

Metody prostorové interpolace

Základní pojmy

Interpolace – skupina metod, které slouží k odhadu neznámých hodnot proměnné v jistých bodech (neměřených) na základě hodnot proměnné v bodech měřených.

Prostorová interpolace – skupina metod, které slouží k vytváření spojitého povrchu (polí) z bodových měření.

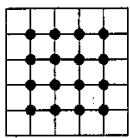
Body mohou být lokalizovány v 1, 2 i 3 rozměrném prostoru. Interpolace se může týkat nejenom bodů, ale i linií a ploch.

Extrapolace – odhad hodnot proměnné vně oblasti definované krajními body měření.

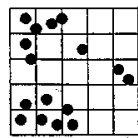
Naprostá většina interpolačních postupů je založena na principu **prostorové autokorelace** – tedy na předpokladu, že hodnoty odhadované veličiny v lokalitách blízkých si budou více podobné než hodnoty v lokalitách vzdálených.

Výběr reprezentativních vzorků (sampling)

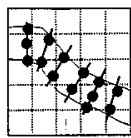
Je důležitý pro výběr interpolačního algoritmu a úspěšnost vlastní interpolace.



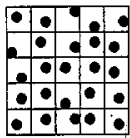
a) regular sampling



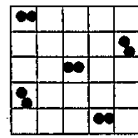
b) random sampling



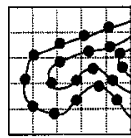
c) transect sampling



d) stratified random sampling



e) cluster sampling



f) contour sampling

Další aspekty ovlivňující úspěšnost interpolace

- způsob prezentace spojitého povrchu (grid, TIN, izočáry, areály)
- dostupné datové zdroje pro interpolaci
- vymezení studované plochy – přirozené a administrativní hranice
- dostupnost bodů měření vně studované plochy

Předpoklady úspěšné prostorové interpolace

- existence dostatečně reprezentativního vzorku měřených dat
- vhodné vlastnosti měřené veličiny a typ dat (ordinální, intervalová, poměrová)
- teoretické i empirické znalosti o povaze prostorové diferenciace studovaného jevu
- znalost podstaty použitelných interpolačních metod
- znalost způsobu výběru nevhodnější metody

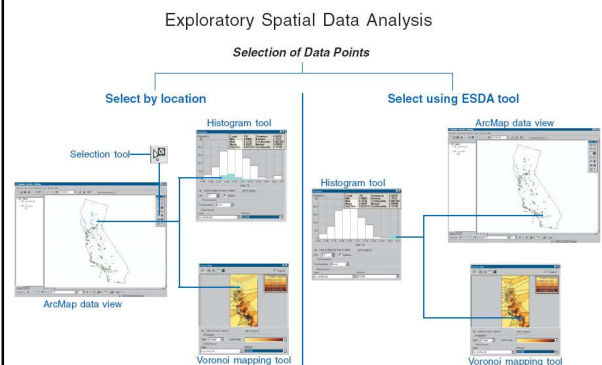
Průzkumová analýza prostorových dat (ESDA).

- **EDA** – Exploratory Data Analysis
- **ESDA** – Exploratory Spatial Data Analysis
- **ESTDA** – Exploratory Spatio – Temporal Data Analysis

• Množina statistických metod a speciálních nástrojů, zvláště grafických metod, používaných k lepšímu porozumění datům, k odhalení jejich důležitých vlastností.

• Jejím cílem je zjistit základní informace o charakteru vstupních dat v tomto případě za účelem následné interpolace.

Průzkumová analýza prostorových dat (ESDA).



Základní postupy průzkumové analýzy prostorových dat

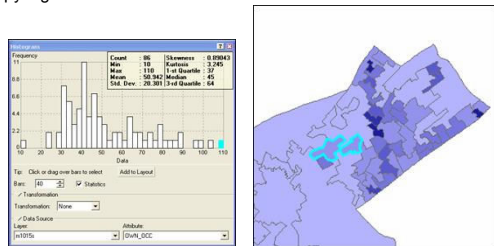
- výpočet základní popisné statistiky včetně momentů vyššího řádu (asymetrie a špičatosti)
- prověření požadavků normality a stacionarity
- analýza rozdělení hodnot - analýza histogramu
- analýza kvantilového grafu (Q-Q grafu)
- zkoumání odlehlých hodnot a jejich případné odstranění
- analýza trendu a jeho případné odstranění
- případná transformace vstupních dat (log)

Základní postupy průzkumové analýzy prostorových dat

- ESDA slouží k průzkumu, deskripci, vizualizaci, zvyrazňování základních rysů dat, jejich distribuce (nejen ve smyslu prostorovém).
- Postupy ESDA slouží k prověření požadavků normality, stacionarity vstupních dat.
- Používá specifických nástrojů (histogram, box plot, scatter plot, Q-Q graf) i statistických charakteristik
- Postupy ESDA mohou vést k nutnosti **úpravy** či **transformace** původních dat.
- Úprava může spočívat v odstranění trendu či odlehlých hodnot, transformace potom např. například v tzv. log-transformaci.
- ESDA je nezbytným předstupněm úspěšné aplikace řady interolačních postupů (např. metod krigingu).

Základní nástroje průzkumové analýzy prostorových dat

Popisná statistika a „mapped histogram“ - propojení mapy a grafu



Hodnocení polohy a prostorového uspořádání typických resp. extrémních hodnot.

Základní nástroje průzkumové analýzy prostorových dat

Voronoi map

- definování přirozených sousedů k vyšetřovanému bodu
- výpočet lokální statistiky (od měr úrovně až po míry entropie)

Základní nástroje průzkumové analýzy prostorových dat

Voronoi map

Entropie – je počítána z hodnot daného polygonu a všech polygonů sousedních. Nejprve jsou všechny polygony rozříděny do pěti tříd.

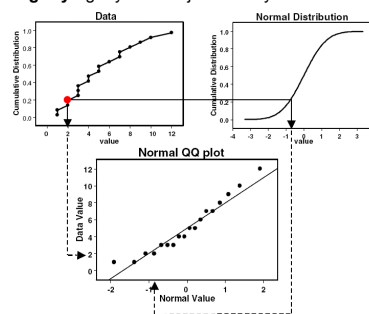
$$Entropie = -\sum p_i * \text{Log}_2 p_i$$

kde p_i je poměr polygonů náležejících do dané třídy z celkového počtu polygonů

- Minimální entropie** – všechny buňky patří do stejné třídy
- Maximální entropie** – každá z buněk náleží k jiné třídě.

Základní nástroje průzkumové analýzy prostorových dat

Kvantilové grafy - grafy zobrazující kvantily dvou rozdělení

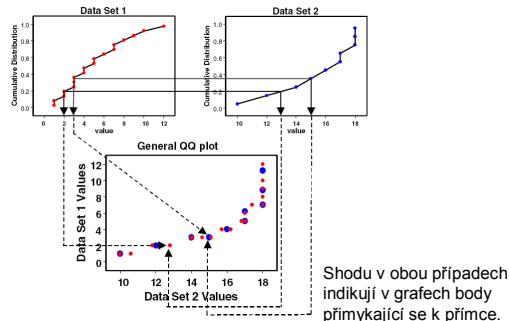


Normální Q-Q graf

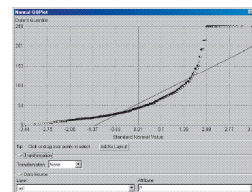
Slouží jako nástroj k posouzení normality vstupních dat.

Základní nástroje průzkumové analýzy prostorových dat

Obecný Q-Q graf – testuje se podobnost rozdělení dvou datových sborů, vynáší se odpovídající si hodnoty kvantilů



Základní nástroje průzkumové analýzy prostorových dat



Takováto data vyžadují transformaci. Základní typy transformací:

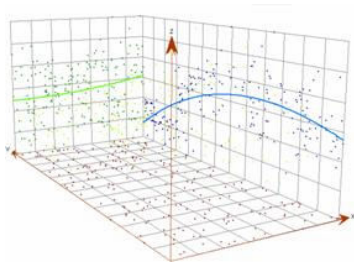
- Box-Cox
- Arcsine
- Logaritmická

Základní nástroje průzkumové analýzy prostorových dat

Analýza trendu

Definování globálního trendu v datech, jeho odhalení a eventuálního odstranění.

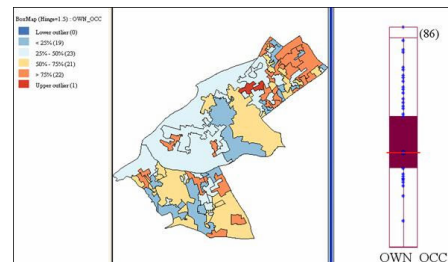
Spočívá v projekci hodnot vyšetřovaných bodů do rovin xz a yz a jejich proložení polynomem n-tého řádu.



Základní nástroje průzkumové analýzy prostorových dat

Krabicové grafy (box plots)

- detekce odlehlých či extrémních hodnot
- lokální a globální odlehlé hodnoty



Základní nástroje průzkumové analýzy prostorových dat

Vykreslení množiny hodnot semivariance či covariance

- Detekce míry prostorové autokorelace, vystižení míry anizotropie, odhlaení odlehlých hodnot.
- **Semivariance (semivariogram)** – empirický semivariogram jako graf míry nepodobnosti.
- V úlohách interpolace je tato veličina důležitá pro objektivní definování **velikosti a tvaru okolí** vyšetřovaného bodu.

Základní nástroje průzkumové analýzy prostorových dat

Vykreslení množiny hodnot semivariance či covariance

Polovina ze sumy čtverců rozdílů hodnot všech dvojic vyšetřovaných bodů vzdálených o určitou hodnotu.

Hodnota empirické **semivariance** proměnné z pro dvojici bodů v poloze x_i a x_j :

$$0,5 * (z(x_i) - z(x_j))^2$$

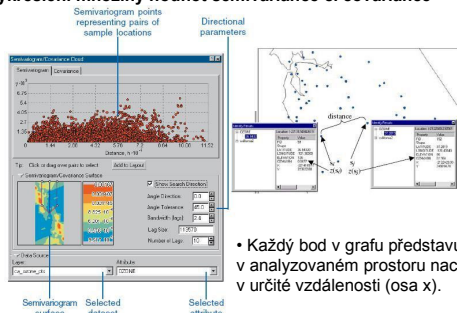
Hodnota empirické **covariance**

$$(z(x_i) - \bar{z})(z(x_j) - \bar{z})$$

Hodnota empirické **crosscovariance**

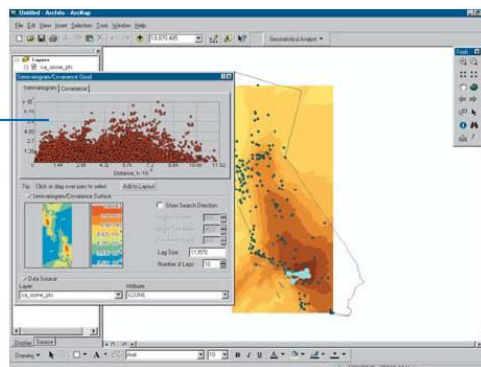
$$(z(x_i) - \bar{z})(y(t_j) - \bar{y})$$

Základní nástroje průzkumové analýzy prostorových dat Vykreslení množiny hodnot semivariance či covariance

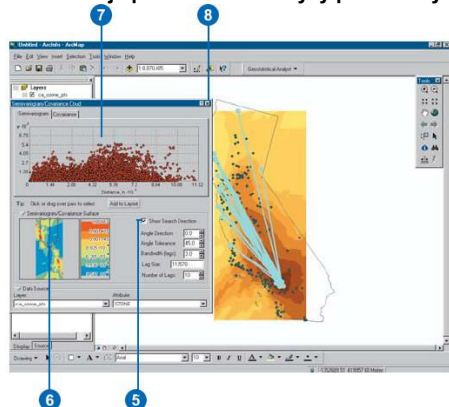


- Každý bod v grafu představuje dvojici bodů v analyzovaném prostoru nacházejících se v určité vzdálenosti (osa x).
- Podobnost hodnot interpolované veličiny je vyjádřena semivariancí (osa y).

Základní nástroje průzkumové analýzy prostorových dat

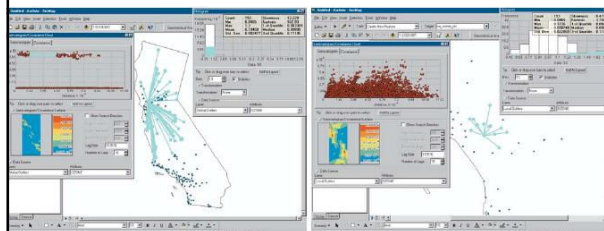
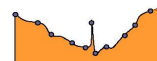


Základní nástroje průzkumové analýzy prostorových dat



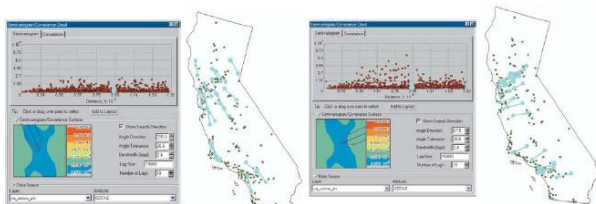
Základní nástroje průzkumové analýzy prostorových dat Detekce odlehlých hodnot (outliers)

- Základní nástroje:
- histogram
 - semivariogram/ covariance cloud
 - Voronoi map



Detekce globální (vlevo) a lokální (vpravo) odlehlé hodnoty.

Základní nástroje průzkumové analýzy prostorových dat Vyšetřování tvaru okolí – izotropní a anizotropní povrch



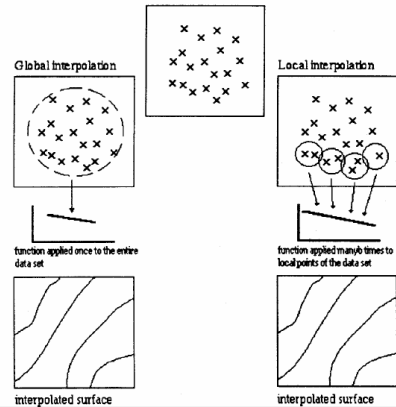
Detekce globální (vlevo) a lokální (vpravo) odlehlé hodnoty.

Explorační analýza prostorových dat (ESDA).

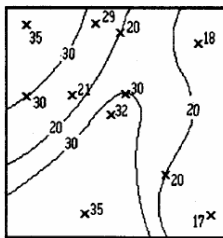
Rozdělení metod prostorové interpolace

- metody interpolace bodů, linií a ploch.
- metody lokální a globální
- metody exaktní a aproximující
- metody spojité a zlomové (abrupt)
- metody deterministické a stochastické

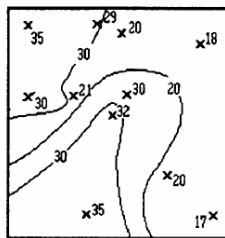
Globální a lokální metody interpolace



Exaktní a aproximující metody interpolace

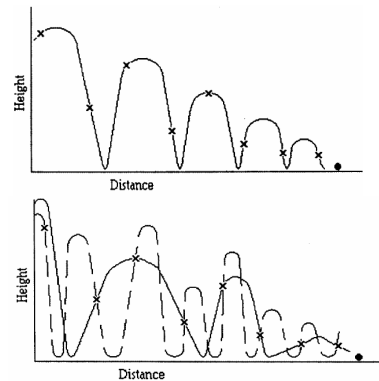


An exact surface which honours the observed values



An approximate surface which does not honour all the observed values

Deterministické a stochastické metody interpolace



Metody analogové interpolace (line threading or eye balling)

- vytváření izolinií na základě spojování míst s obdobnými hodnotami jevu založené na expertním odhadu
- využívají empirie, obecné teorie a znalosti místních zvláštností
- expertní systémy

Základní omezení (s ohledem na počítačové zpracování):

- problém zpracování velkého množství bodů
- problém subjektivního přístupu
- problém časové náročnosti

Globální interpolátory využívající klasifikačních modelů a ANOVA analýzy

Stacionarita - předpoklad, že míry úrovně a variability výběrového souboru nezávisí na velikosti výběru a rozmístění jednotlivých měřených bodů

Klasifikace homogenními polygony - k interpolaci v rámci studovaného území využít externě definovaných prostorových jednotek (regionů).

Obecný model

$$z(x_0) = \mu + \alpha_k + \varepsilon$$

z - hodnota atributu v lokalitě x_0

μ - celkový průměr atributu na zpracovávaném území

α_k - odchylka mezi μ a průměrem v regionu k

ε - reziduum, šum

Model předpokládá, že v rámci každého regionu (třídy) k mají hodnoty interpolovaného atributu normální rozdělení. Průměrný atribut pro třídu k je roven:

$$\mu + \alpha_k$$

a je určen z výběrových měření v rámci třídy k .

Uvedený přístup vychází z několika předpokladů:

- kolísání hodnot z v rámci jednotlivých tříd je náhodné
- měřená hodnota v rámci každé mapované třídy se vyznačuje stejně velikou náhodnou složkou ε
- studované atributy mají normální rozdělení
- veškeré prostorové změny se dějí na hranicích mezi jednotlivými třídami, změny se dějí skokem, ne postupně

Globální interpolátory využívající analýzy trendu

Princip - mnohonásobná regrese hodnot atributu vs. geografické souřadnice.

Metodou nejmenších čtverců jsou nalezeny nejhodnější koeficienty pro daný polynom n -tého řádu.

Předpokládá se normální rozdělení.

lineární trend:

$$z = b_0 + b_1x + b_2y$$

kvadratický trend:

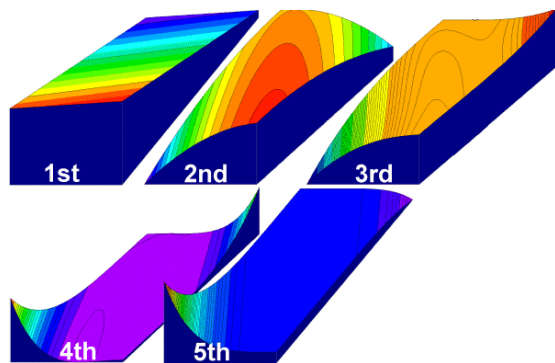
$$z = b_0 + b_1x + b_2y + b_3x^2 + b_4xy + b_5y^2$$

kubický trend:

$$z = b_0 + b_1x + b_2y + b_3x^2 + b_4xy + b_5y^2 + b_6x^3 + b_7x^2y + b_8xy^2 + b_9y^3$$

b – koeficienty, x, y – souřadnice bodů

Interpolace trendové složky polynomy 1 až 5 stupně



Globální interpolátory využívající regresní analýzy

Princip - existuje vazba mezi hodnotami interpolované veličiny a vybranými jinými atributy studovaného prostoru (např. teplota a nadmořská výška, koncentrace znečištění a vzdálenost od zdroje).

Forma - empirický model závislosti interpolované veličiny na hodnotách jedné či několika veličinách nezávislých:

$$z(x) = b_0 + b_1P_1 + b_2P_2 + \varepsilon$$

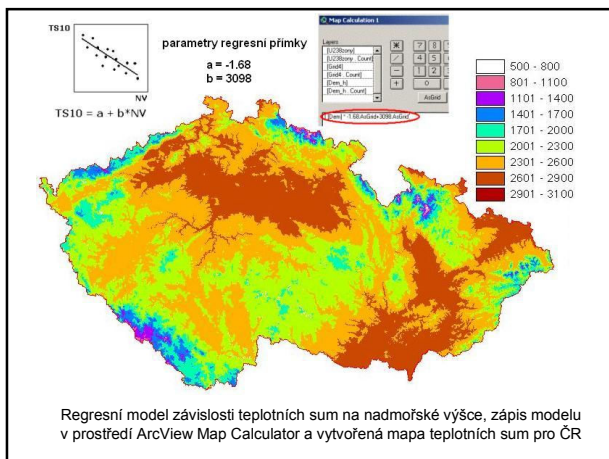
$b_0 \dots b_n$ - regresní koeficienty

$P_1 \dots P_n$ - nezávislé proměnné

Sestavení regresní závislosti je založeno na metodě nejmenších čtverců.

Výsledný model může být lineární i nelineární.

Jako nezávislé proměnné lze kombinovat geografické souřadnice s jinými atributy.



Metody lokální interpolace (lokální interpolátory)

Globální interpolátory - lokální efekty = náhodný šum

Lokální interpolátory - hledaná hodnota je určena z určitého počtu měření z předem definovaného okolí počítaného bodu.

Obecný postup se sestává z následujících kroků:

1. definování velikosti a tvaru zájmového okolí
2. nalezení měřených bodů v tomto okolí
3. nalezení matematické funkce vystihující kolísání hodnot nacházejících se v okolí daného bodu
4. výpočet hodnoty pro uzly regulérní sítě (grid)

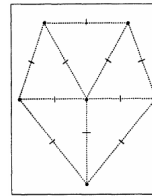
Pro lokální interpolace jsou důležité následující skutečnosti:

- druh použité interpolační funkce
- velikost, tvar a orientace okolí
- počet bodů v okolí zahrnutých do výpočtu
- rozložení uvažovaných bodů (regulérní či nepravidelné)
- možné začlenění externí informace např. o obecném trendu

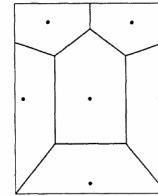
Metoda nejbližšího souseda (thiessenovy polygony)

Princip - hodnoty atributů v neměřených místech jsou určeny z hodnot nejbližšího místa měřeného.

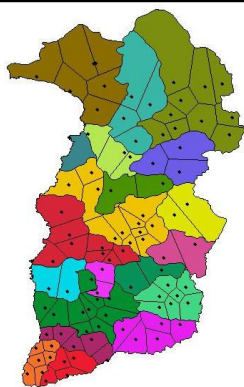
Zpracovávané území rozděleno na nepravidelné trojúhelníky (Delaunay triangulace) a z nich jsou poté definovány tzv. thiessenovy polygony.



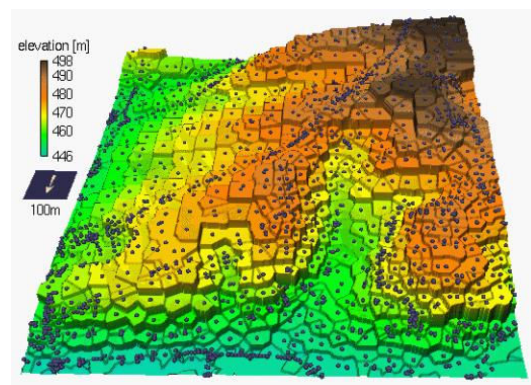
Neighbouring points identified and territory markers established at the half way point along the line connecting pairs of points



Bisecting lines drawn through territory markers and a Thiessen polygon mosaic established by linking the bisecting lines



Příklad interpolace množiny nepravidelně rozmístěných bodů v ploše metodou thiessenových polygonů



Příklad interpolace spojité veličiny metodou thiessenových polygonů

Metody konstrukce nepravidelných trojúhelníků (TIN)

- Exaktní metoda vhodná pro nepravidelně rozmístěné body měření.
- Body jsou spojeny liniemi a vytváří síť nepravidelných trojúhelníků.
- Hodnoty v bodech na počátku a konci linií jsou známy, lze použít jednoduchou lineární závislost k interpolaci bodů mezi dvěma body na linii.
- TIN je metoda interpolace i způsob vizualizace spojitých povrchů.
- Metoda vhodná pro povrchy vyznačující se náhlými změnami spádu (fluvialně erodované povrchy).

Proces vytváření spojitého povrchu metodou TIN zahrnuje:

- výběr charakteristických bodů (ne z jakékoliv množiny nepravidelně rozmístěných bodů lze vytvořit TIN)
- způsob propojení bodů do trojúhelníkové sítě
- způsob modelování povrchu uvnitř trojúhelníků

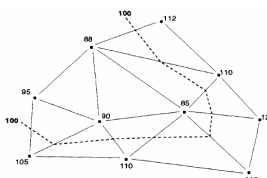
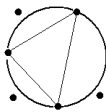
Výběr bodů a algoritmy pro výběr bodů:

- algoritmus Fowler and Little
- VIP algoritmus
- Drop heuristic algoritmus

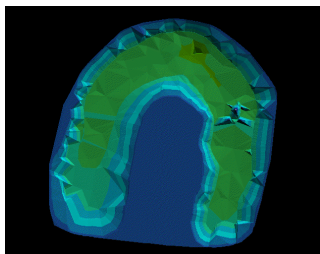
<http://www.ncgia.ucsb.edu/giscc/units/u056/>

Způsob propojení bodů do TIN - Delaunay triangulace:

Metody není možné použít k extrapolaci – výsledný povrch má plochu, která vznikne spojením vnějších měřených bodů („hull“).



TIN je model vhodný k následné konstrukci izoliní.



Metoda inverzní vzdálenosti

Princip - hodnota atributu v určitém bodě je váženým aritmetickým průměrem hodnot okolních měřených bodů.

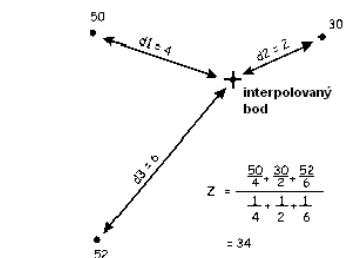
Váhy jsou určeny pro každý bod jako inverzní vzdálenost měřeného bodu od bodu interpolovaného.

Obečný vzorec pro odhad hodnoty Z:

$$\hat{Z} = \frac{\sum_{i=1}^n w_i z_i}{\sum_{i=1}^n w_i}$$

Váhy se určují ze vztahu: $w = \frac{1}{d^k}$ nebo $w = e^{-kd}$

Hodnoty vah w_i představují funkci vzdálenosti d . Hodnota exponentu k se nejčastěji volí 1 či 2.



Odhad hodnoty v bodě metodou inverzní vzdálenosti

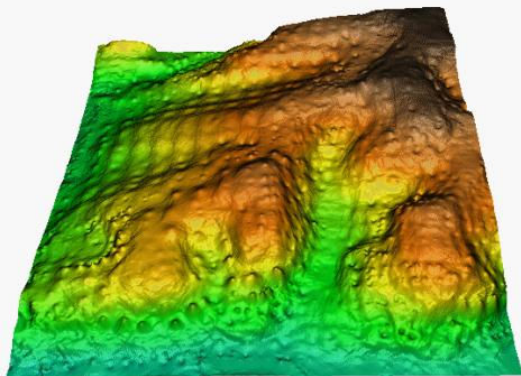
Metoda inverzní vzdálenosti efekt „průměrování“- potlačení lokálních extrémů



Problém generování koncentrických struktur kolem interpolovaných bodů (tzv. „bulls eyes“)

Způsob definování okolí

- > izotropní povrch - kruhové okolí interpolovaného bodu, pro odhad hodnoty bereme všechny body bez ohledu na směr
- > anizotropie - body v jistém směru mohou mít na interpolovanou hodnotu jinou váhu než ve směru jiném - okolí tvaru elipsy
- > minimální a maximální počet bodů pro výpočet nové hodnoty
- > rozmístění bodů v rámci definovaného okolí (kvadranty, oktanty)
- > IDW je senzitivní na shluky měřených bodů a také na odlehle hodnoty



Příklad interpolace spojité veličiny metodou inverzní vzdálenosti

Prostorové klouzavé průměry

Modifikace metody IDW

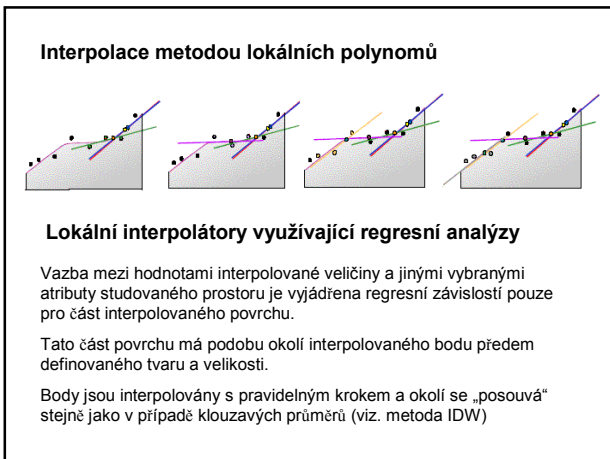
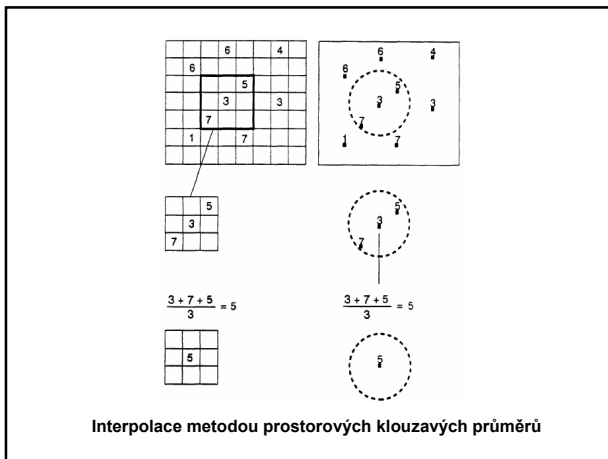
Nová hodnota může být prostým průměrem, váženým průměrem, modální hodnotou.

Definování velikosti, tvaru a charakteru okolí.

Počet bodů v okolí (min, max) - 4 až 12 bodů, optimum 6 až 8 bodů.

Metody je vhodné použít za těchto podmínek:

- > existuje nejistota s ohledem na reprodukovatelnost výsledků opakovaných měření v daném bodě (vlastní proměnlivost pole hodnot měření)
- > samotná technická stránka měření je zatížena jistou chybou
- > je známo, že skutečné prostorové pole daného jevu vykazuje kromě obecného trendu také lokální variabilitu.



Splínové funkce

Matematicky definované křivky, které po částech a exaktně interpolují jednotlivé body povrchu, jsou lokálními interpolátory.

Zajišťují kontinuální spojení jednotlivých částí interpolovaného povrchu.

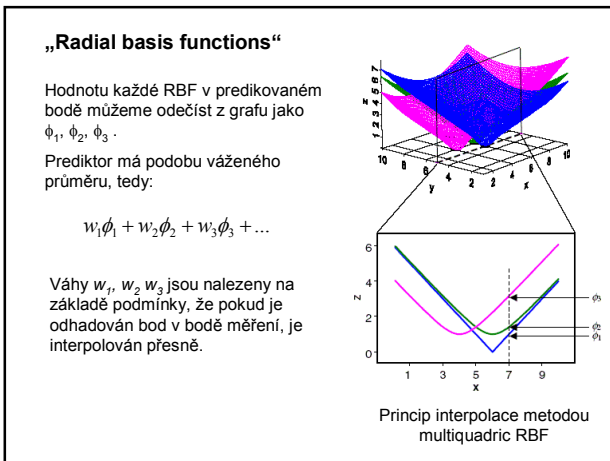
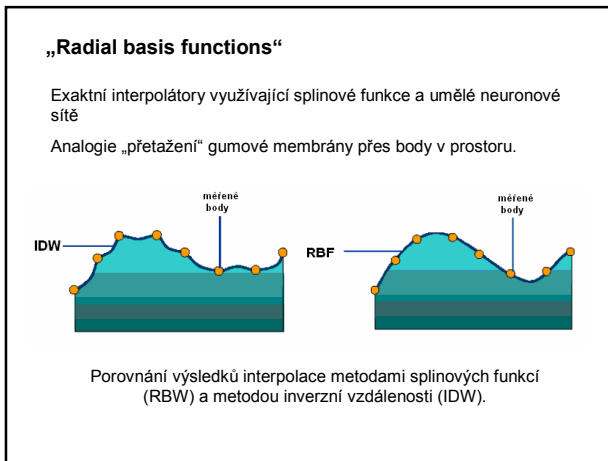
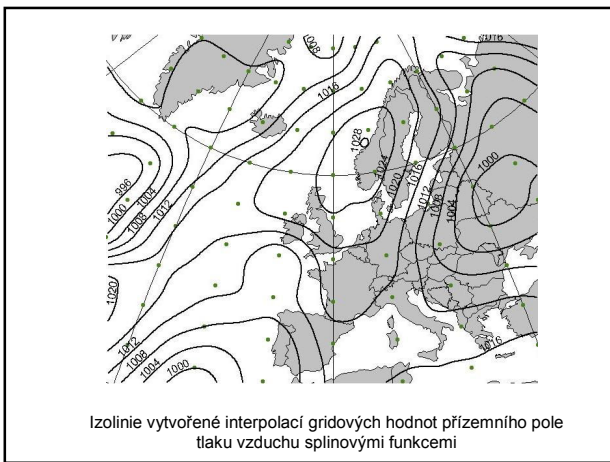
Lze modifikovat část povrchu bez přepočtu celého povrchu (toto neumožňují trendy).

Pro interpolování linií se používá tzv. **kubických splínů**, pro interpolování povrchů se využívá jejich 2D analogie označované jako „**thin plate splines**“

Nahrazují části povrchů interpolované přesným splínem lokálně shladenou průměrnou hodnotou.

Povrch je interpolován tak, aby procházel co nejbližší měřeným bodům a také aby zachoval **podmínku minimální křivosti**.

Interpolované povrchy jsou často značně shladené, jsou vhodné pro interpolaci jevů, které se mění spjitě.



Multiquadric RBF

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

$B_i(x, y)$ – radiální funkce vzdálenosti $d_i(x, y)$

$d_i(x, y)$ – relativní vzdálenost měřeného bodu v místě x, y , od místa odhadu x, y

R^2 – vyhlazovací parametr

Pro funkce $B_i(x, y)$ jsou během výpočtu v každém interpolovaném bodě stanovovány váhy řešením soustavy lineárních rovnic.

Čím větší je hodnota R , tím více je shlazený je výsledný interpolovaný povrch.

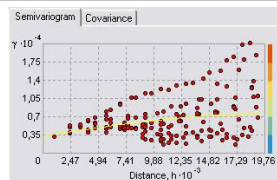
„Radial basis functions“

Parametry konkrétní interpolující funkce jsou optimalizovány výpočtem chyby RMSPE.

RBF jsou exaktní metodou a jsou vhodné pro hladké povrchy generované z velkého počtu bodů (např. modely terénu).

Nehodí se pro interpolaci jevů, které se výrazně mění v prostoru a dále pro interpolaci jevů, u nichž existuje jistá míra nejistoty ohledně přesnosti měřených bodů.

Kriging



Lokální metody interpolace, které optimalizují výběr bodů okolí, ze kterých je odhadována nová hodnota.

K této optimalizaci se provádí tzv. **strukturní analýza** založená na studiu tzv. strukturních funkcí – např. semivariogramu.

Semivariogram z empiricky zjištěných dat je nahrazen teoretickým modelem a parametry tohoto modelu jsou použity ve vlastním krigování.

Kriging je založen na odhadu závislosti průměrné změny v hodnotách studované veličiny a vzdálenosti měřených bodů.

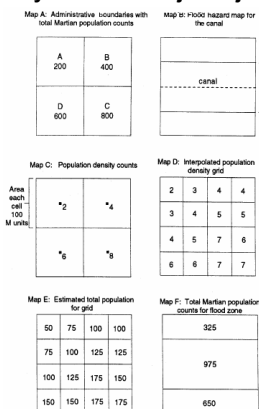
Metody prostorové interpolace ploch

Mnoho jevů se vztahuje k plošným jednotkám spíše než k bodům (hustota obyvatelstva, kvalita pitné vody...).

Metody řeší způsob, jakým lze odhadnout hodnoty jistého jevu na základě hodnot jiného jevu vázaných na plošné jednotky.

- plošné jednotky se shodují
- zdrojové jednotky jsou podmnožinou jednotek výstupních
- metody zachovávající objem studovaného jevu (volume preserving)
- metody nezachovávající objem studovaného jevu (non-volume preserving)

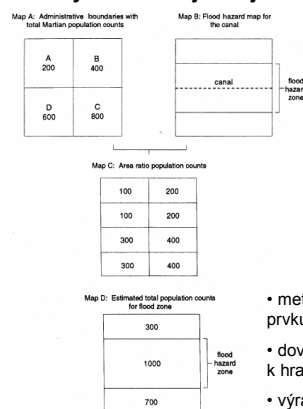
Metody nezachovávající objem studovaného jevu



Jaká je hustota obyvatelstva uvnitř záplavové zóny?

1. výpočet hustoty obyvatelstva pro každou plochu
2. určení centroidu každé plochy
3. interpolace hustoty obyvatelstva výše popsány metodami

Metody zachovávající objem studovaného jevu



„pycnophylatic method“

1. Proveďte se překrytí cílových zón (oblastí) přes oblasti zdrojové.
2. Určete poměrnou část cílové zóny, která spadá do zóny zdrojové.
3. Celková hodnota atributu v cílové zóně je určena v závislosti na plošném zastoupení zón zdrojových.

- metoda zachovává sumu studovaného prvku
- dovoluje kontinuální změnu směrem k hranicím každé třídy.
- výrazně mění min a max hodnoty