Lesnická a dřevařská fakulta

2013, Brno
Ing. Tomáš Mikita, Ph.D.
Doc. Ing. Martin Klimánek, Ph.D.

Mendelova univerzita v Brně

Využití GIS a DPZ pro krajinné inženýrství
přednáška č.6

Prostorová interpolace dat,geostatistika, digitální modely terénu

evropský sociální fond v ČR
EVROPSKÁ UNIE
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY
OP Vzdělávání pro konkurenceschopnost
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Prezentace vznikla za podpory projektu reg. č. CZ 1.07/2.2.00/15.0080

Prostorová interpolace

- procedura odhadu neznámých hodnot ze známých hodnot v okolí
    - interpolace
    - extrapolace

- úkolem interpolace je určit vhodnou funkci y=f(x), která v daných (tzv. uzlových) bodech nabývá známých hodnot

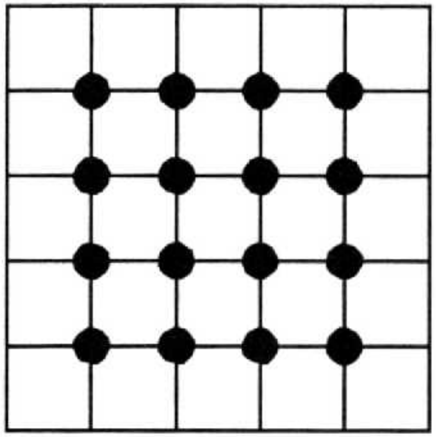
- interpolační funkce f(x) - shoda s původními hodnotami v uzlových bodech

- aproximující funkce f(x) - nahrazuje původní funkci fo(x) s jistou přesností

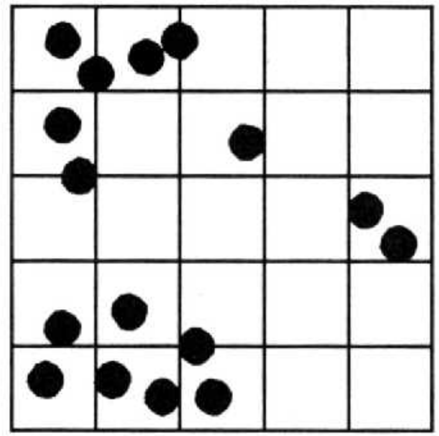
- Dodržují se známé hodnoty?
    - interpolační (exaktní) metody (deterministické)
    - aproximující metody (stochastické)

Formát a rozmístění zdrojových dat

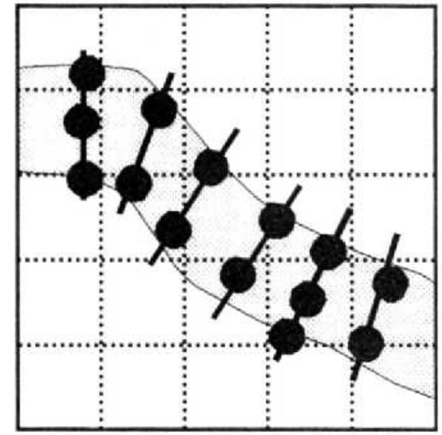
- Rozmístění (prostorová lokalizace) vstupních dat je důležitým kritériem při volbě vhodné interpolační metody (některé interpolační metody při nevhodném rozmístění vstupních dat poskytují chybné výsledky nebo je nelze vůbec použít).
- Rozmístění vstupních dat, se kterými se uživatel v praxi nejčastěji setkává jsou na obrázku (viz dále). Jako vhodný kompromis mezi pravidelným a náhodným rozmístěním je možné použít stratifikované náhodné rozmístění, kdy je zajištěno pokrytí celé zájmové oblasti (rozdělené v pravidelném rastru). Shluky (skupiny) se nejčastěji využívají při analýze prostorové variability. Specifickými problémy se potom vyznačuje interpolace na profilech a izoliniích, zvláště pokud je vyžadována extrapolace nebo je množství dat nedostatečné.



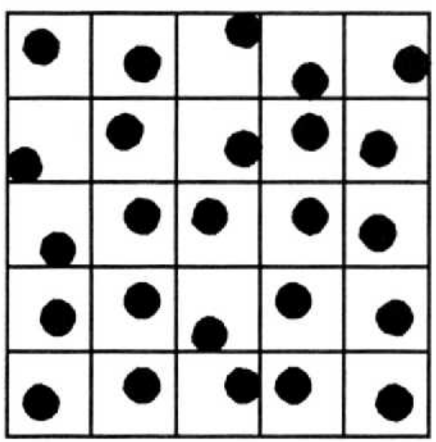
pravidelné rozmístění



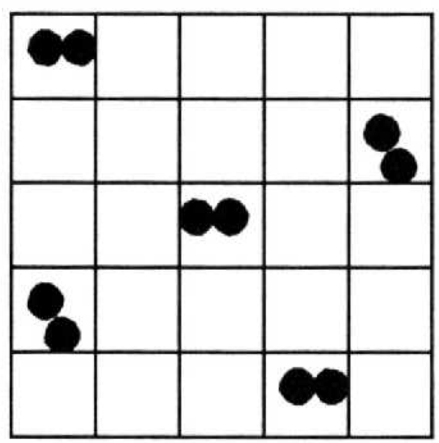
náhodné rozmístění



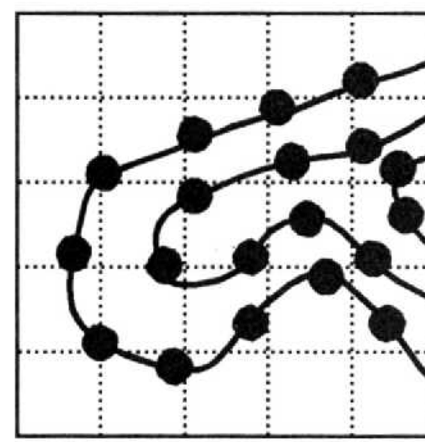
transekty (profily)



stratifikované náhodné

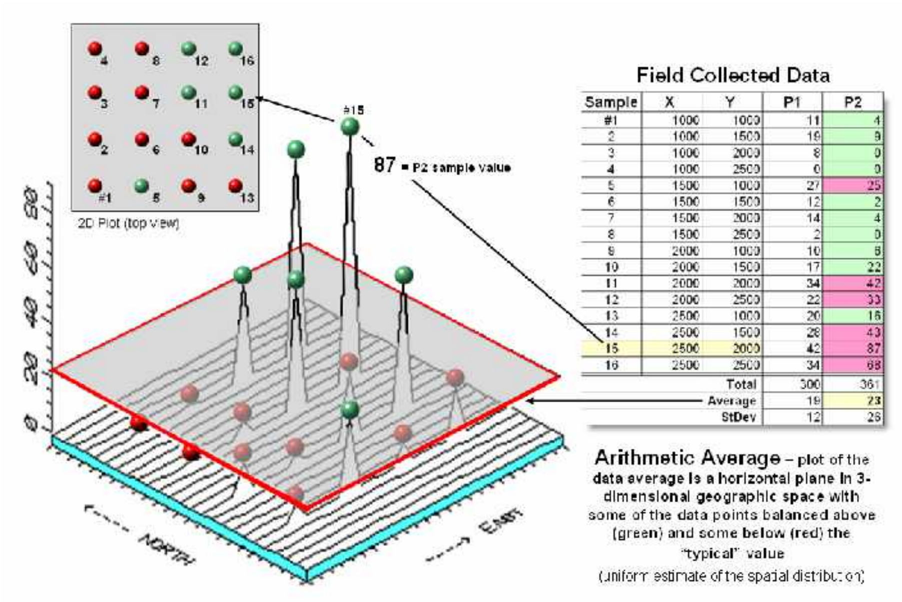


shluky (skupiny)

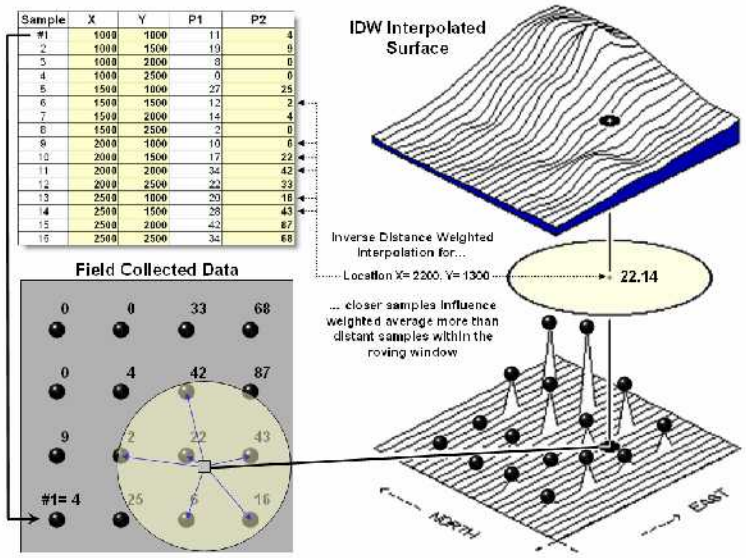


vrstevnice (izolinie)

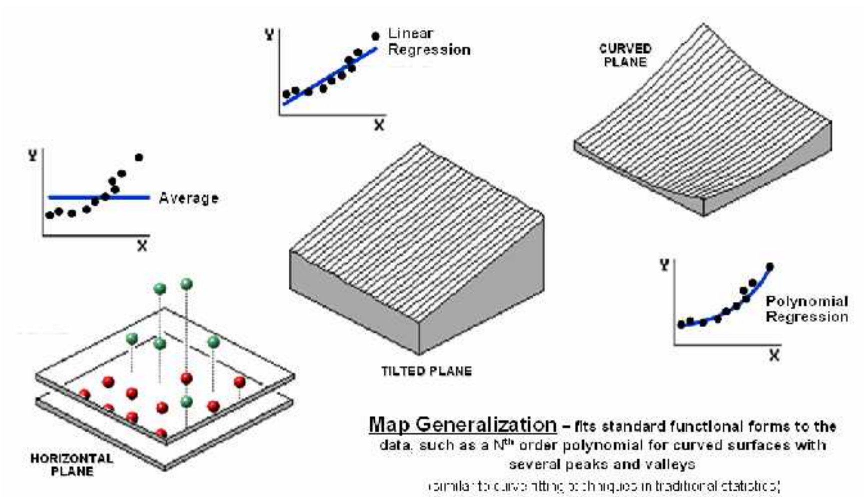
(Burrough and McDonnell 1998)

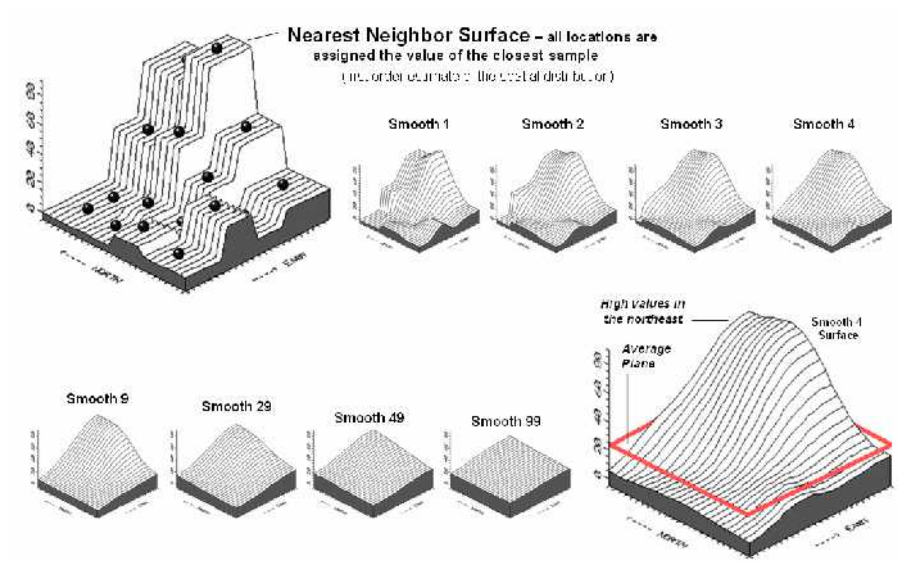


Sample	X	Y	P1	P2
#1	1000	1000	11	4
2	1000	1500	19	9
3	1000	2000	8	0
4	1000	2500	0	0
5	1500	1000	27	25
6	1500	1500	12	2
7	1500	2000	14	4
8	1500	2500	2	0
9	2000	1000	10	6
10	2000	1500	17	22
11	2000	2000	34	42
12	2000	2500	22	33
13	2500	1000	20	16
14	2500	1500	28	43
15	2500	2000	42	87
16	2500	2500	34	68
		Total	300	361
		Average	19	23
		StDev	12	26



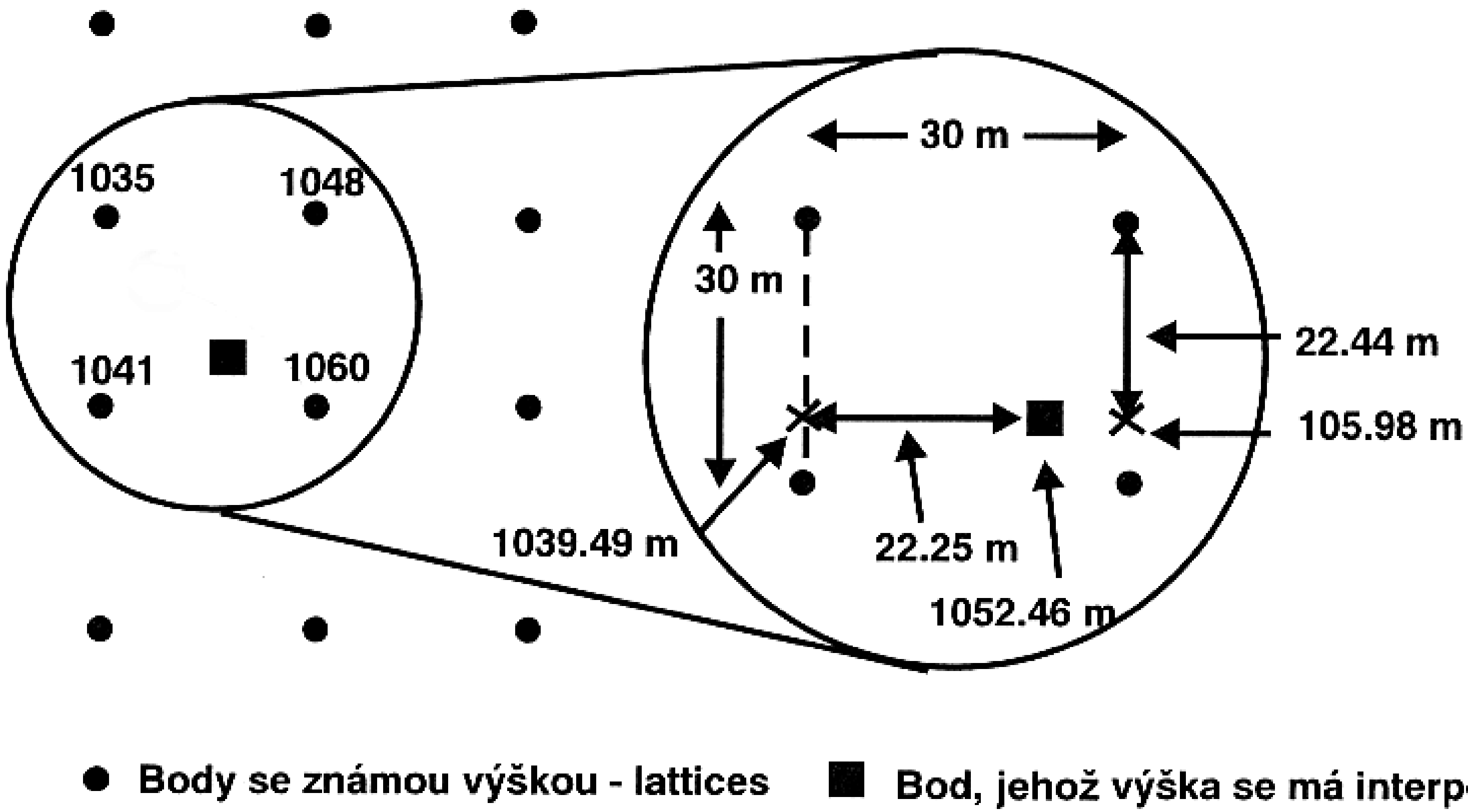
Sample	X	Y	P1	P2
#1	1000	1000	11	4
2	1000	1500	19	9
3	1000	2000	8	0
4	1000	2500	0	0
5	1500	1000	27	25
6	1500	1500	12	2
7	1500	2000	14	4
8	1500	2500	2	0
9	2000	1000	10	6
10	2000	1500	17	22
11	2000	2000	34	42
12	2000	2500	22	33
13	2500	1000	20	16
14	2500	1500	28	43
15	2500	2000	42	87
16	2500	2500	34	68



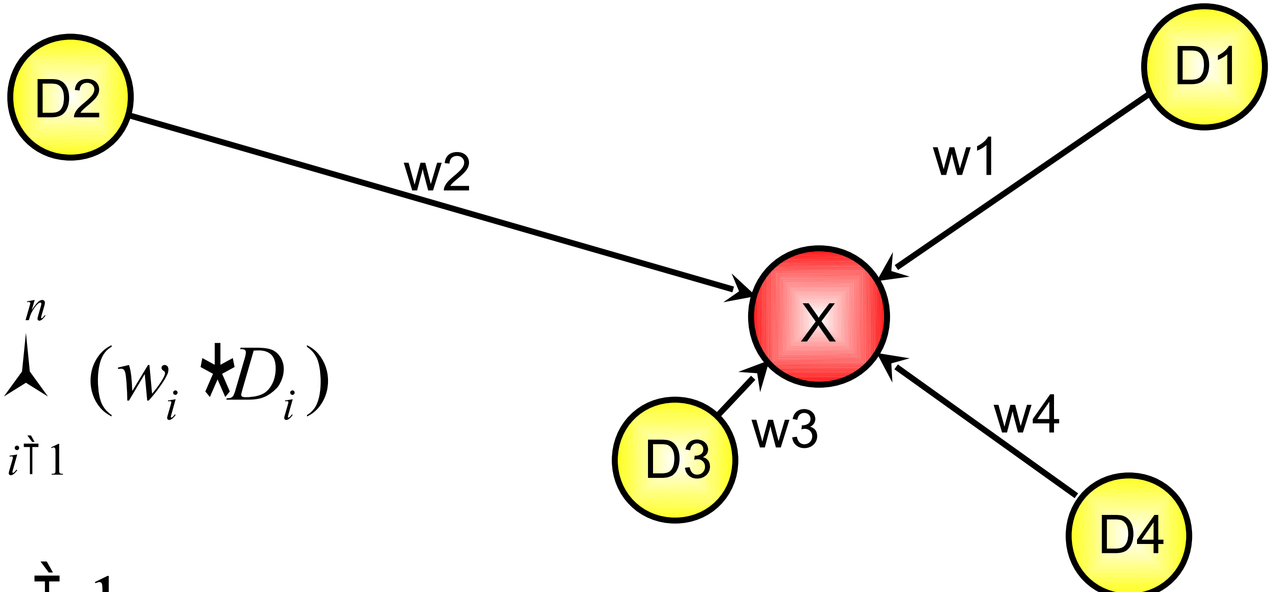


http://www.innovativegis.com/basis/MapAnalysis/

Bilineární interpolace



Vážený průměr



$$X = \sum_{i=1}^{n} (w_i * D_i)$$

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

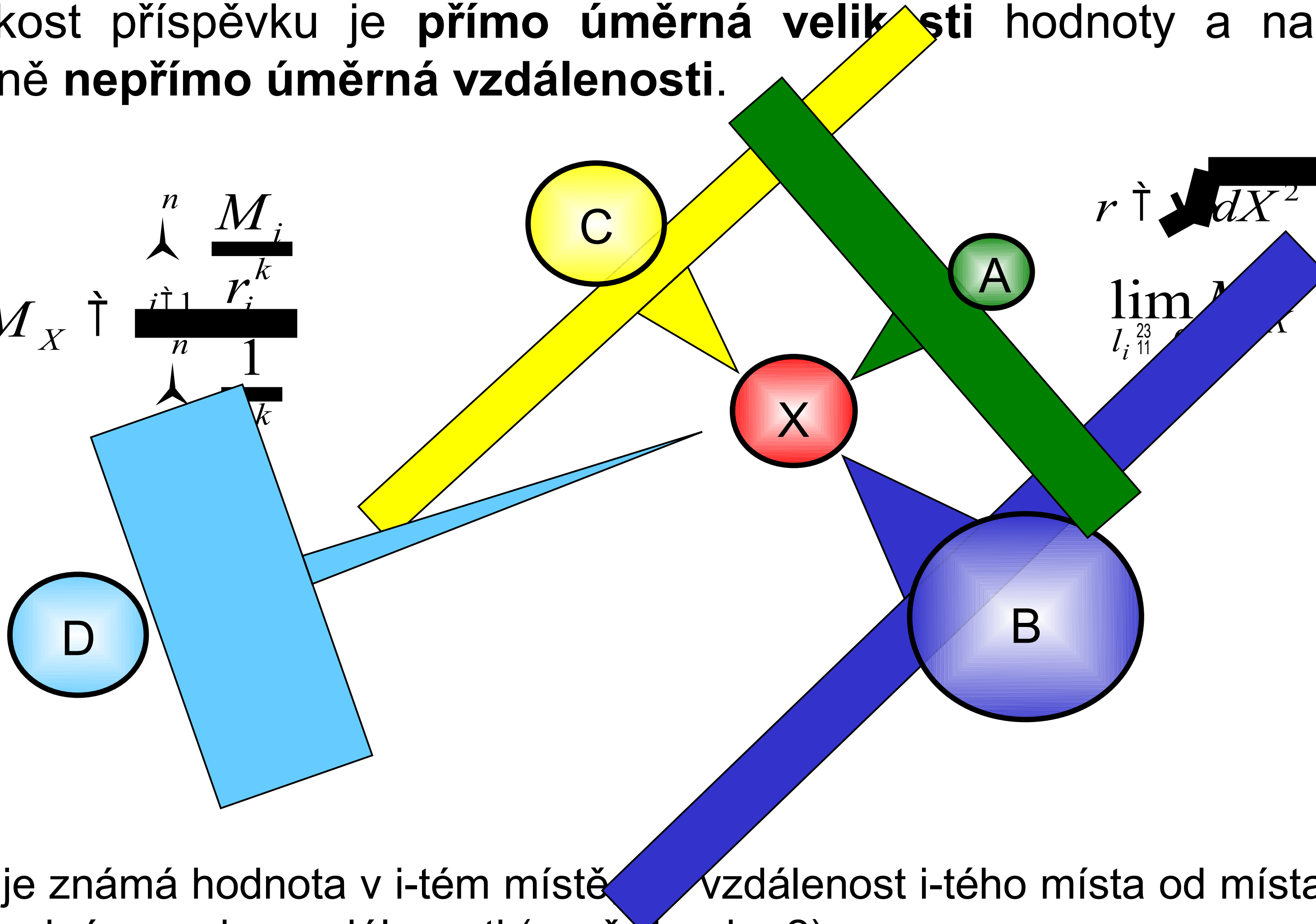
Metoda inverzních vzdáleností (ID, IDW) – princip

Velikost příspěvku je **přímo úměrná velikosti** hodnoty a na druhé straně **nepřímo úměrná vzdálenosti**.

$$M_X = \frac{\sum_{i=1}^{n} \frac{M_i}{r_i^k}}{\sum_{i=1}^{n} \frac{1}{r_i^k}}$$

$$r = \sqrt{dX^2 + dY^2}$$

$$\lim_{r_i \to 0} M_X = M_i$$



„Mi“ je známá hodnota v i-tém místě … vzdálenost i-tého místa od místa X a „k“ je vhodná mocnina vzdálenosti (např. 1 nebo 2).

Metoda inverzních vzdáleností (ID) – varianty:

- s výběrem z kvadrantů nebo oktantů – rozdělení prohledávané oblasti do 4 nebo 8 sektorů
- se směrovým vážením – rozdělení prohledávané oblasti do nepravidelných sektorů. Vstupní parametry:
  - úhlové rozpětí sektorů,
  - váhy přiřazené jednotlivým sektorům
  - = (zavádění anizotropie pole)
- z každého sektoru je vybrán 1 bod (nebo „k“ bodů) a na tuto množinu vybraných bodů je aplikována metoda ID

Shepardova metoda (SH) – metoda inverzních vzdáleností + vyrovnání metodou nejmenších čtverců

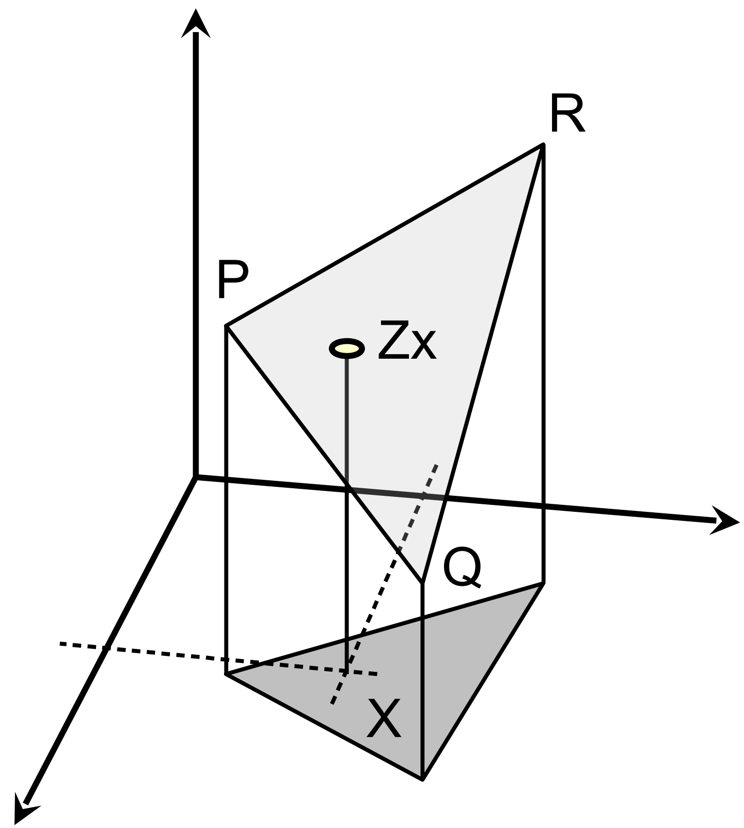
Metoda triangulace s lineární interpolací (TR)

- poskytuje odhad neznámé hodnoty pomocí lineární závislosti. Lineárním útvarem je tedy rovina. Rovnice roviny obsahuje tři koeficienty - to znamená, že pro její určení jsou zapotřebí tři známé body:

  P=[x1,y1,z1], Q=[x2,y2,z2], R=[x3,

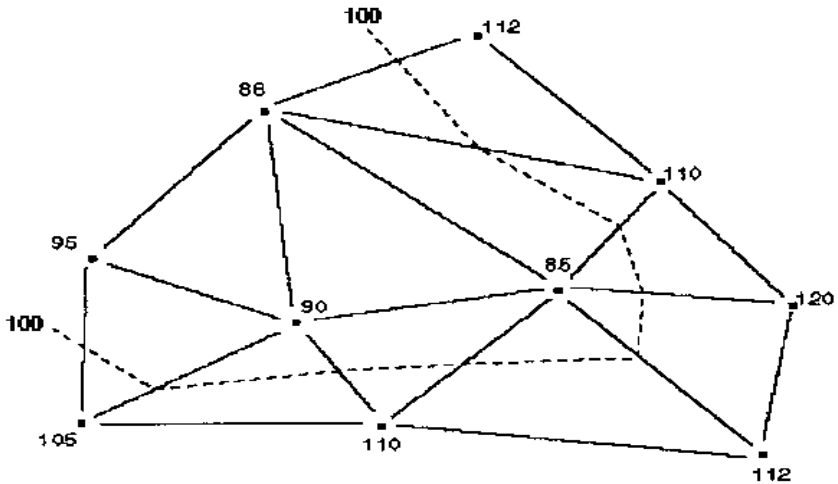
- tři body P, Q a R jsou vybrány tak, aby bod X ležel uvnitř trojúhelníka získaného jako průmět trojúhelníka PQR. Těmito body je určena rovina, jejíž koeficienty lze získat řešením soustavy rovnic. Hledaný odhad Zx je dán vztahem:

$$Z_x = ax_0 + by_0 + c$$



Metoda triangulace s lineární interpolací

- dodržuje naměřené hodnoty
- výhodná je pro modelování nespojitostí v poli (zlom, hrana)
- nevýhodná pro nespojitost vypočtených hodnot (kostrbatý průběh izolinií) – různé stupně vyhlazení
- interpolace izolinií



Metoda triangulace s lineární interpolací

- Skrývá jistý prvek náhodnosti.

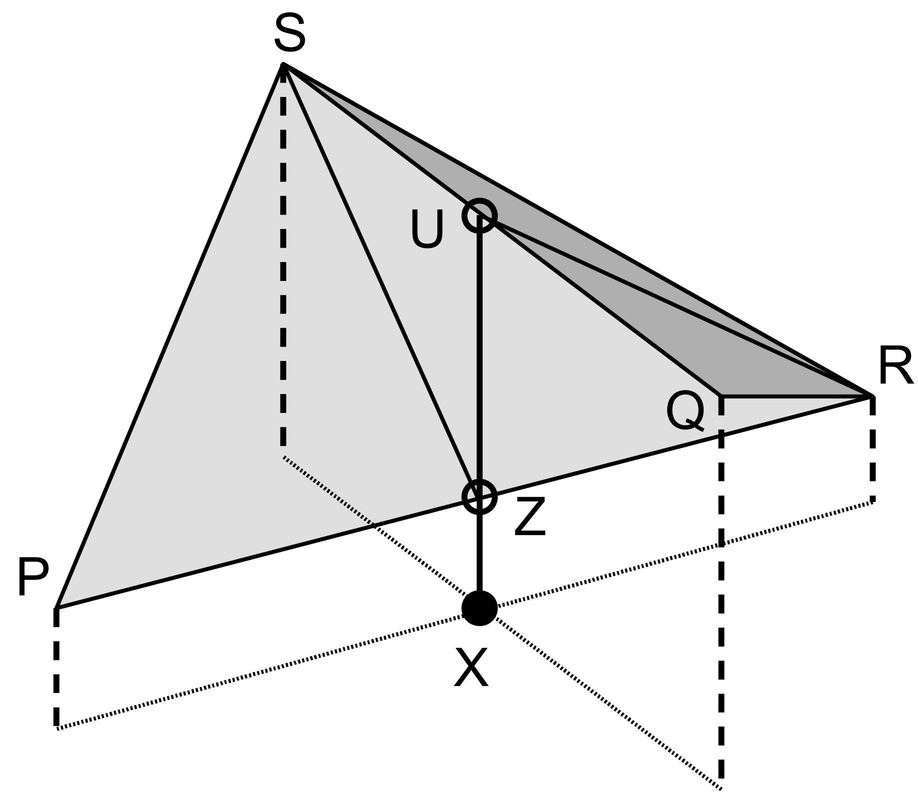
  Uvažujme následující data:

P=[1,1,1], Q=[3,1,10], S=[3,1,10], R=[3,3,1]. Odhad pro X=[2,2,?].

- Náhodnost spočívá ve volbě trojúhelníka:

  Volbou PRS, je odhadem hodnota 1.

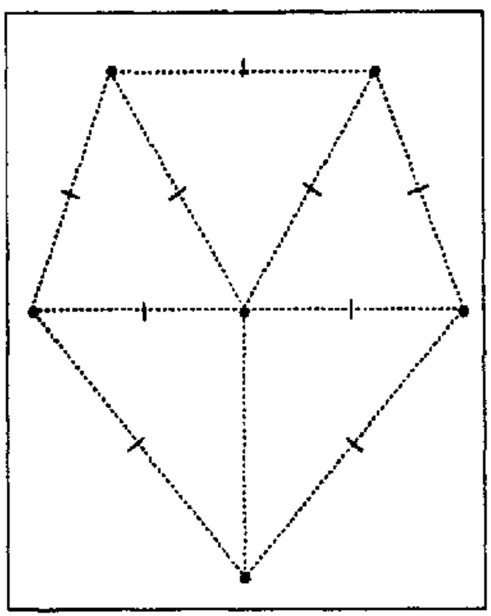
  Volbou QRS, je odhadem hodnota 10.

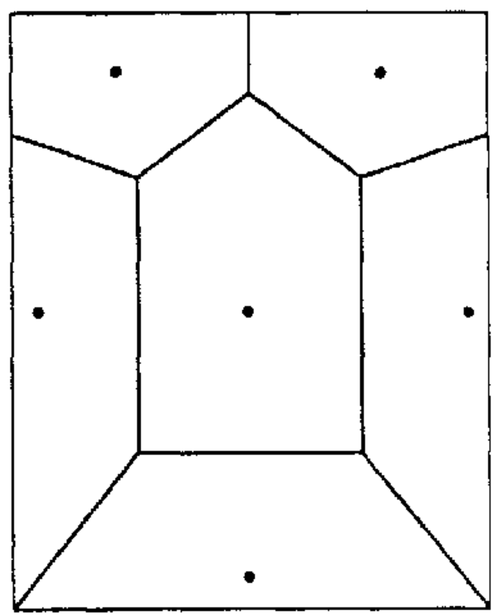


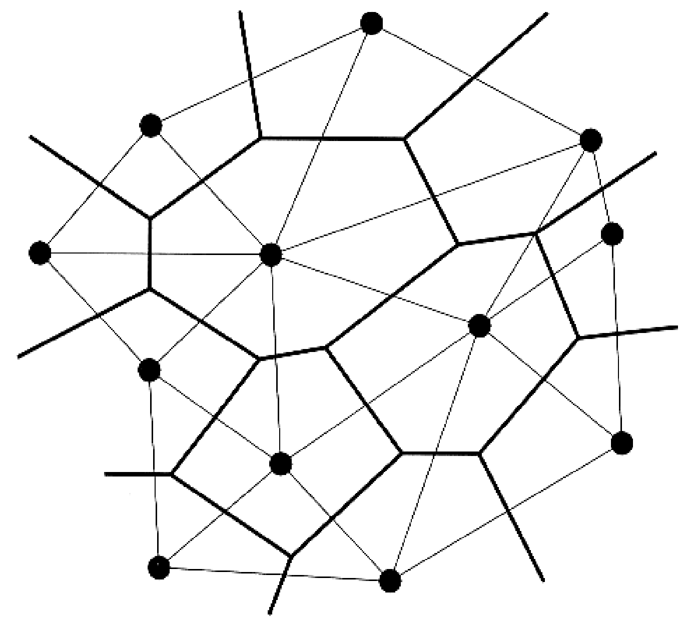
Problém je tedy ve velkém rozdílu dat mezi „sousedními“ hodnotami. Bohužel právě takový charakter má mnoho geoprostorových dat, a proto zvláště na taková data není vhodné metodu aplikovat. S dobrými výsledky ji lze použít v případech, kdy hodnoty nemají příliš velký rozptyl.

Thiessenovy (Dirichlet, Voronoi) polygony (TP)

Představují přesnou metodu interpolace, která vychází z předpokladu, že neznámé hodnoty bodů odpovídají hodnotě nejbližších známých bodů. Zahrnuje šíření teritoria sdruženého s bodem, které pokračuje tak dlouho, až se narazí na obdobně zpracovávané teritorium sousedního bodu. Je-li rozmístění bodů nepravidelné, výsledkem bude mozaika polygonů.







Thiessen polygons (thick lines)

Delaunay triangulation (thin lines)

Metoda minimální křivosti (MC)

- spline-funkce – vychází z interpolace pomocí (nejčastěji kubických) funkcí. „Spojuje“ tedy dvojice daných bodů segmenty kubické křivky (ten je dán čtyřmi body). Z prvních čtyř bodů se spočte kubická křivka a první dva body se spojí jejím segmentem. Pak se z druhého až pátého bodu spočte kubická křivka a druhé dva body se spojí jejím segmentem, atd.

- polynomické funkce, na hranách spojité (spojitost interpolující funkce a stanoveného počtu jejich derivací)

- povrch je interpolován po částech

- hladké povrchy

- míra aproximace je dána stanovením tolerance a počtem iterací

Metoda radiálních funkcí (RF)

- hladký povrch
- dodržuje naměřené hodnoty
- multikvadriková metoda

$$D_i(x,y) = \sqrt{d_i(x,y)^2 + R^2}$$

Di(x,y) je radiální funkce vzdálenosti di(x,y)
di(x,y) je relativní, anizotropní vzdálenost mezi místem měření (xi ,yi) a místem odhadu (x, y)
R2 je vyhlazovací faktor

V každém místě jsou stanoveny optimální váhy řešením soustavy lineárních rovnic.

Fourierova analýza (FA)

- obecný interpolační postup, který se používá v geostatistických metodách pro odhad průběhu povrchu pomocí série funkcí sinus a kosinus
- nejvhodnější pro datové soubory, které vykazují periodicitu – analýza signálů, modelování klimatických změn, pohyb (vln v oceánu, sněhových závějí, pískových dun apod.) a zpracování obrazu (odstranění šumu, korekce); redukce variability (kontrastů) v DMT pro snadnější identifikaci tvarů terénu (hřbety, údolí)

$$f(x) = a_0 + \sum_{k=1}^{n} \left[a_k \cos(k\omega x) + b_k \sin(k\omega x) \right]$$

$$amplituda = \sqrt{a^2 + b^2}$$

$$fáze = \tan^{-1}(b/a)$$

$$frekvence(spektrum) = \ln(1 + amplituda^2)$$

$\omega = 2\pi/T$

k = koeficient harmonické funkce

a, b = koeficienty Fourierovy transformace

Nyquistova frekvence – nejvyšší frekvence = nejkratší vlna = 2·pixel

Geostatistické metody - krigování (KR) – úvod

- „teorie regionalizovaných proměnných“, která byla vytvořena G. Matheronem a D. G. Krigem (metoda interpolace dat získávaných při vyhledávání a průzkumu ložisek nerostných surovin)
- vychází ze zjištění, že prostorová variabilita řady geoprostorových prvků je příliš nepravidelná než, aby mohla být modelována pomocí vyhlazovacích matematických funkcí – optimalizovaná soustava vah v každém kroku s požadavkem minimálního rozptylu
- základ spočívá v nalezení průměrné hodnoty změn v závislosti na změně vzdálenosti mezi měřenými body – vyhodnocování strukturálních funkcí
- optimalizuje prostorovou interpolaci tak, že rozděluje prostorovou variabilitu do tří komponent:
  - deterministickou variabilitu (různé trendy)
  - prostorově autokorelovanou variabilitu
  - nekorelovaný šum
- charakter prostorově korelované variability lze znázornit jako semivariogram, který poskytuje informaci pro optimalizaci interpolačních vah a prohledávacích poloměrů

Krigování (KR) – úvod

Výchozí předpoklady pro strukturální analýzu a krigování:

- normální distribuce
  - transformace nebo použití nelineárních technik
- průměrná hodnota konstantní
  - trend, univerzální krigování
- prostorová autokorelace konstantní

Krigování (KR) – experimentální semivariogram

Semivariogram je základní geostatistický nástroj pro vizualizaci, modelování a využití prostorové autokorelace regionalizované proměnné. Semivariance je míra variance.

$$\gamma(h) = \frac{1}{2n} \sum_{i} \left[z(x_i) - z(x_i + h) \right]^2$$

kde $\gamma(h)$ je semivariance proměnné *z* pro vzdálenost *h*

Základní schéma výpočtu pro vzdálenost **h** sestává z následujících kroků:

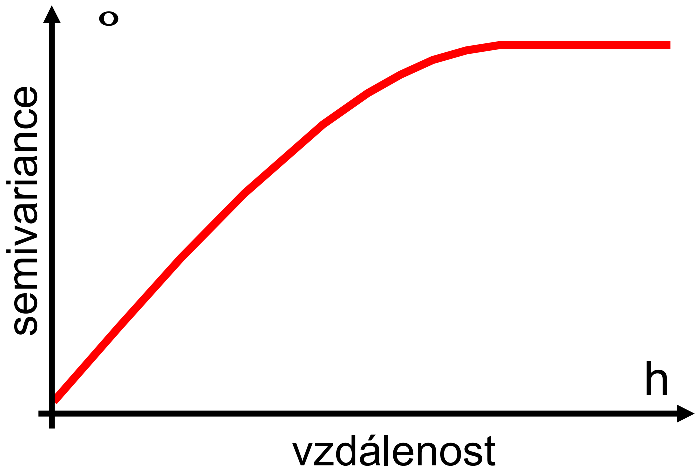
- v úvahu se vezmou všechny takové páry **zi** a **zj**, jejichž vzdálenost padne do třídy pro **h**
- vypočtou se rozdíly hodnot těchto párů
- sečtou se kvadráty těchto rozdílů
- součet se vydělí dvojnásobkem počtu párů

Tak se pro všechny hodnoty **h** získá řada hodnot nazývaných **experimentální semivariance**.

Krigování (KR) – teoretický semivariogram

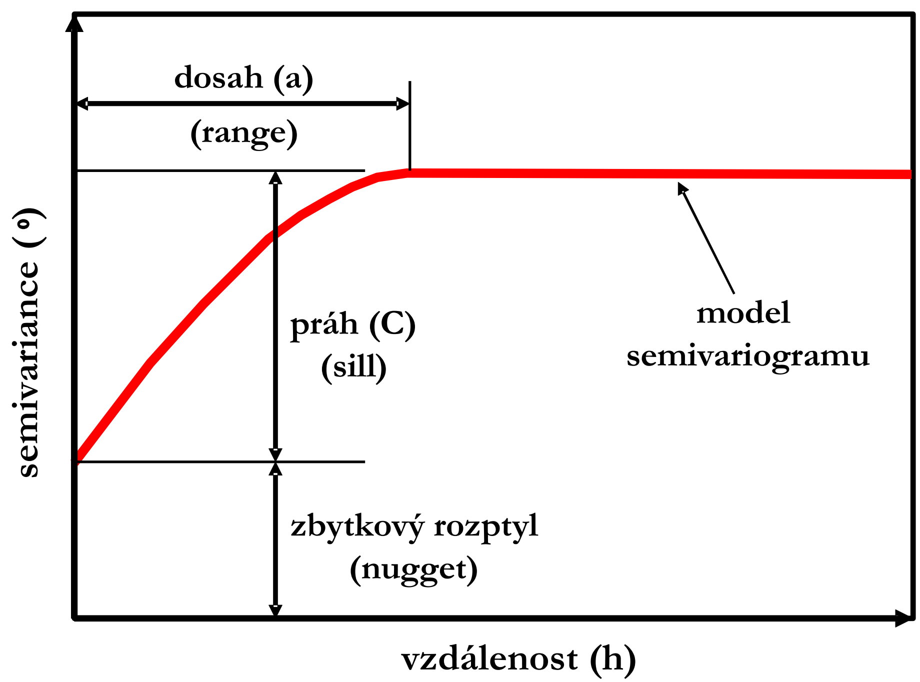
- Funkční vztah, který pokud možno dobře sleduje (modeluje) experimentální semivariogram, se nazývá **teoretický semivariogram**
- Chování semivariogramu je možno (víceméně) intuitivně popsat takto:

  - velmi blízká data mají velmi malou odchylku
  - data ve větších vzdálenostech mají větší odchylky, avšak odchylky pro velmi vzdálená data a velmi velice vzdálená data se už příliš neliší
  - od jisté vzdálenosti už vzájemné odchylky nerostou (např. také proto, že vzdálenost překračuje rozměry zkoumané plochy nebo tělesa)



Krigování (KR) – strukturní analýza

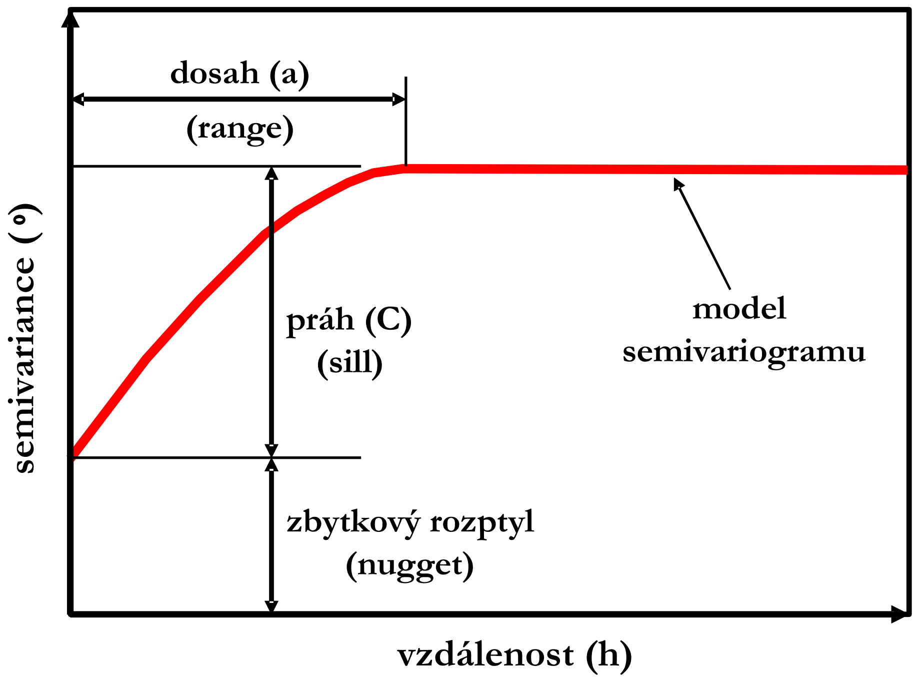
- proces hledání teoretického semivariogramu pro daný experimentální semivariogram je nazýván **strukturní analýzou**. Model nalezený pro danou množinu dat závisí jak na experimentálních, tak teoretických předpokladech. Vlastnosti, které prakticky vedou k určení konkrétního teoretického modelu, jsou:
  - přítomnost nebo absence „ploché části“ semi-variogramu, to znamená, že při zvětšující se vzdálenosti se hodnoty variance nemění - existuje tzv. práh (sill); je dán konstantou „**C**“
  - vzdálenost, ve které semi-variance dosáhne prahové hodnoty - tzv. dosah (range); je dán konstantou „**a**“
  - chování v počátku (tj. semi-variance mezi velmi blízkými body)



Krigování (KR) – strukturní analýza

- dosah (range) je mírou korelace uvnitř množiny dat; „dlouhý“ dosah indikuje vysokou korelaci, „krátký“ korelaci nízkou
- hodnota prahu (sill) je rovna celkovému rozptylu

- velmi často nenabývají experimentální semivario-gramy v počátku nulové hodnoty; protínají osu *y* v nenulové hodnotě, která je nazývána zbytkový rozptyl (nugget effect). To může ukazovat na rozptyl menší než je „vzorkovací“ vzdálenost nebo na malou přesnost měření (např. jsou v datech obsaženy dva vzorky ze stejného místa, pokaždé s jinou hodnotou).



Krigování (KR) – modely semivariogramů

- Sférický model

$$\gamma(h) = C\left(\frac{3h}{2a} - \frac{h^3}{2a^3}\right) \qquad \text{pro } h \leq a$$

$$\gamma(h) = C \qquad \text{pro } h > a$$

- Exponenciální model

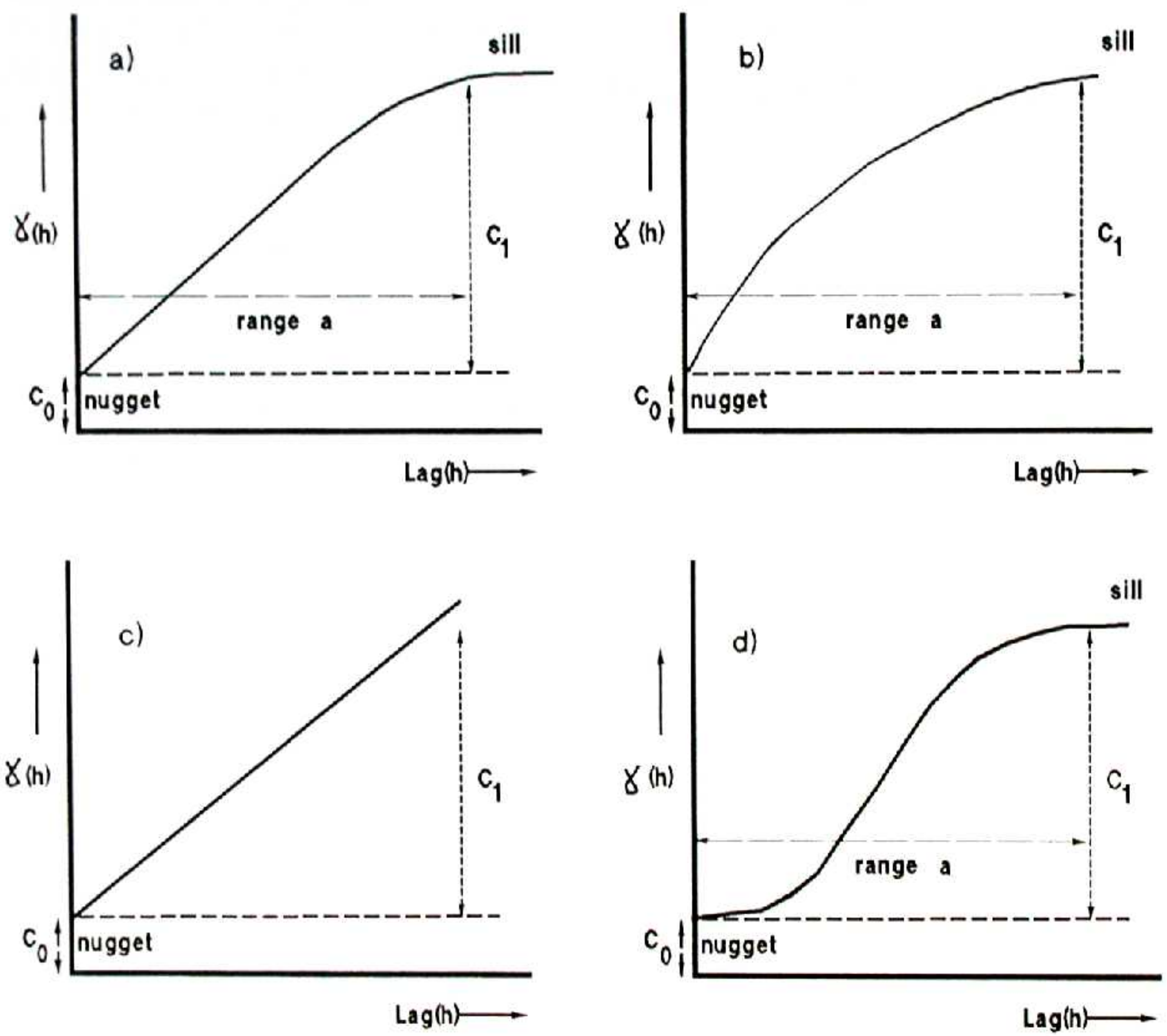
$$\gamma(h) = C\left[1 - e^{-h/a}\right]$$

- Gaussův model

$$\gamma(h) = C\left(1 - e^{-\frac{h^2}{a^2}}\right)$$

- Složené modely – každý zdroj variability je popsán vlastní strukturní funkcí

Krigování (KR) – modely semivariogramů



- a) sférický
- b) exponenciální
- c) lineární
- d) gaussovský

Krigování (KR) – směrovost semivariancí

- mnohé fenomény nevykazují prostorovou isotropii rozptylu - naopak se chovají anisotropně – rozsah influence je různý v různých směrech

- pro data, vykazující anisotropii, se využívá postupu spočívajícího v konstrukci dílčích semivariogramů pro stanovené směrové tolerance

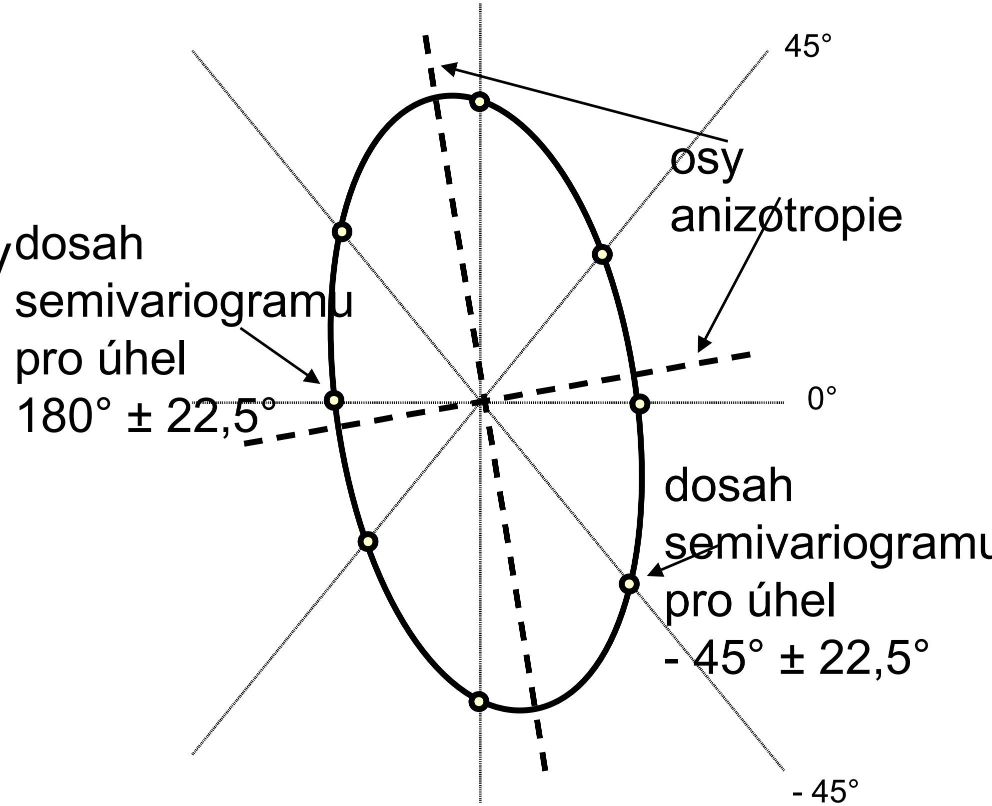
V nejjednodušším případě rozdělíme všesměrové pole rovnoměrně:

- například na směry sever $^{238}_{93}$22.5o, jih $^{238}_{93}$22.5o, východ $^{238}_{93}$22.5o, západ $^{238}_{93}$ 22.5o, severovýchod $^{238}_{93}$ 22.5o, jihovýchod $^{238}_{93}$ 22.5o, jihozápad $^{238}_{93}$22.5o, severozápad $^{238}_{93}$22.5o

a zkonstruujeme osm dílčích směrových semivariogramů tak, že do každého z nich zahrneme pouze ty dvojice daných bodů, jejichž směrový vektor padne do intervalu směrů daného dílčího semivariogramu.

Krigování (KR) – směrovost semivariancí

- Při směrově symetrických datech stačí zkonstruovat pouze polovinu počtu semivariogramů. Po zjištění jejich dosahů se tyto dosahy zanesou do růžicového diagramu obsahujícího ty směry, pro něž byly dílčí semivariogramy sestaveny.
- Anizotropická množina dat je charakterizována směrem maximální variance a směrem minimální variance. Tyto směry jsou směry hlavní a vedlejší poloosy tzv. elipsy anizotropie. Elipsa anizotropie je pak zjistitelná jako elipsa, která aproximuje dosahy vynesené do shora zmíněného růžicového diagramu.



Krigování (KR) – varianty

- simple kriging – jednoduché krigování – známá průměrná hodnota veličiny
- cokriging – odhad proměnné na základě hodnot korelovaných veličin, vztahy popsané pomocí vzájemného semivariogramu
- lognormální krigování
- indikátorové krigování – transformace kvantitativních údajů na kvalitativní (ano/ne) – výsledkem lokálního odhadu je pravděpodobnost; (absolutní nebo relativní indikátory - problém konzistence)
- probability kriging – kombinace indikátorů s původními údaji
- soft kriging – kombinace hodnoty s dalšími údaji

Podmíněná stochastická simulace (SS)

1. Vyberte náhodně dosud nezpracovaný bod (neznámé hodnoty)

2. Vypočítejte pro něj pomocí jednoduchého krigování odhad hodnoty a rozptyl odhadu s využitím známých (nebo již vypočítaných) hodnot.

3. Stanovte náhodnou hodnotu z distribuce pravděpodobnosti určené zjištěným průměrem (zjištěný odhad hodnoty) a rozptylem (zjištěný rozptyl odhadu). Výsledek představuje simulovanou hodnotu pro dané místo. Bod se stává místem se známou hodnotou a vstupuje do dalších výpočtů.

4. Opakujte první 3 kroky, až je pokryto celé území.

5. Opakujte první 4 kroky tolikrát, kolik simulací je potřebné provést pro dostatečně věrohodný odhad modelu (např. 100 krát).

6. Ze simulovaných hodnot vypočítejte průměrnou hodnotu a rozptyl (resp. směrodatnou odchylku) pro každé místo.

Volba metody a parametrů

- Průzkumová analýza dat
  - zkoumání distribuce hodnot,
    - statistická (normalita, extrémní hodnoty),
    - prostorová (existence trendu, anizotropie),
  - uspořádání sítě měření.

- Transformace hodnot měření – normální distribuce (případně použití nelineárních technik)

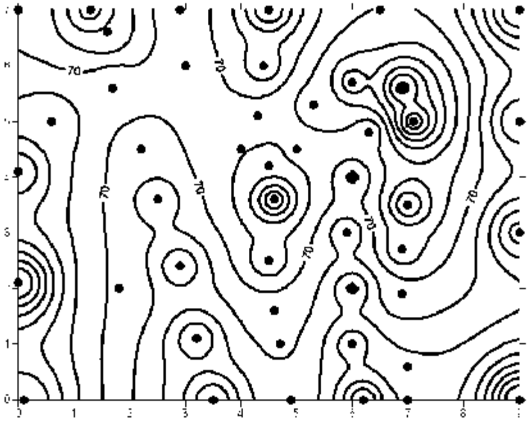
- Eliminace trendu, model trendu, výpočet reziduálních hodnot (zvlášť výpočet základní hodnoty z průběhu trendové funkce a interpolace odchylky)

- Návrh vhodných parametrů pro vyhledávání s ohledem na existující síť (u některých metod nutné pro získání korektního výsledku, jinde pro zrychlení výpočtu):
  - volba sektorů
  - počet bodů
  - dosah hledání
  - tvar prohledávané oblasti (kruh, elipsa)

- Ověření výsledku interpolace (verifikace) – hodnocení rozptylu, srovnání s jinými metodami, s původně naměřenými hodnotami:
    - ručně provedená interpolace
    - bumerangová metoda (cross validation)
    - použití referenční funkce (známý vývoj v ploše)
- mapa chyb (střední chyba)
- změna průměrné hodnoty (rozptylu)
- velikost rozptylu u výsledku
- porovnání s původně naměřenými hodnotami
    - (požadavek exaktní interpolace)
- průběh pole

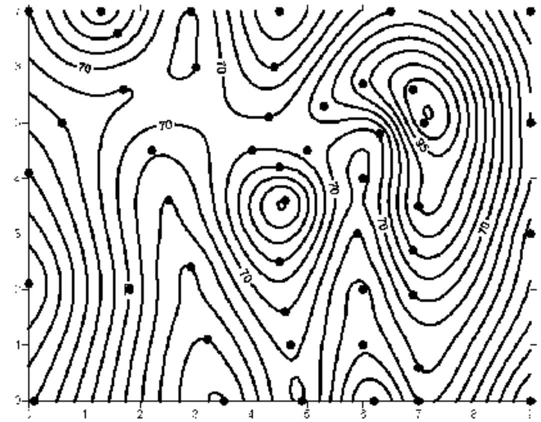
Volba metody a parametrů

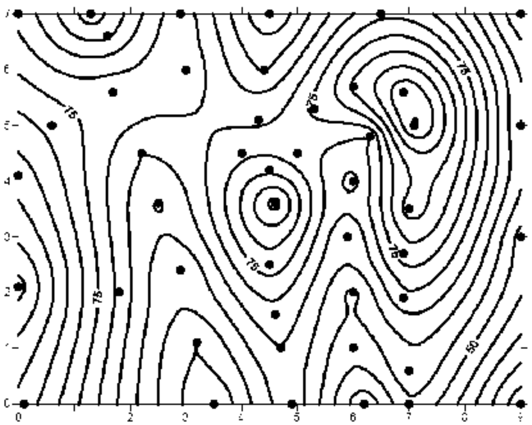
- Na čem tedy závisí výsledek ?
  - cíl interpolace (exaktní interpolace – aproximace – průběh pole)
  - počet a rozmístění známých měření
  - statistických charakteristikách zkoumaného souboru měření
  - vazbě principu zvolené metody a vývoje sledovaného pole
- Nelze spoléhat na 1 metodu !



ID

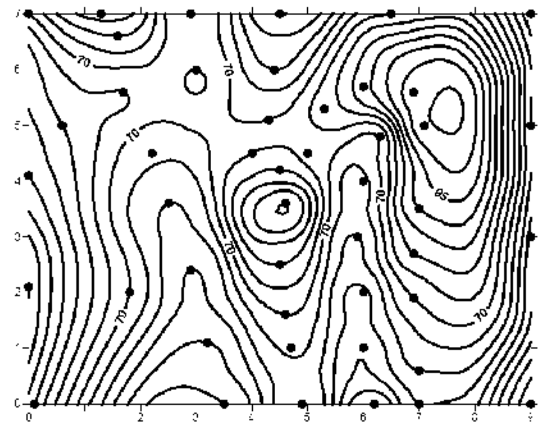
RF

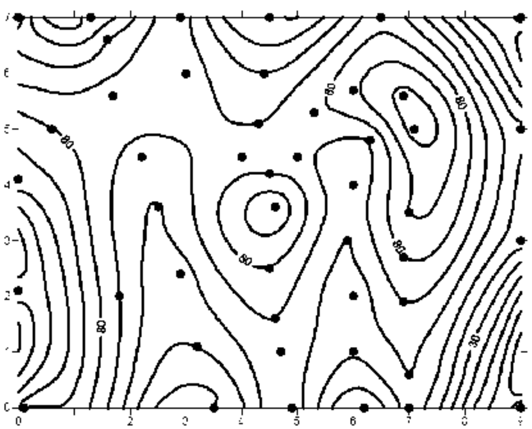




KR

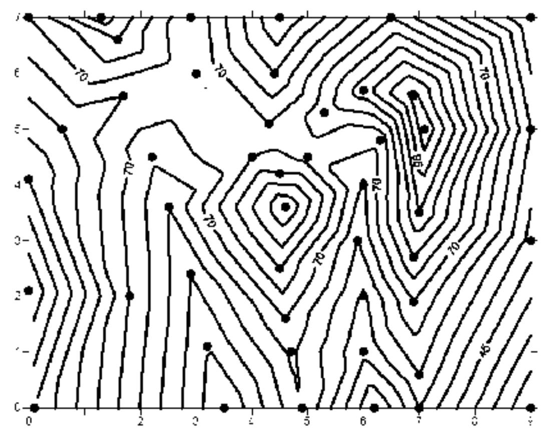
SH

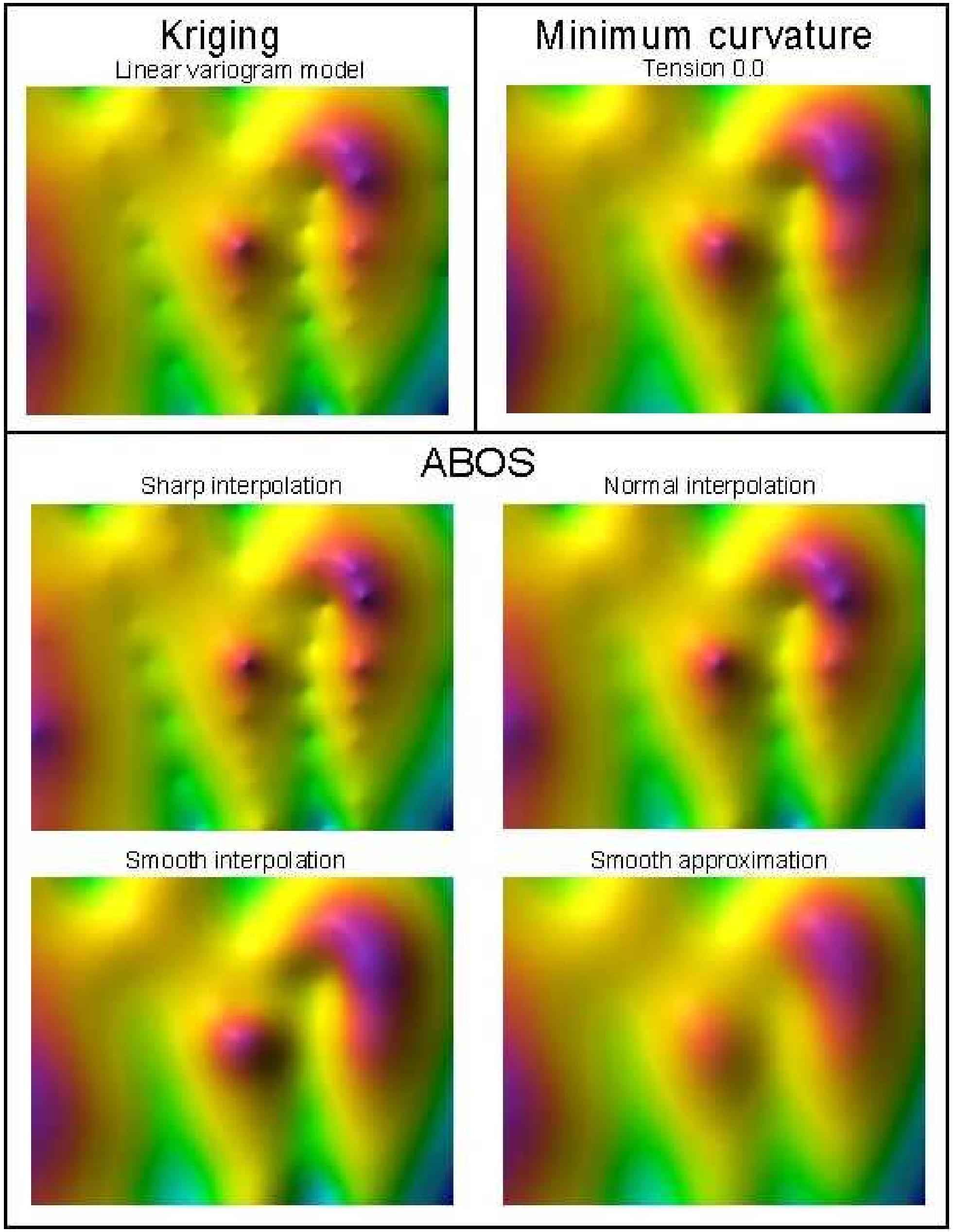


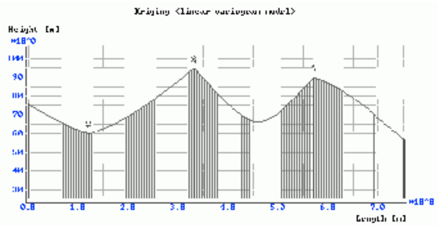


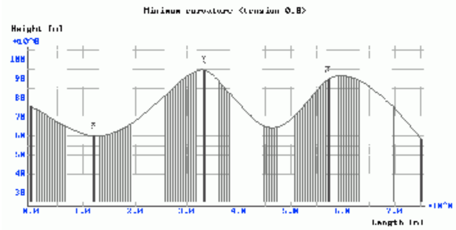
MC

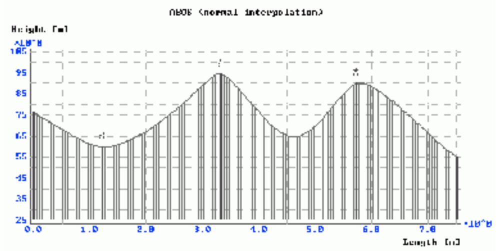
TR

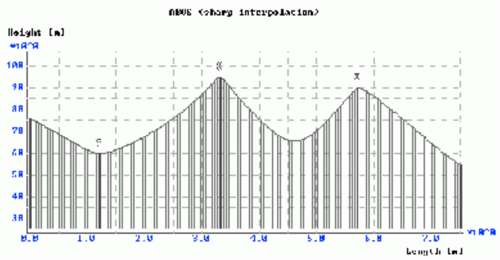


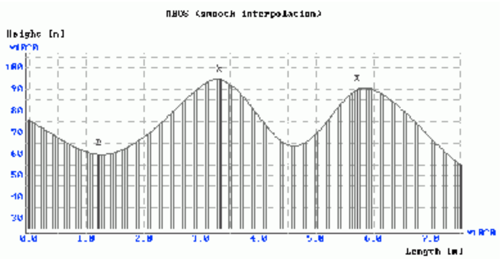

Kriging
Linear variogram model
Minimum curvature
Tension 0.0
ABOS
Sharp interpolation
Normal interpolation
Smooth interpolation
Smooth approximation





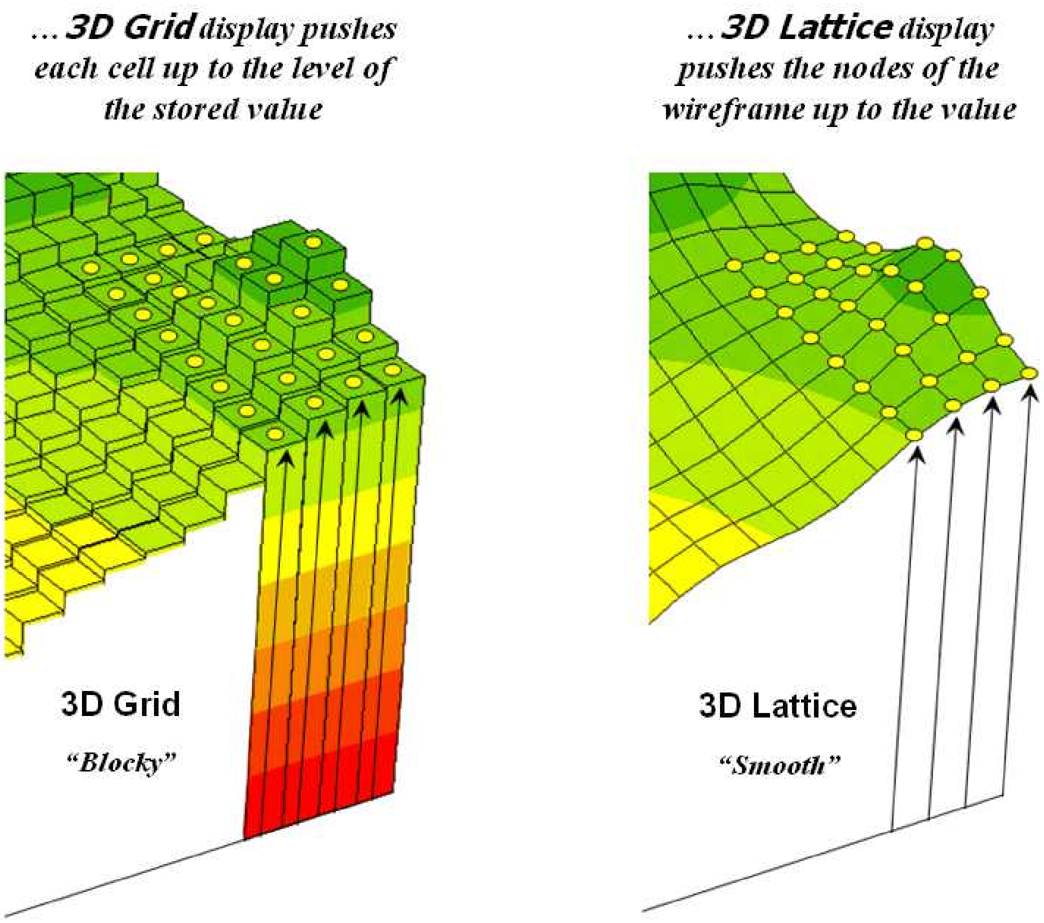






Digitální modely terénu a prostorová interpolace

- Rastrový model



http://www.innovativegis.com/

Polyedrický model

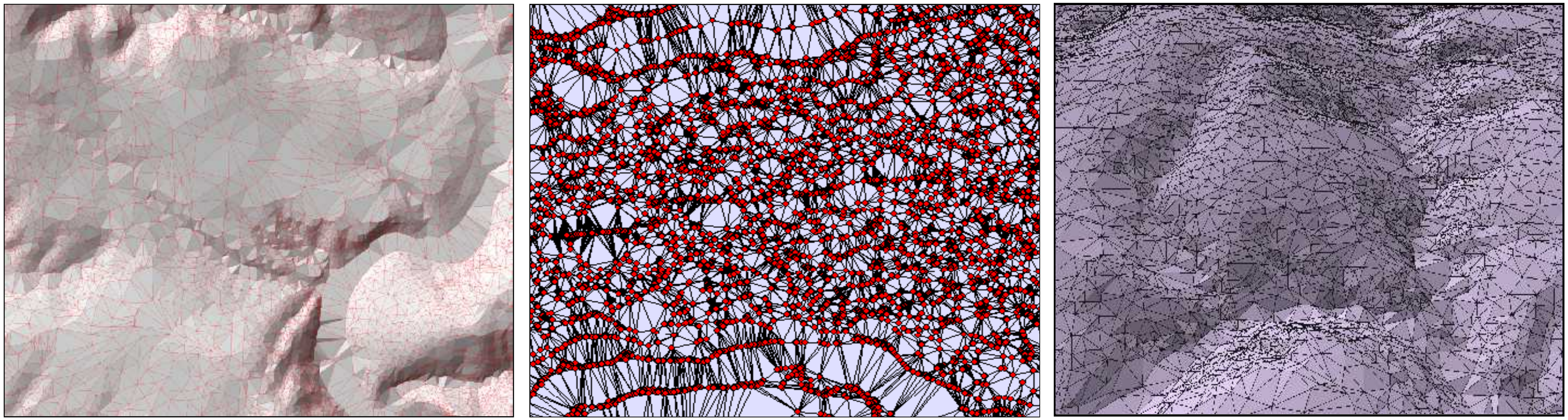
- Elementárními ploškami polyedrického modelu jsou trojúhelníky, které k sobě přiléhají a tvoří tak mnohostěn, přimykající se k terénu.
- Vrcholy mnohostěnu jsou body na terénní ploše, interpolace plochy se obvykle provádí lineárně po trojúhelnících.
- Tento přístup, označovaný jako triangulace či nepravidelná trojúhelníková síť, v originále Triangulated Irregular Network (TIN)*, je v současné době u vektorově orientovaných GIS nejrozšířenější.

---

* Počátky této metody se datují do 70. let 20. století, kdy prof. Thomas Poiker (*Simon Fraser University*) inicioval založení skupiny zabývající se touto problematikou. Původní název byl „*Independent Data of Irregularly Oriented Terrain*“ (IDIOT), protože však někteří oslovení vědci měli výhrady k této zkratce, byla přejmenována na TIN. Zanedlouho potom Randolph Franklin publikoval počítačový algoritmus na vytváření povrchů za využití TIN struktury nazvaný TIN 73 (Thurston 2003).

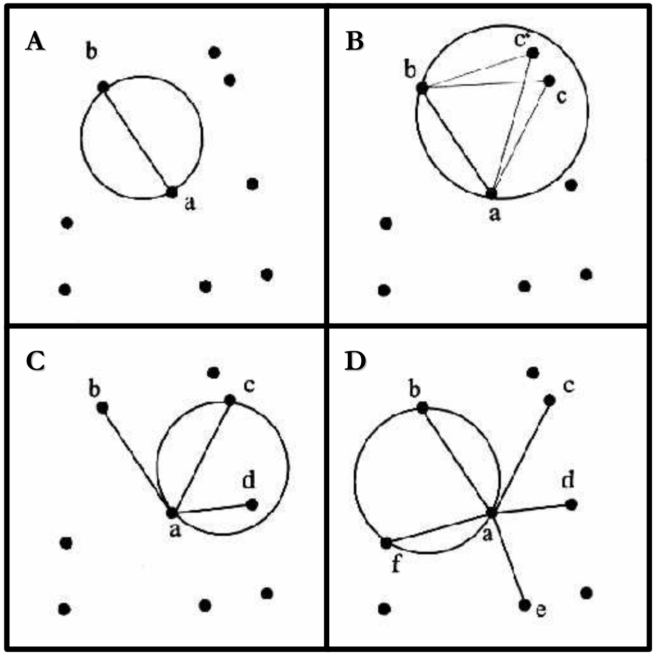
Terén je v tomto případě reprezentován trojúhelníky, čili sadou vrcholů (vertices, vertex), hran (edges) a plošek (faces). Vstupní zdrojová data jsou vyjádřena ***X***, ***Y*** a ***Z*** souřadnicemi vrcholů, každá hrana tak spojuje 2 vrcholy a rozděluje 2 plošky, každá trojúhelníkovitá ploška je potom ohraničena 3 hranami a je považována za rovinný útvar (Kraus 2000).

V některých případech však nejsou z hlediska přesnosti rovinné trojúhelníky dostačujícím řešením a jejich povrch může být modelován pomocí specifických algoritmů jako jsou například Bezierův plát (Pfeifer and Pottmann 1996) nebo Coonsova plocha.



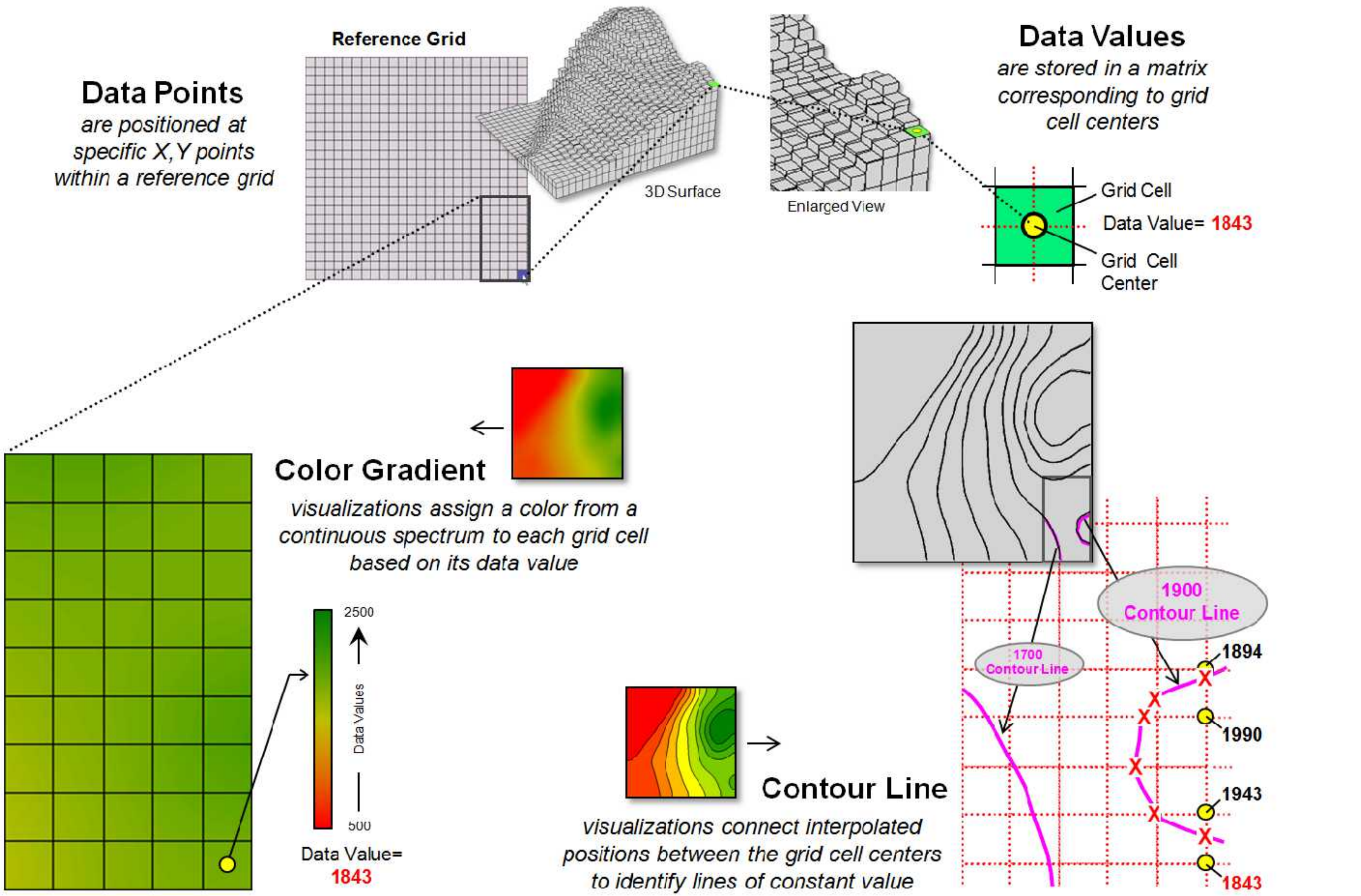
Geometrie vektorových modelů typu TIN

Nejdůležitější fází pro tento model je samotný typ triangulace, která se týká pouze rovinných souřadnic vstupních bodů. Nejčastěji používaným kritériem je tzv. **Delaunayho podmínka**, tedy, že vnitřek kruhu opsaného libovolnému z trojúhelníků sítě neobsahuje žádný další (čtvrtý) bod sítě (Preparata and Shamos 1990). Toto kritérium je zároveň ekvivalentní pro maximalizaci velikostí všech úhlů v trojúhelníku (Heitzinger 1996).

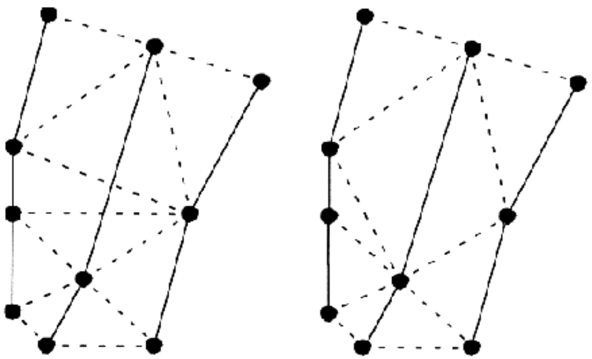


Postup triangulace

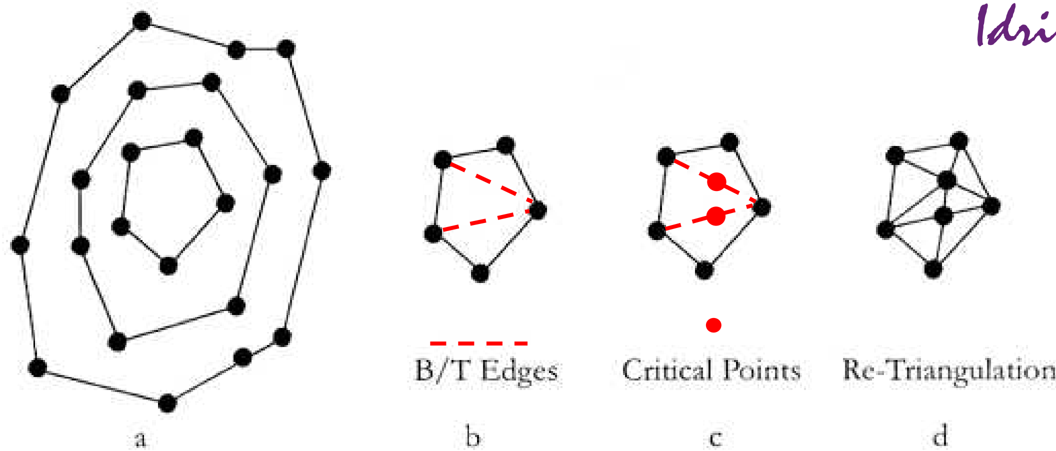
Postup triangulace znázorňuje obrázek (Raper 1993). K danému bodu ***b*** najdeme nejbližšího souseda ***a***, tzv. Thiessenovy sousedy (A). Rozšíříme hledací kružnici procházející přes tyto body tak, aby do ní padly další body (B). Vybereme z nich ten, u kterého úhel ke spojnici východiskových bodů ***ab*** je větší, tedy ***c*** (C). Pokračujeme v hledání do té doby, až v kružnici opět najdeme bod ***b*** (D).



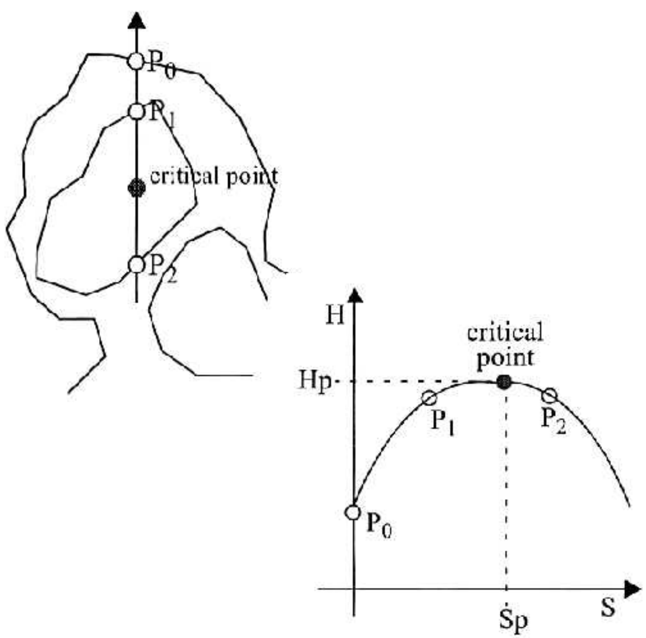
http://www.innovativegis.com/

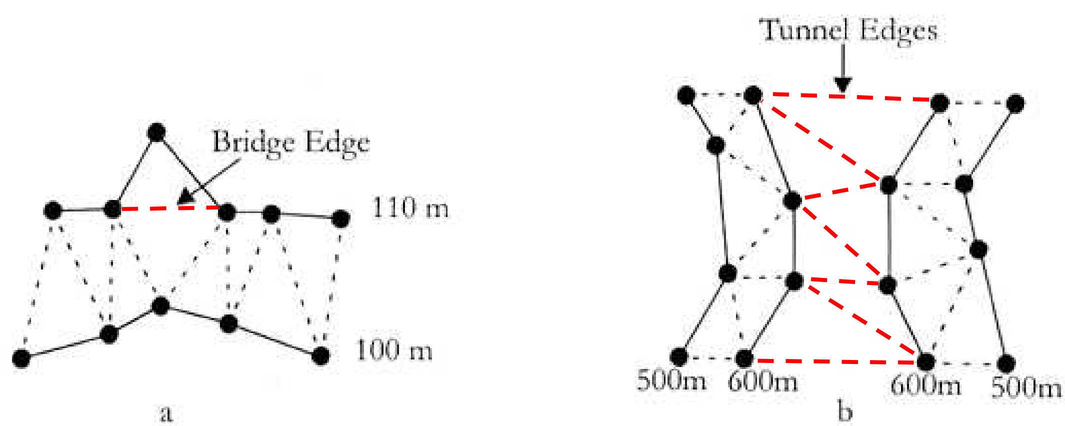


Unconstrained (left) and constrained (right) Delaunay triangulations. Solid lines represent isolines.



a: Contours at the top of a hill; b: triangulation of highest contour, with B/T edges identified; c: placement of critical points on B/T edges;





a: Contours at a stream; b: contours at a "saddle" feature. Contours are shown with solid lines, constrained triangle edges with dashed lines. B/T edges are shown in red.

IN struktura v GIS Idrisi

Srovnání datových reprezentací DMT

	Rastrové modely	Triangulace
Aproximace povrchu	Velmi variabilní (závisí na typu zvoleného algoritmu prostorové interpolace dat).	Obtížně modifikovatelná (vstupní data se stávají vrcholy trojúhelníků jejichž plochy jsou nejčastěji považovány za rovinné).
Geomorfologie	Nutné zavedení zlomových linií je relativně obtížné – definováním oblastí bez aproximace povrchu nebo zavedením hybridního modelu.	Nutné zavedení zlomových linií je relativně jednoduché – definováním zlomových linií jako povinných hran trojúhelníkovité sítě.
Vstupní data	Omezení vyplývající z maticové struktury modelu, obvyklé dilema mezi velikostí gridu (pixelu) odrážející se ve velikosti dat a přesnosti. Ideální pro extrémní objemy pravidelně uspořádaných dat (laserové skeno-vání).	Relativně neomezené parametry, závislé pouze na kvalitě trojúhelníkovité sítě vzhledem k rozmístění (hustotě) vstupních dat.
Použitelnost	Omezení vyplývající z 2,5D repre-zentace. Velké množství relativně jednoduchých algoritmů pro mnoho aplikací.	Struktura modelu je svou podstatou jednodušší. Algoritmy pro aplikace jsou však složitější vzhledem k výpočtům v topologickém modelu.