{1}{2N(h)}\sum_{(i,j)\left|h_{ij}\approx h\right.}(v_i-v_j)^2$$

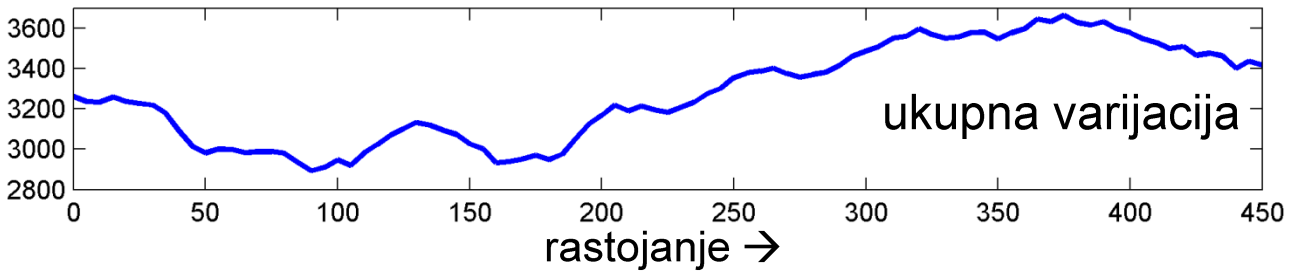
Kriging princip

Vrednosti Z na lokaciji s modeluju se kao suma:

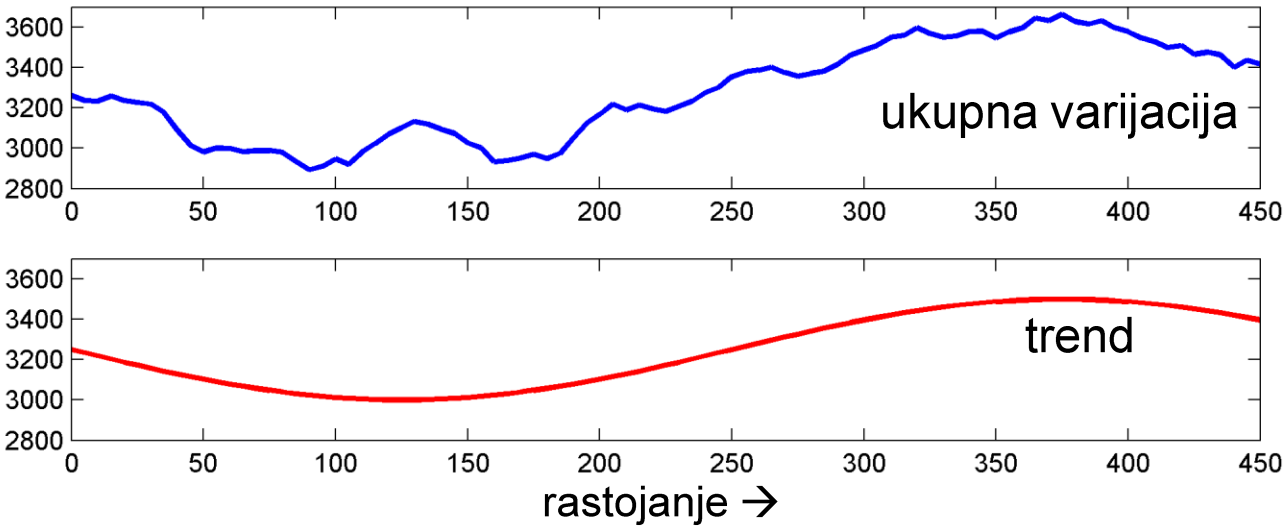
- trenda (m)
- prostorne autokorelacije (ε')
- Greške sa normalnom raspodelom (Gaussian error (ε''))

$$Z(s) = m(s) + \varepsilon'(s) + \varepsilon''(s)$$

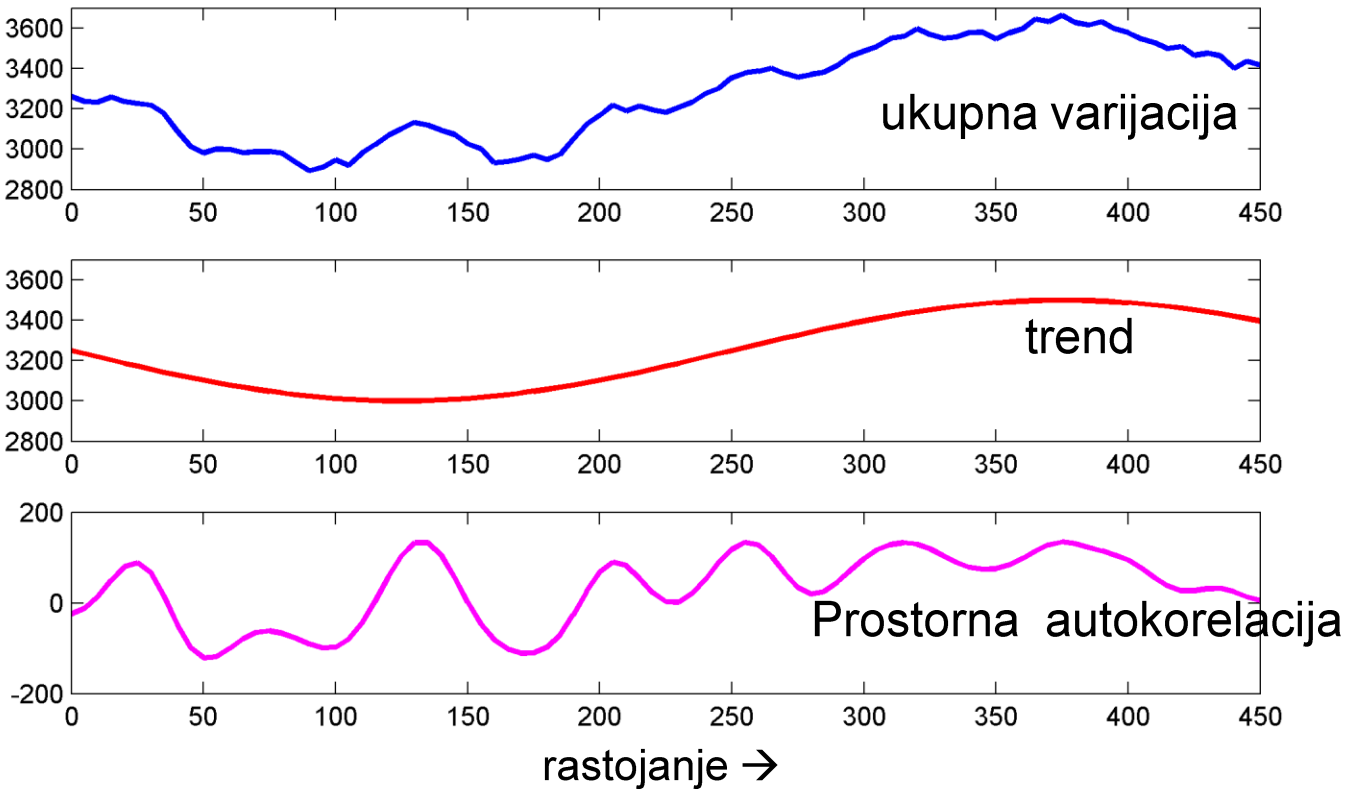
Kriging princip



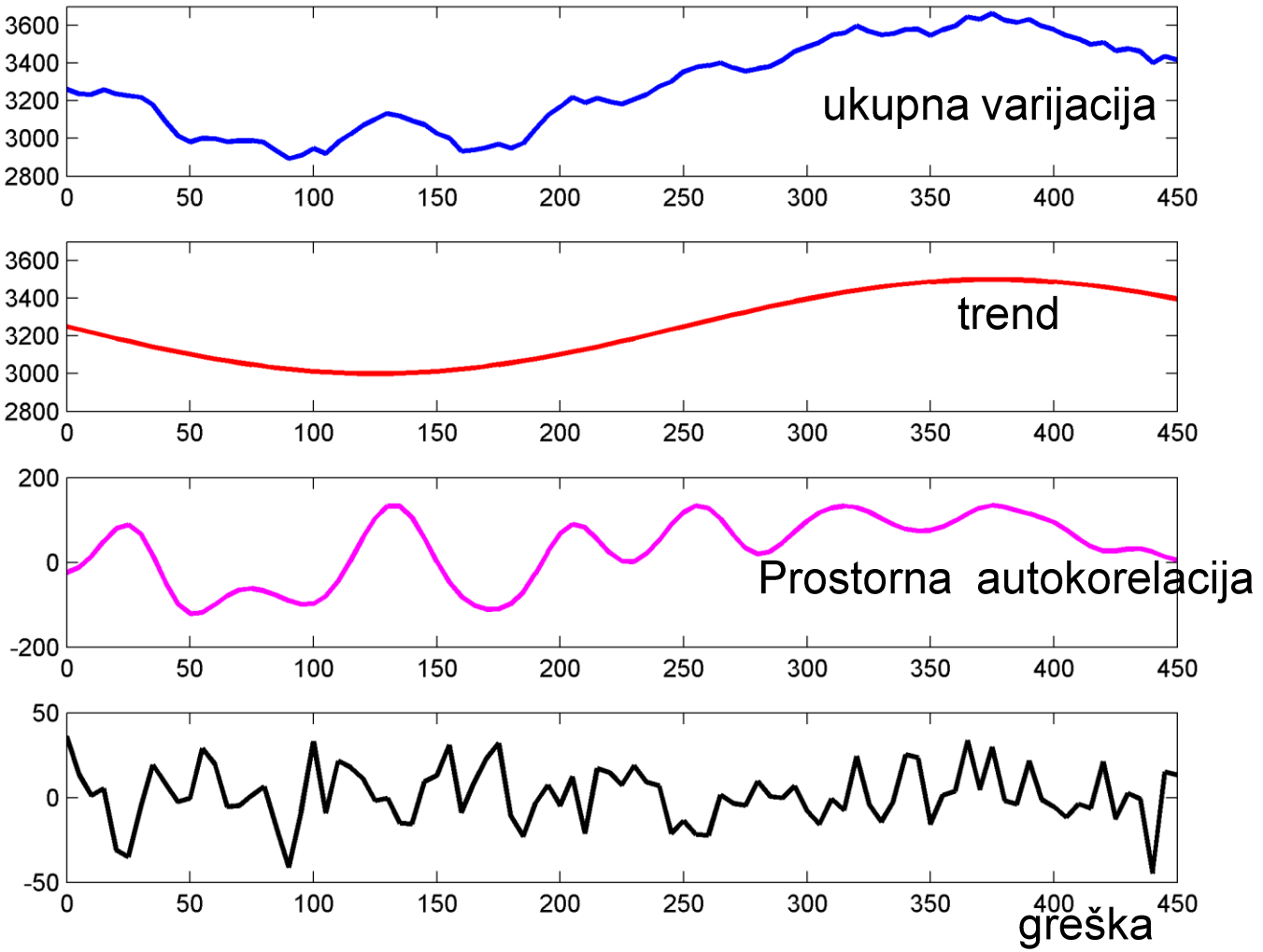
Kriging princip



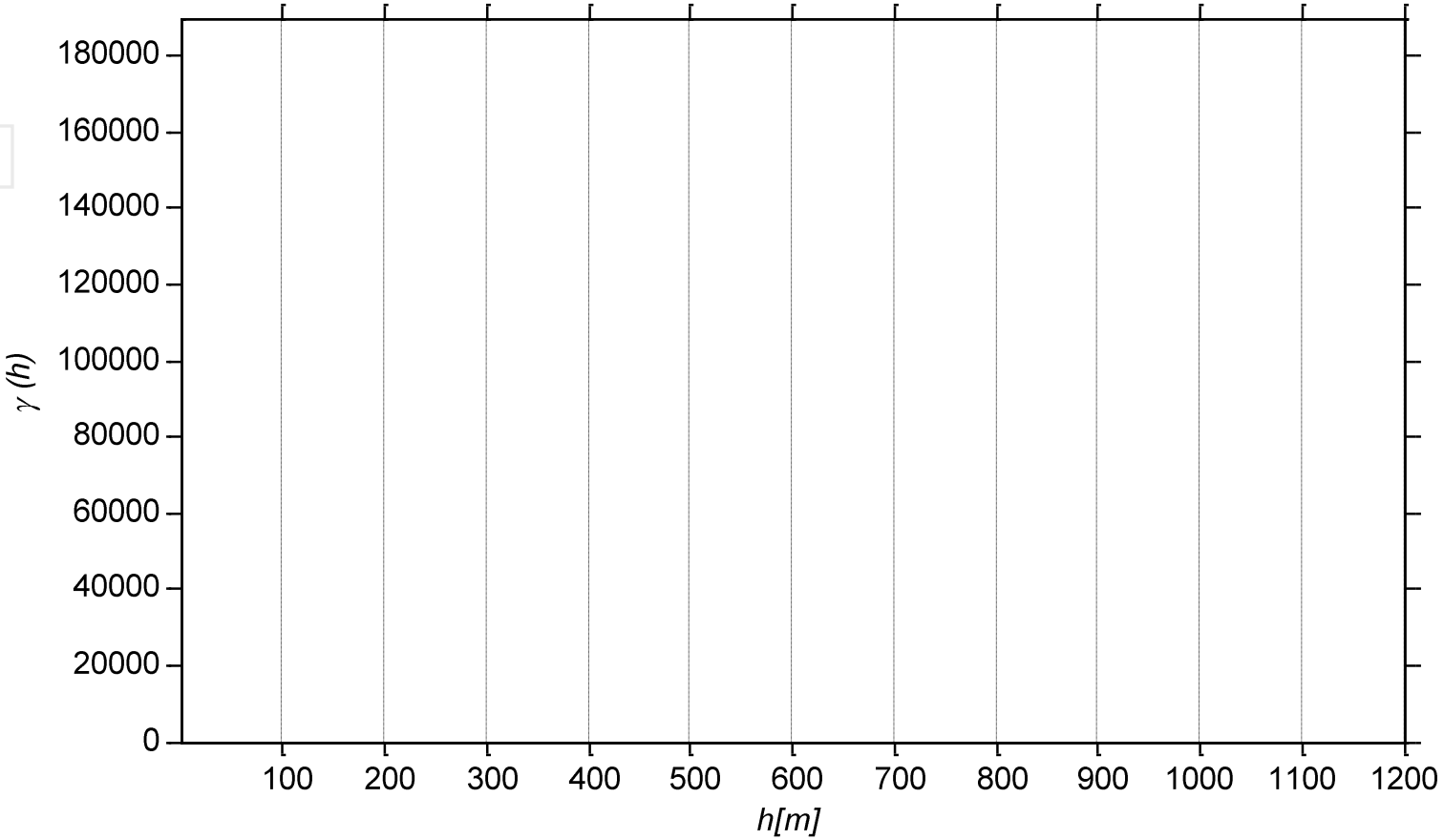
Kriging princip



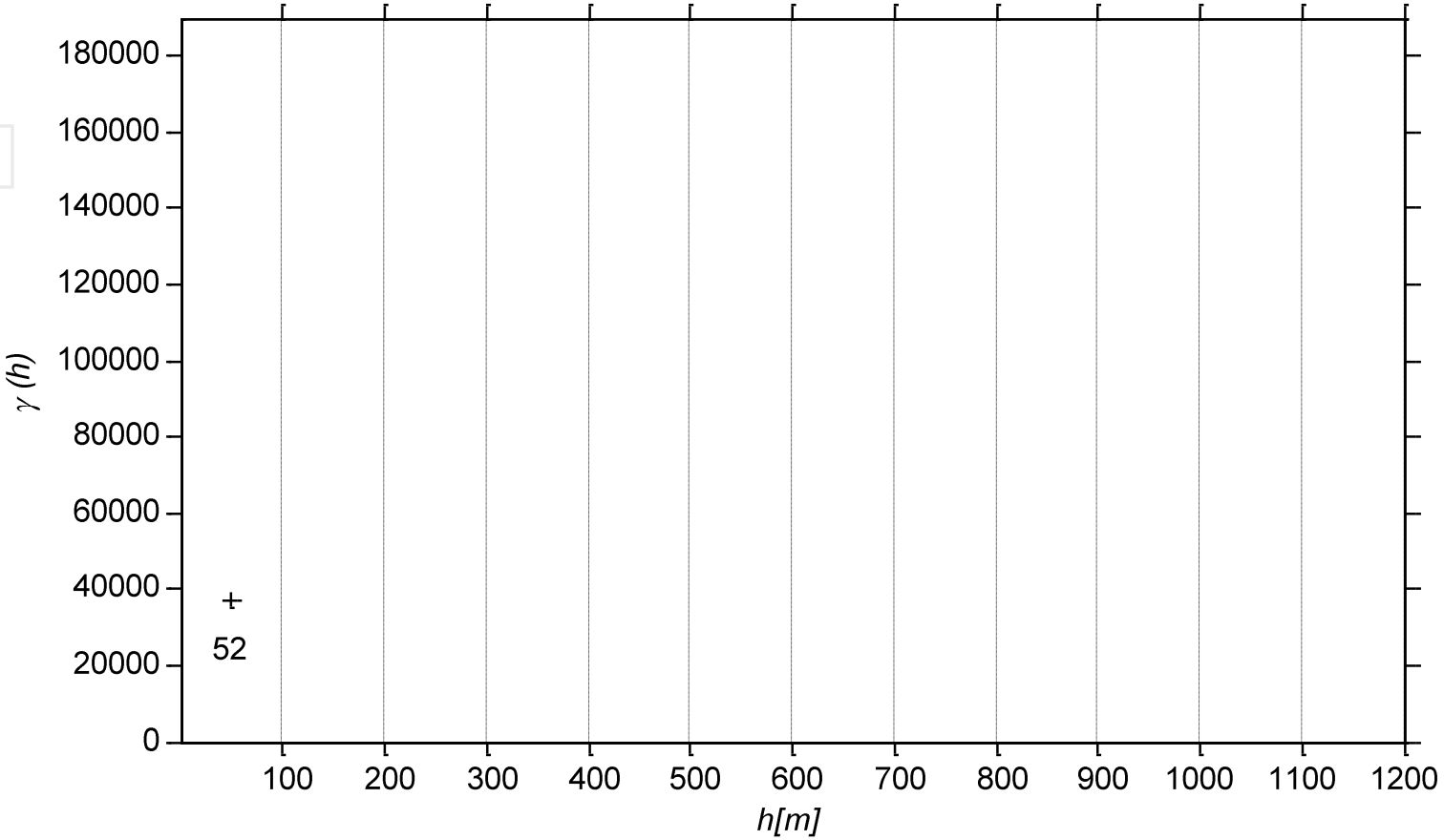
Kriging princip



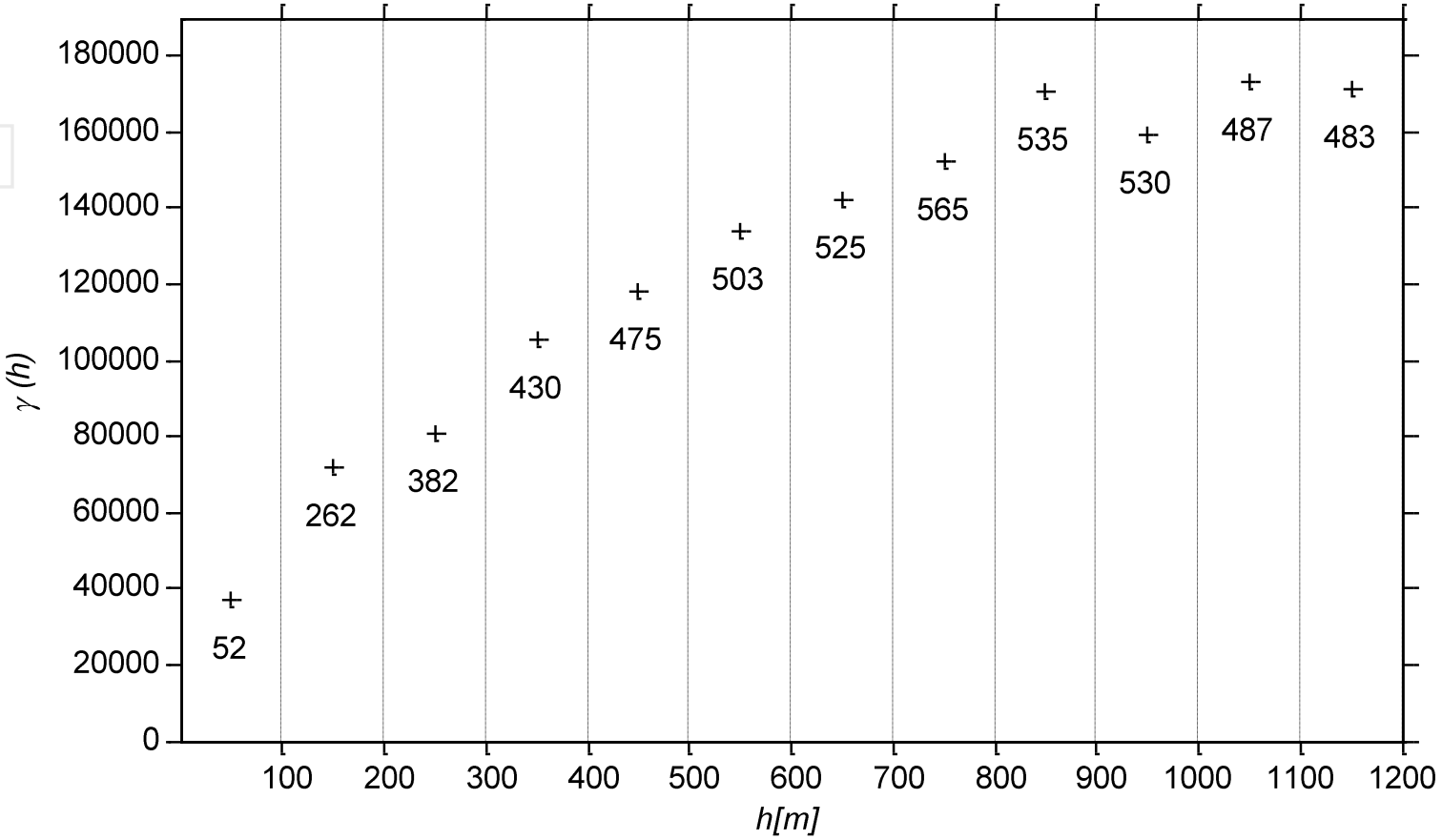
Eksperimentalni variogram



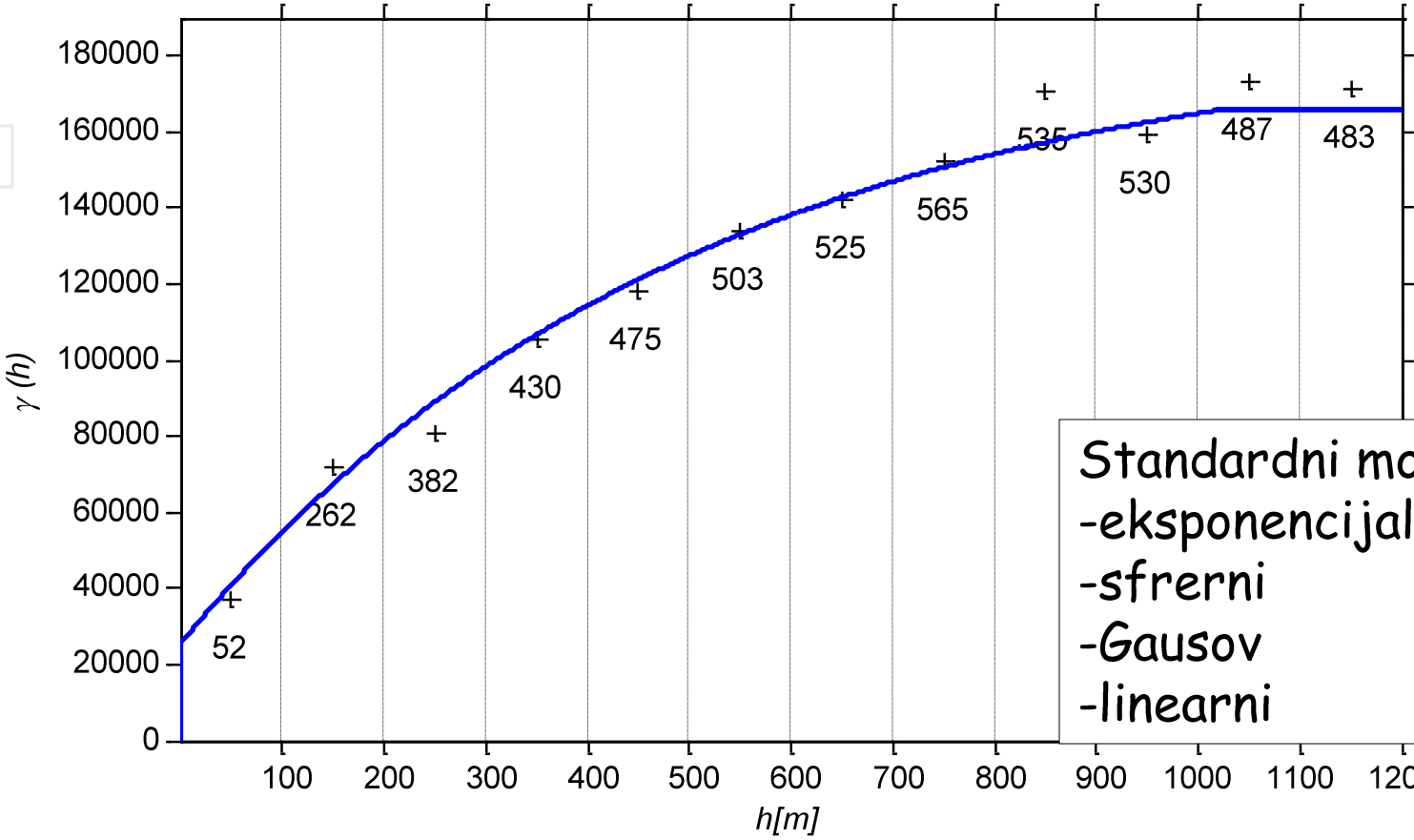
Eksperimentalni variogram



Eksperimentalni variogram

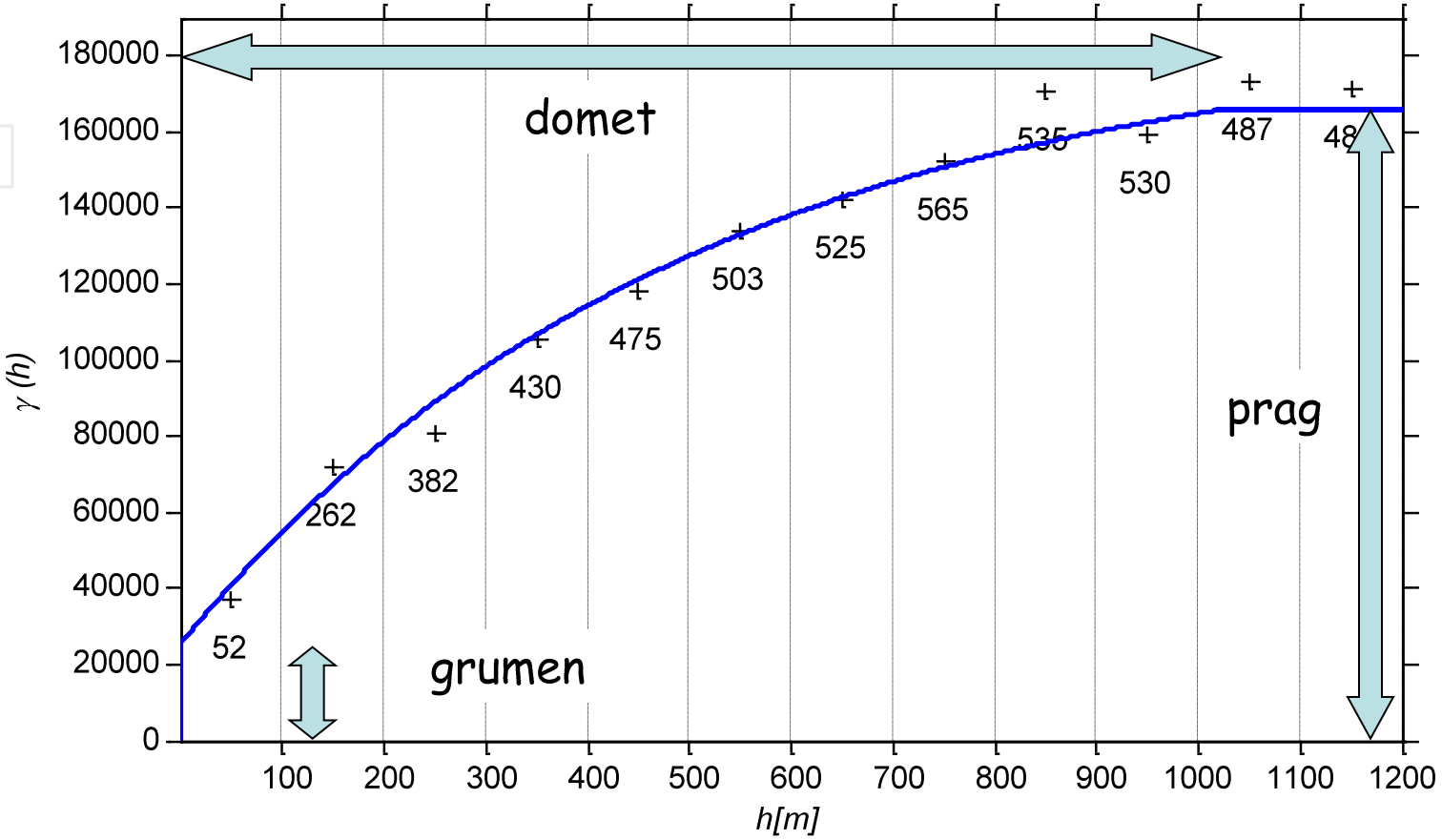


Eksperimentalni variogram

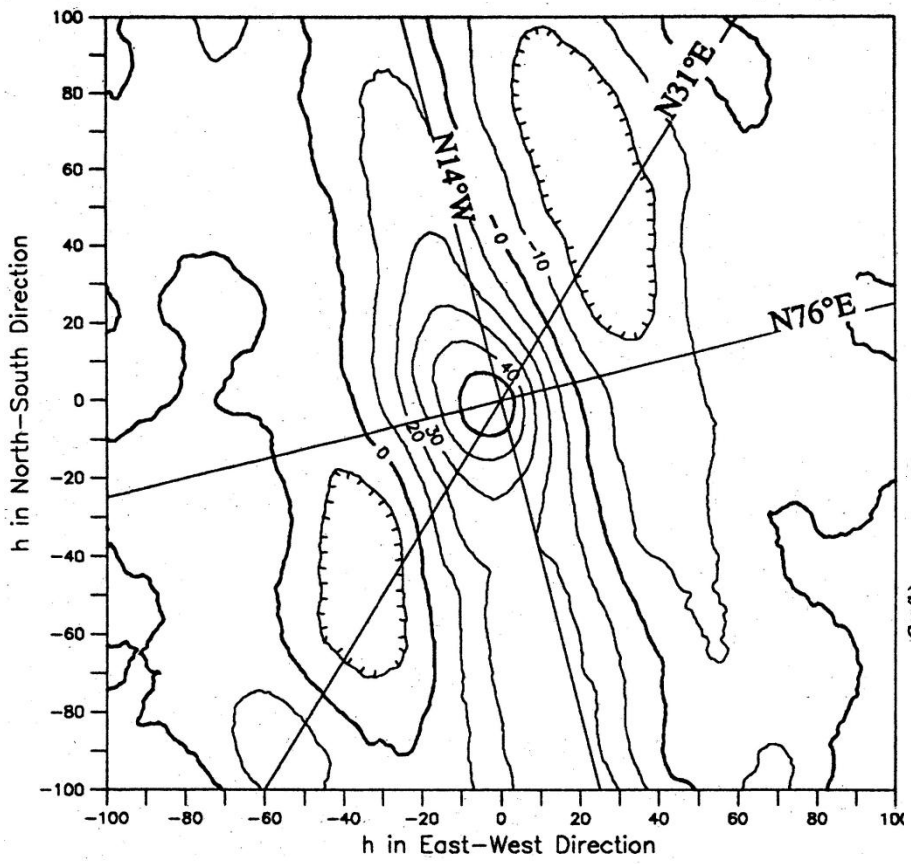


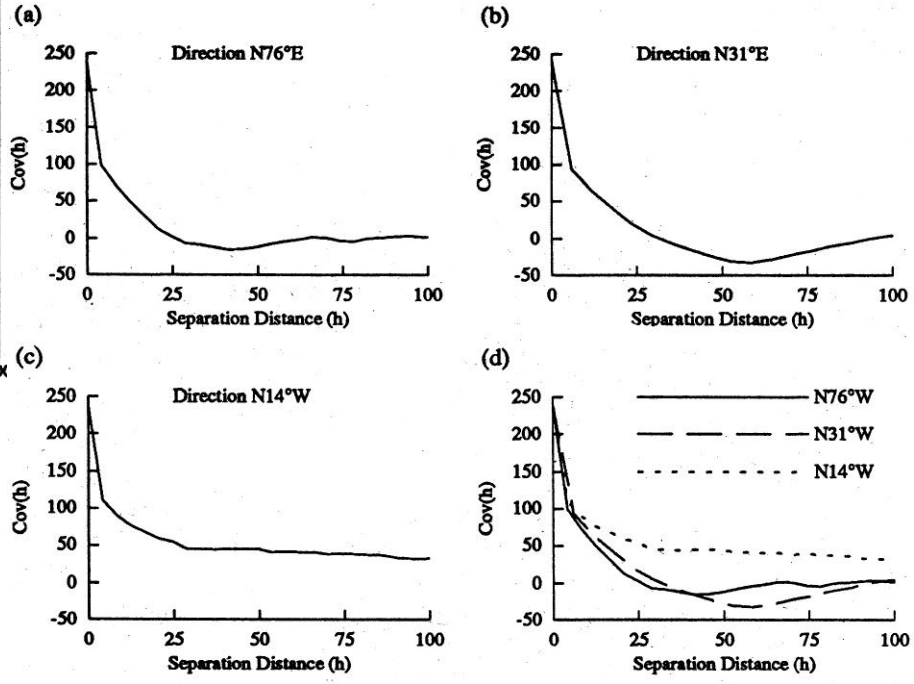
Standardni modeli:
-eksponencijalni
-sfrerni
-Gausov
-linearni

Eksperimentalni variogram



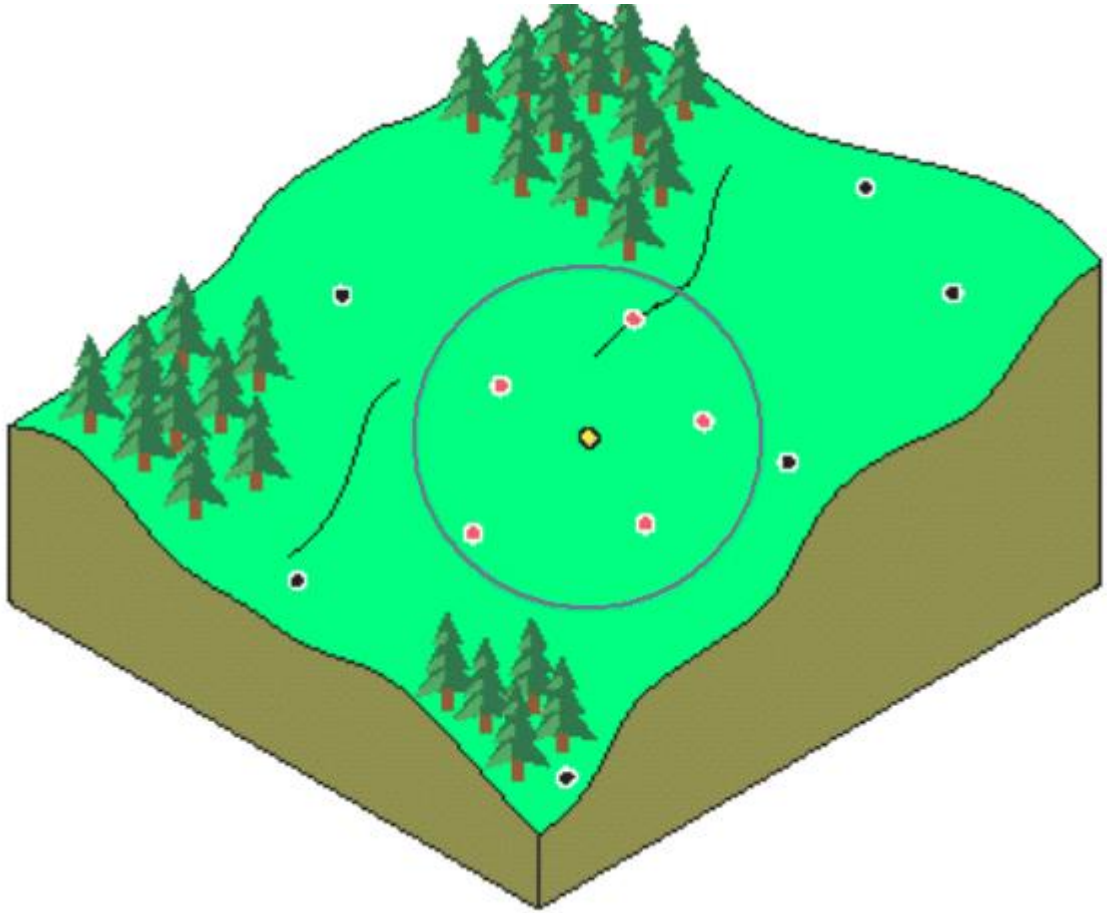
Anizotropija





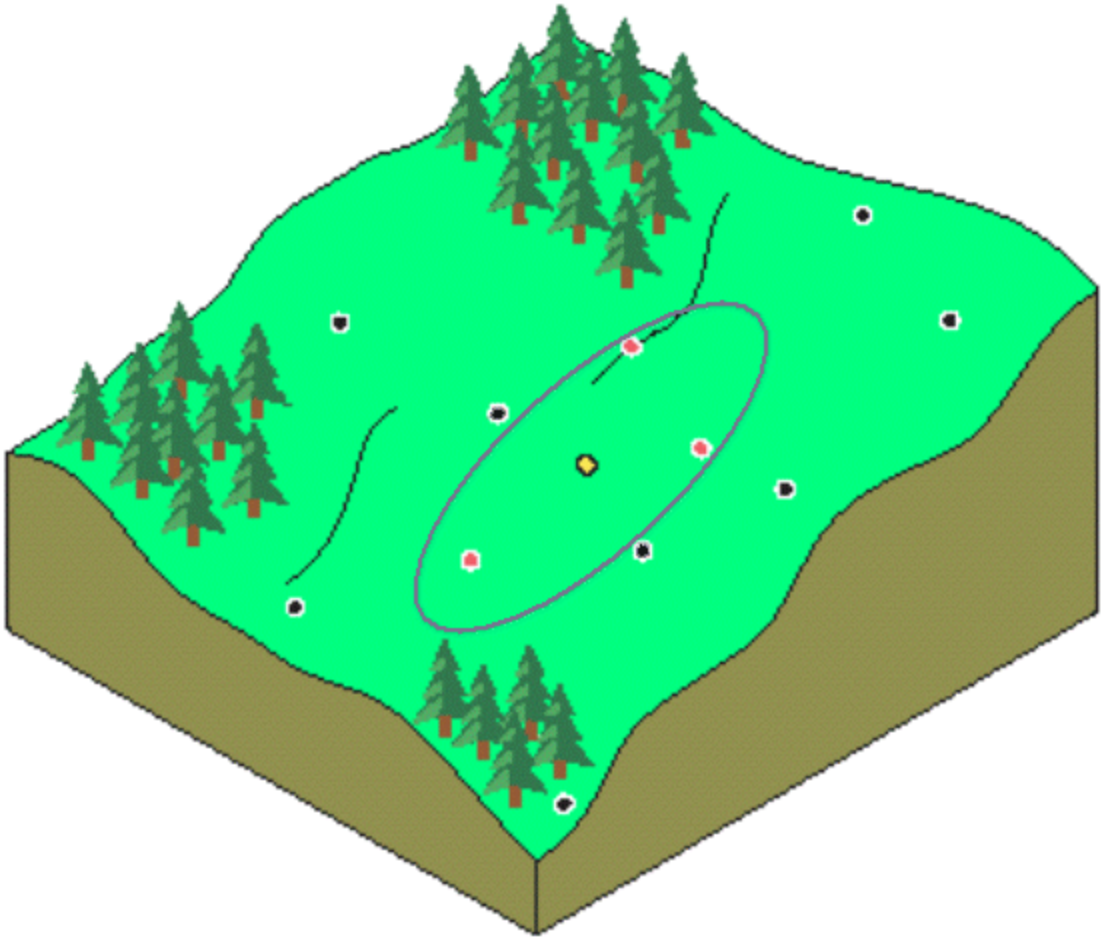
Slučaj izotropije

- Ukoliko ne postoji poseban pravac u kojem je uticaj okolnih uzoraka od većeg značaja tada se prozor za pretraživanje zadaje u obliku kružnice sa zadatim radijusom.



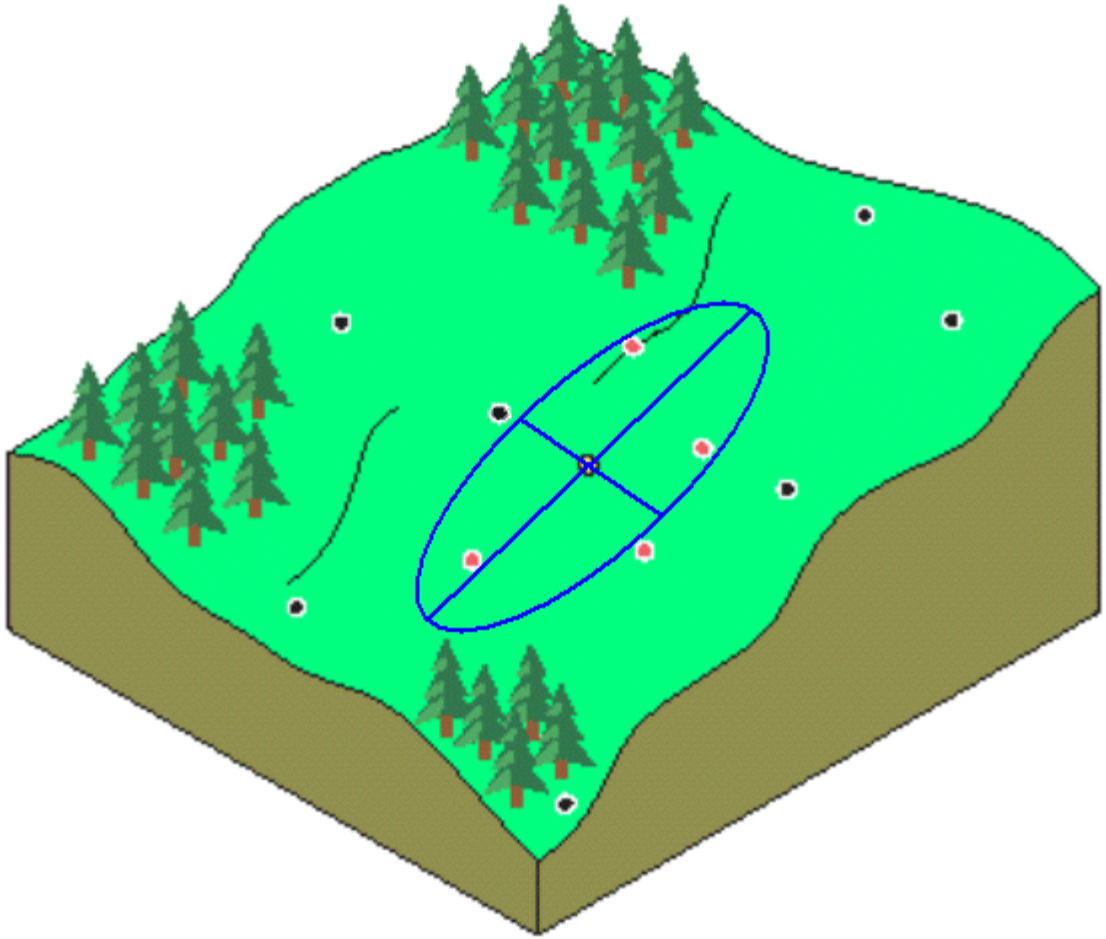
Definisanje smera uticaja

- U slučaju anizotropije, kada je uticaj podataka u određenom pravcu značajniji u odnosu na druge (npr. pravac površinskog oticanja voda), tada se umesto prozora u obliku kružnice može definisati i elipsa.



Izbor tačaka po sektoru

- U okviru prozora za pretraživanje definišu se sektori (kvadranti ili oktanti) u okviru kojih se zadaje mnimalni i maksimalni broj tačaka čije vrednosti se koriste pri lokalnoj intepolaciji.



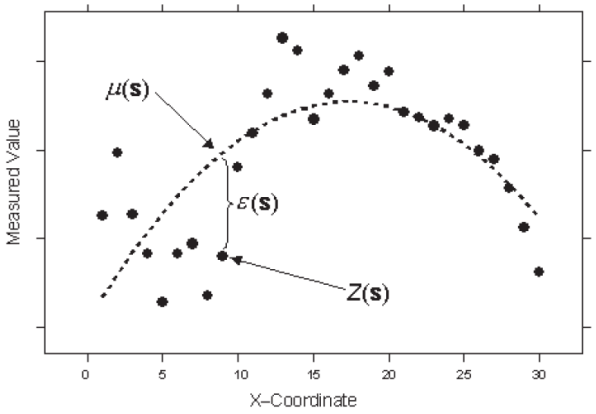
Kriging metode

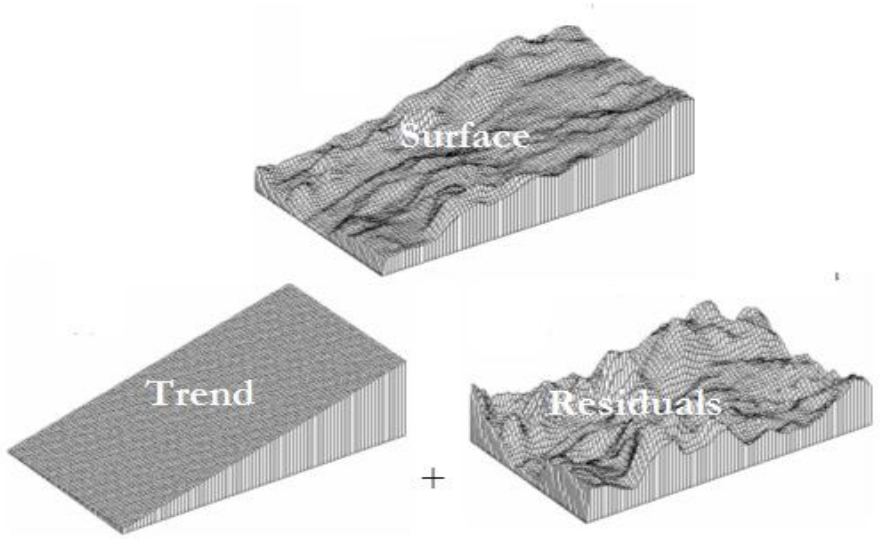
- Universal Kriging – Kriging sa trendom
- Indicator Kriging – sa definisanim pragom vrednosti
- Stratified Kriging – regionalizovani pristup
- Co-Kriging – korišćenje pridruženih podataka

Kriging metode

- ***Ordinary Kriging*** najopštiji model i najčešće korišćeni. Predpostavlja da je srednja vrednost promenljive konstantna veličina (ne postoji trend među merenim podacima).
- ***Simple Kriging*** predpostavka poznavanja srednje vrednosti promenljive (konstantna veličina).
- **Universal Kriging** predpostavlja prisustvo trenda među merenim podacima, npr. preovlađujući vetrovi gde se trend može modelovati determinističkim polinomskim funkcijama. U ovom slučaju se trend oduzima od merenih uzoraka (dobijaju se reziduali) i tada se vrši ocena prostorne korelacije.

Universal Kriging





$$Z(s) = \mu(s) + \varepsilon(s)$$

Trend se ocenjuje u funkciji prostornih koordinata, UK je dobro rešenje kada je evidentna prisutna nestacionarnost među merenim podacima, tj srednja vrednost podataka varira u prostoru.

Procena pomoću dodatnih podataka

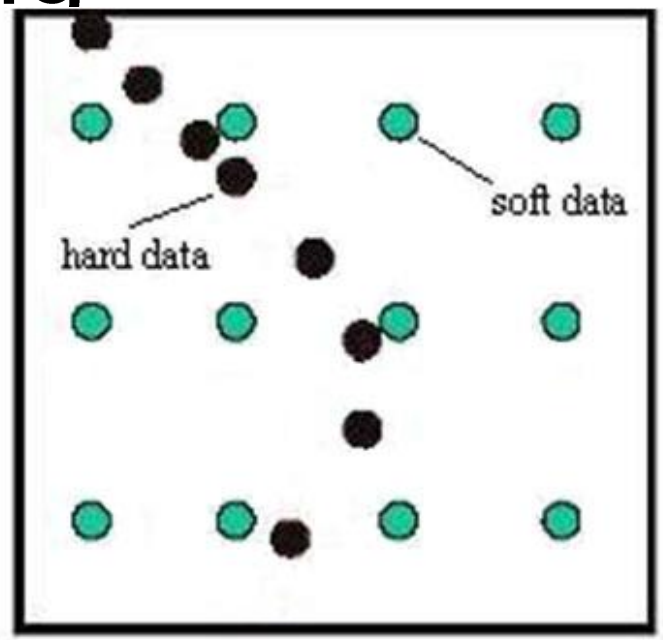
- Co-kriging
- Regresioni kriging ili kriging sa eksternim trendom (with external drift)

Cokriging

- Želimo da procenimo ciljanu promenljivu koja je obično ili teška za uzorkovanje i laboratorijsku obradu, pa imamo samo mali broj uzoraka.
- Međutim, imamo sekundarnu promenljivu (obično laku za uzorkovanje) koja je u korelisana sa ciljanom promeljivom.

Cokriging

$$\hat{u}_0 = \sum_{i=1}^{n} a_i \cdot u_i + \sum_{j=1}^{m} b_j \cdot v_j$$



$u_1, \ldots, u_n$ - Primarna promenljiva (*hard data*)

$v_1, \ldots, v_m$ - sekundarna promeljiva (*soft data*)

Cokriging

- Cokriging sistem jednačina:

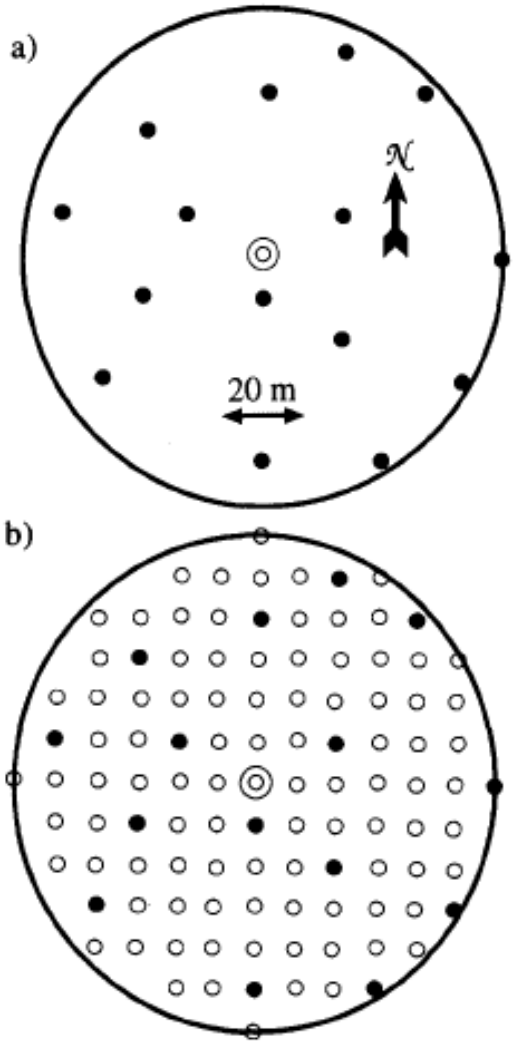
$$\sum_{i=1}^{n} a_i Cov\ U_i U_j + \sum_{i=1}^{m} b_i Cov\ V_i U_j + \mu_1 = Cov\ U_0 U_j \quad \text{za } j = 1,\ldots n$$

$$\sum_{i=1}^{n} a_i Cov\ U_i V_j + \sum_{i=1}^{m} b_i Cov\ V_i V_j + \mu_2 = Cov\ U_0 V_j \quad \text{za } j = 1,\ldots m$$

$$\sum_{i=1}^{n} a_i = 1$$

$$\sum_{i=1}^{m} b_i = 0$$

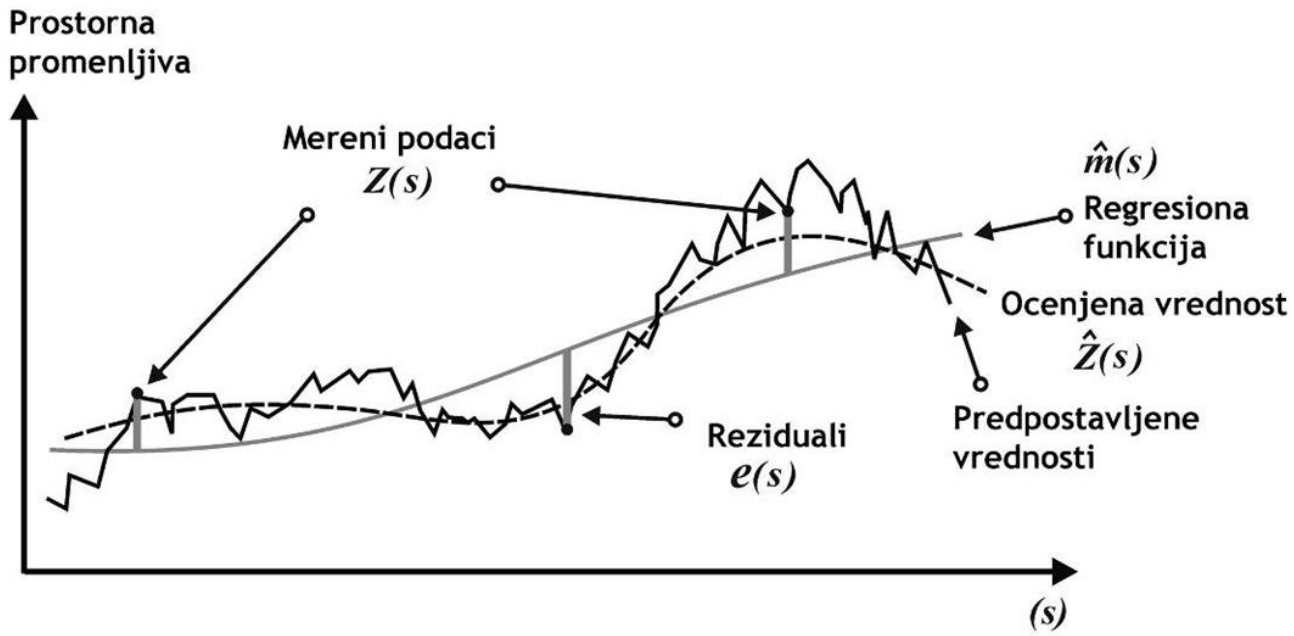
Cokriging



a)Isotopic

b)Heterotropic

Regresioni kriging



$$\hat{z}(s_0) = \hat{m}(s_0) + \hat{e}(s_0)$$

$$\hat{z}(s_0) = \sum_{k=0}^{p} \hat{\beta}_k \cdot q_k(s_0) + \sum_{i=1}^{n} w_i(s_0) \cdot e(s_i);$$

$$q_0(s_0) = 1$$

$\hat{m}(s_0)$ -fitted deterministic part

$\hat{e}(s_0)$ -interpolated residual

$\hat{\beta}_k$ -the estimated drift model coefficients

$q_k(s_0)$ -is the kth external explanatory variable or predictor at location s_0

Pomoćne (eksterne) promenljive

- Digitalni modeli terena
- Primarni topografski parametri (nagib, ekspozicija, zakrivljenost ...).
- Sekundarni topografski parametri (topographic wetness index, stream-power indexes, radiation index, natural drainage networks...).
- Podaci iz satelitskih snimaka (NDVI, ili drugi vegetacioni indeksi...).
- Digitalizovane tematske karte (biljni pokrivač ,namena zemljišta...).