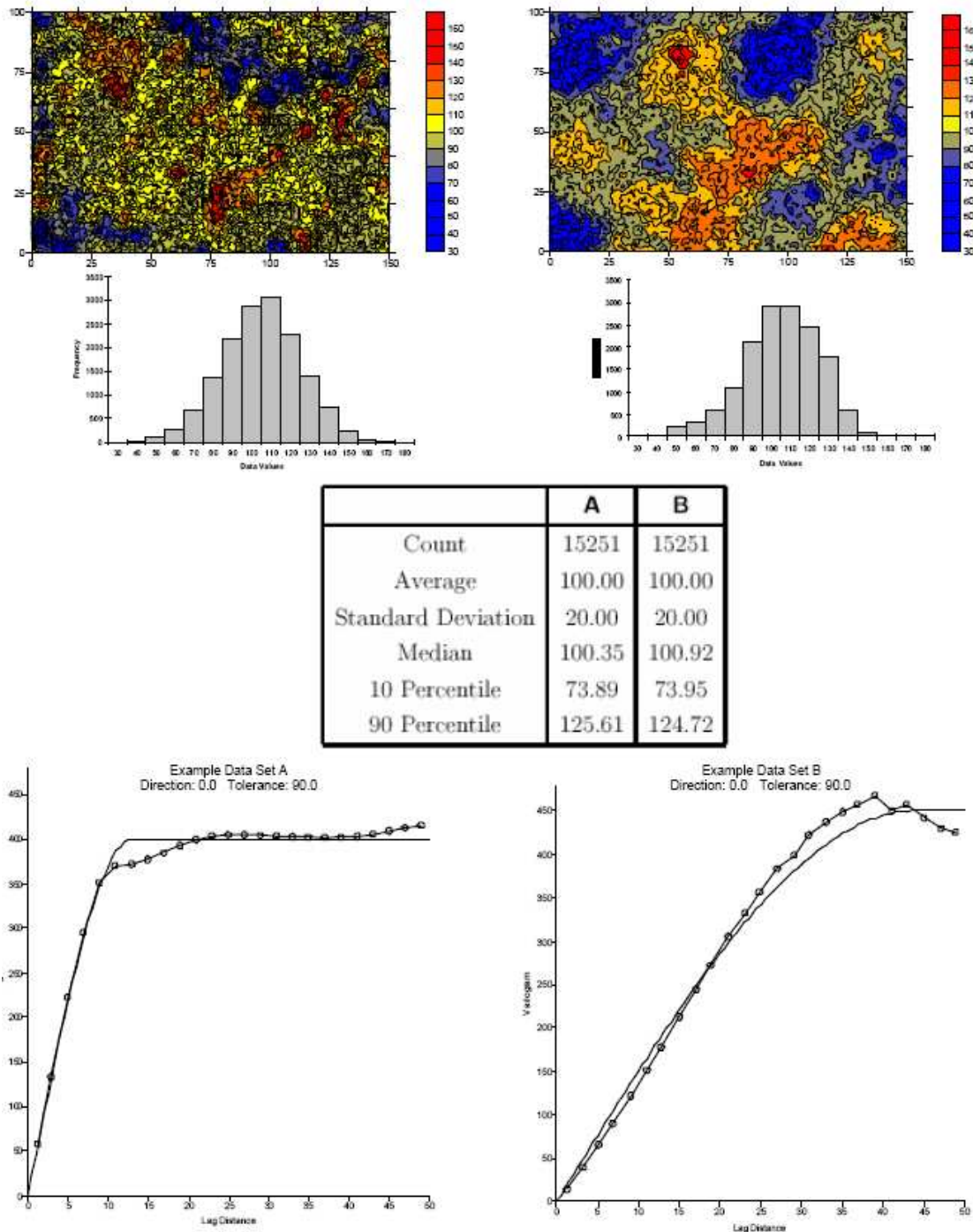


Geostatistika – vymezení pojmu

Geostatistika v užším slova smyslu – skupina interpolačních algoritmů založených na metodě krigingu. V širším slova smyslu – statistická analýza prostorově lokalizovaných dat.

Pomocí „klasických“ statistických metod lze vhodně analyzovat především atributová data – jejich kvantitativní či kvalitativní vlastnosti. Velmi omezeně však jimi lze charakterizovat prostorové vlastnosti objektů a jevů. Tyto prostorové vlastnosti jako např. spojitost jevů, prostorovou autokorelaci, prostorové uspořádání (strukturu) lze charakterizovat právě pomocí geostatistických metod.



Obr. 1.1. Prezentace prostorového rozšíření spojitého jevu metodami popisné statistiky a pomocí tzv. semivariogramu.

Na obrázku jsou znázorněny dva příklady zcela rozdílného prostorového rozšíření jistého spojitého jevu – např. koncentrace znečištění území jistou látkou. Z níže uvedené tabulky základních popisných charakteristik i histogramů nelze zjistit žádný podstatný rozdíl v prostorovém uspořádání studovaného jevu v obou porovnávaných mapách. Ten je však patrný pokud prostorové rozšíření charakterizujeme pomocí tzv. semivariogramu, který patří k základním nástrojům strukturní analýzy a geostatistických metod.

Geostatistika v širším slova smyslu představuje především:

- Konstrukce spojitých polí tzv. deterministickými metodami
- Koncept prostorové autokorelace
- Strukturní analýzu a popis prostorové autokorelace strukturními funkcemi
- Konstrukci spojitých polí metodami krigingu
- Statistický popis prostorově lokalizovaných dat (geografických objektů)
- Statistický popis prostorového uspořádání objektů (bodů, linií, ploch)
- Objektivní metody klasifikace

Metody prostorové interpolace

1. Základní pojmy

Prostorová interpolace slouží k odhadu hodnot určitého jevu či jeho intenzity v libovolném místě studované plochy, pro niž existují známé hodnoty tohoto jevu pouze v určitých lokalitách (meteorologické stanice, výškově zaměřené body apod.) Metod tedy lze využít ke konstrukci spojitých polí, k následné analýze prostorových dat – morfometrické a hydrologické modelování, optimální lokalizace apod.)

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

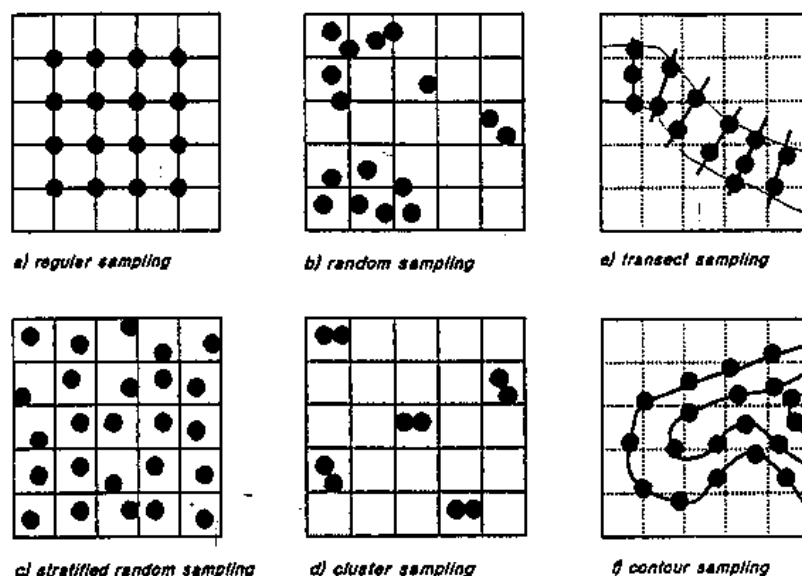
Prostorová interpolace – skupina metod, které slouží k vytváření spojitých povrchů (polí) z bodových měření. Body mohou být lokalizovány v 1, 2 i 3 rozměrném prostoru. Interpolace se může týkat nejenom bodů, ale i linií a ploch. V rámci interpolace je často řešen také problém **extrapolace** – tedy odhad 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

Lokalizace měřených (odměrných) bodů v zájmovém území. Rozmístění (tzv. sampling) je důležité pro výběr interpolačního algoritmu a úspěšnost vlastní interpolace.

Rozmístění

- Pravidelné - může být zavádějící v případě zcela rovnoměrně rozmístěného jevu, který je studován (stromy, ...)
- Náhodné - ze statistického hlediska je korektnější. Má ale i zápory - problematická lokalizace a zaměření jednotlivých míst než u pravidelně rozmístěných uzlů mřížky. Náhodné a nerovnoměrné rozmístění nemusí vystihovat základní rysy v rozložení měřené charakteristiky a může být i nákladnější.



Obr. 1 Možné způsoby rozmístění reprezentativních vzorků

Jistým kompromisem mezi pravidelným a náhodným rozmístěním může být rozmístění stratifikované náhodné (stratified random). Shlukové uspořádání umožňuje studovat jev na několika měřítkových úrovních. V řadě případů je z různých důvodů (např. ekonomických, dostupnost, ...) prováděno měření pouze v omezené míře (profily – transepty). Ve velké části interpolačních úloh je rozmístění měřených bodů předem dáno, bez možnosti ho výrazně ovlivnit vhodnou lokalizací a výběrem (např. síť meteorologických stanic apod.)

Prezentace spojitých polí - grid, TIN, izočáry, areály

Možné datové zdroje pro interpolaci

- bodová měření v terénu
- digitalizované izolinie či polygony
- stereopár leteckých fotografií či družicových obrazových záznamů

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 nejvhodnější metody

Běžné problémy interpolace:

- vymezení studované plochy – přirozené a administrativní hranice
- dostupnost bodů měření vně studované plochy

Průzkumová analýza dat (EDA – Exploratory Data Analysis)

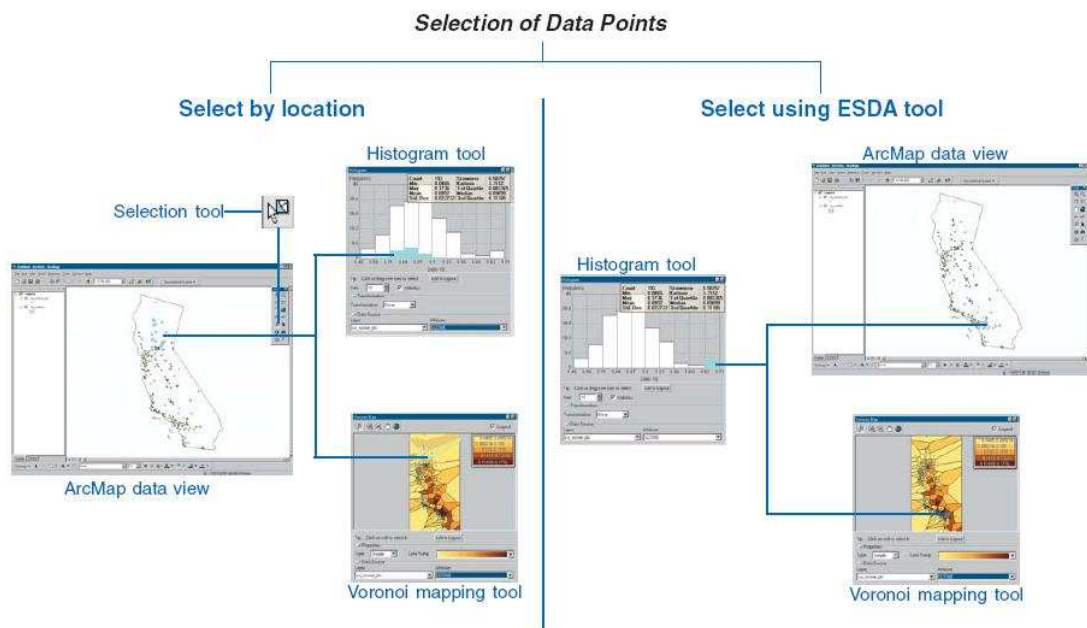
- **ESDA** – Exploratory Spatial Data Analysis
- **ESTDA** – Exploratory Spatio – Temporal Data Analysis

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. Postupy a nástroje ESDA jsou využívány i v obecné prostorové analýze dat (studium prostorové autokorelace, pattern detectors).

EDA slouží k průzkumu, deskripci, vizualizaci, zvýrazňování základních rysů dat, jejich distribuce (nejen ve smyslu prostorovém). Postupy EDA slouží k prověření požadavků normality, stacionarity vstupních dat. K těmto účelům používá specifických nástrojů (histogram, box plot, scatter plot, Q-Q graf). Deskriptivní metody používají jako **měr úrovně** ne „průměry“ ale mediánu, počítají momenty vyššího řádu (asymetrie a špičatosti). Postupy EDA 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 interpolačních postupů (např. metod krigingu).

Exploratory Spatial Data Analysis



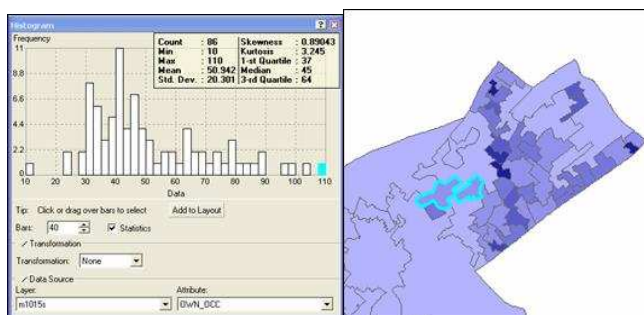
Nástroje EDA jsou často propojeny s vlastní mapou (*ESRI, Using ArcGIS Geostatistical analyst*).

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 nástroje ESDA

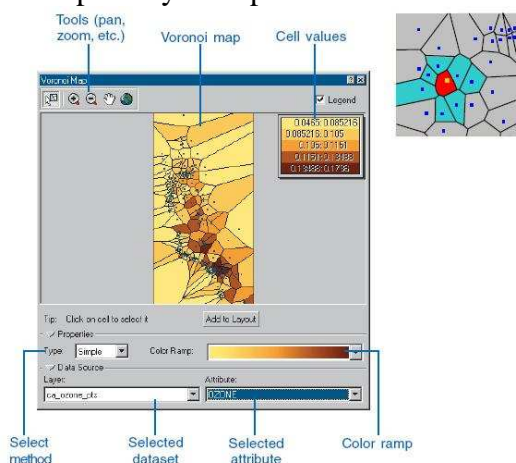
Popisná statistika a „mapped histogram“ - propojení mapy a grafu



Význam základních měř studovaného atributu je stejný jako v případě „klasické“ popisné statistiky. Propojení histogramu s mapou dovoluje hodnotit polohu a prostorové uspořádání typických resp. extrémních hodnot.

Voronoi map

Slouží k definování tzv. přirozených sousedů k vyšetřovanému bodu. Z vyšetřovaného bodu a všech přirozených sousedů lze počítat lokální statistiku – od měř úrovně (prostá hodnota atributu daného bodu) průměr, medián, směrodatná odchylka atributů polygonu daného bodu a všech sousedů), shlukování až po míry entropie.



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

$$Entropie = -\sum p_i * \log - p_i$$

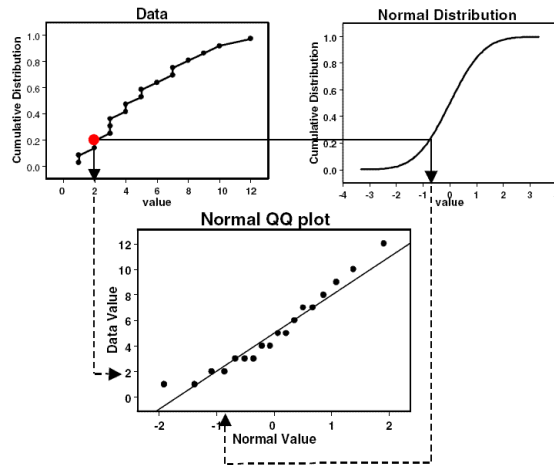
kde p_i je poměr polygonů náležejících do dané třídy z celkového počtu polygonů

Minimální entropie – všechny buňky patří do stejné třídy

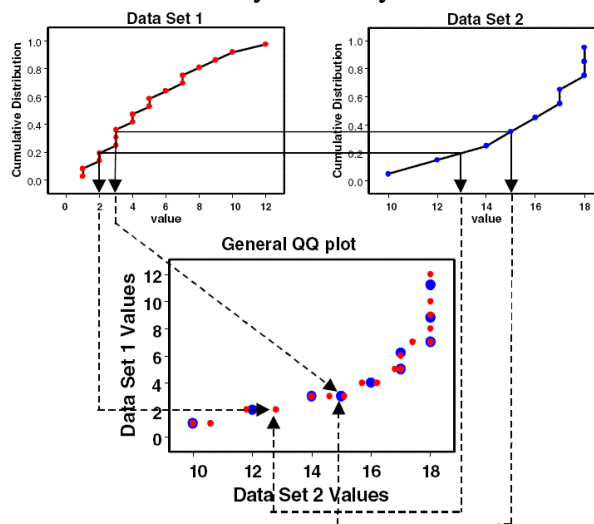
Maximální entropie – každá z buněk náleží k jiné třídě.

Kvantilové grafy - grafy zobrazující kvantily dvou rozdělení

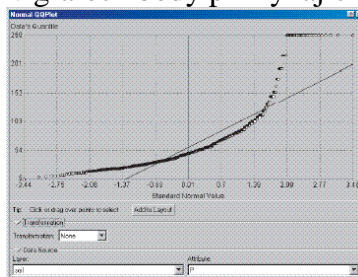
Normální Q-Q graf – vynáší se odpovídající si hodnoty kvantilů (kumulativní četnosti) vyšetřovaných dat a hodnoty kvantilů normálního rozdělení definovaného parametry vstupních dat (kumulativní distribuční funkce). Slouží jako nástroj k posouzení normality vstupních dat.



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



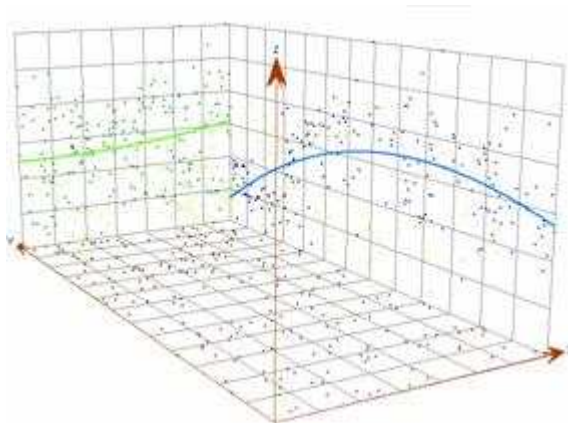
Shodu v obou případech indikují v grafech body přimykající se k přímce.



Takováto data vyžadují transformaci. Základní typy transformací:

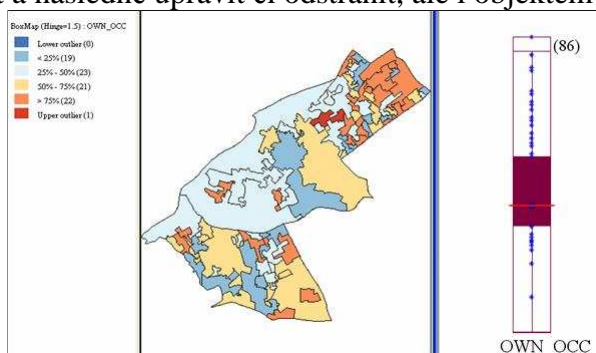
- Box-Cox
- Arcsine
- Logaritmická

Analýza trendu – za účelem 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.



Některé z metod interpolace vyžadují odstranění trendu a modelování takto upravených (**reziduálních**) hodnot. Slouží jako nástroj k posouzení stacionarity vstupních dat.

Krabicové grafy (box plots) a detekce odlehlých či extrémních hodnot – odlehlé hodnoty ve smyslu jejich polohy – odlehlé v porovnání s hodnotami okolními (local spatial outliers), nemusí být odlehlé absolutně (global outliers) Klasickými metodami bez zahrnutí prostorového aspektu je lze těžko identifikovat – např. box plot). Mohou být chybou potřebnou identifikovat a následně upravit či odstranit, ale i objektem studia.

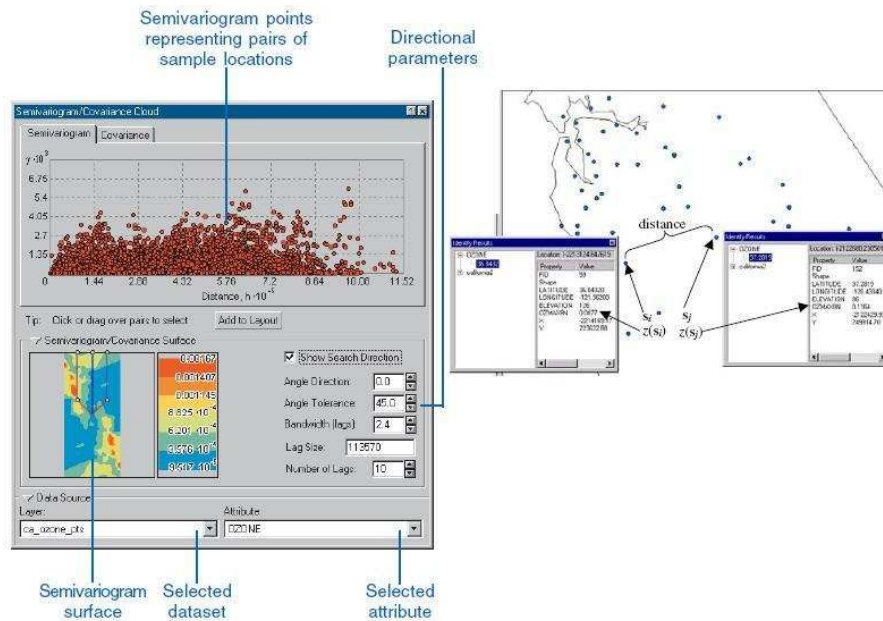


Vykreslení množiny hodnot semivariance či covariance

Slouží k detekci míry prostorové autokorelace, k vystižení míry anizotropie a k odhlazení odlehlých hodnot.

Semivariance (semivariogram) – empirický semivariogram jako graf míry nepodobnosti.

Slouží k vystižení míry prostorové autokorelace. V úlohách interpolace je tato veličina důležitá pro objektivní definování **velikosti a tvaru okolí** vyšetřovaného bodu. Počítá se jako polovina ze sumy čtverců rozdílů hodnot všech dvojic vyšetřovaných bodů vzdálených o určitou hodnotu. Semivariance na ose y a vzdálenost na ose x. Každý bod v grafu představuje dvojici bodů v analyzovaném prostoru nacházejících se v určité vzdálenosti (osa x). Podobnost hodnot interpolované veličiny je vyjádřena semivariací (osa y).



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

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

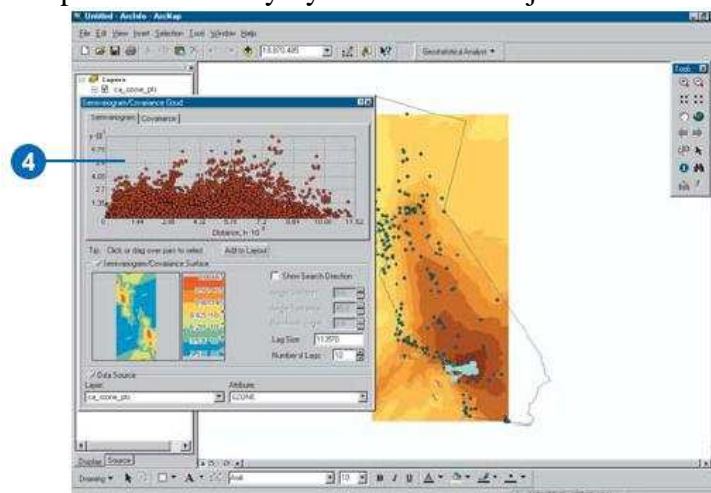
Hodnota empirické **covariance**

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

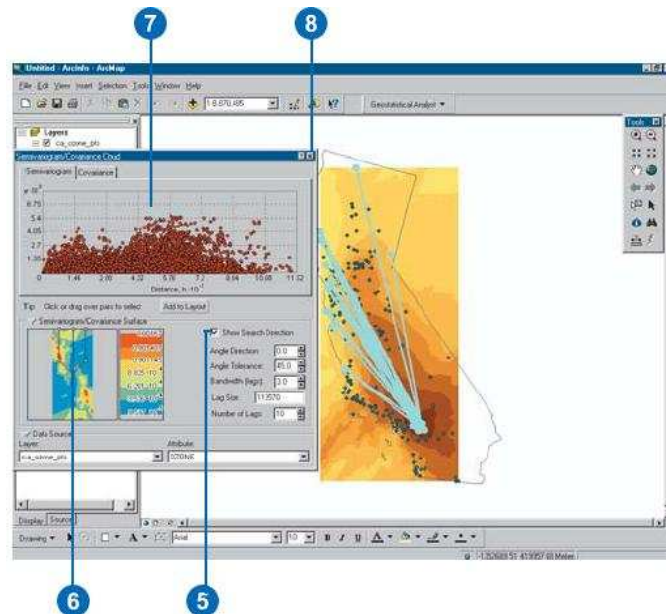
Hodnota empirické **crosscovariance**

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

Obecný model prostorové autokorelace: blízké body podobné, více vzdálené – méně podobné. Uvedený graf umožňuje identifikovat body, které se výrazně odlišují od tohoto obecného schématu. Zajímají nás především odchylky u bodů nacházejících se blízko sebe.



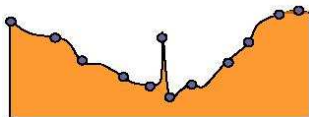
Z různých důvodů nemusí být hodnoty prostorové autokorelace obdobné s ohledem na orientaci spojnice vyšetřovaných bodů. Vysoké hodnoty semivariance mohou být vázány na body nacházející se vzájemně v určitém směru. To potom svědčí o tzv. anizotropii. Pro interpolaci to znamená asymetricky definované okolí vyšetřovaného bodu.



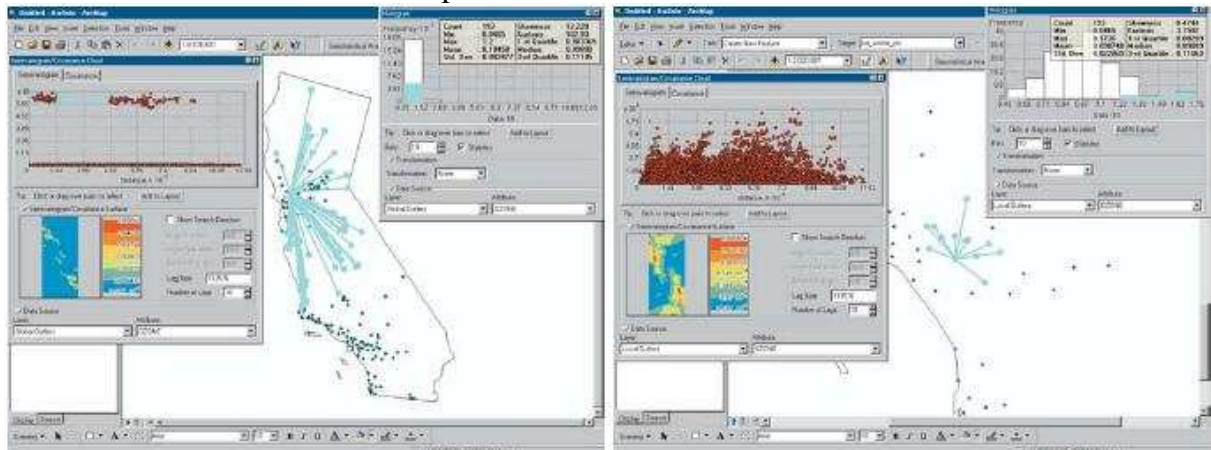
Povrch semivariogramu – povrch tvořený četnostmi (bins) hodnot semivariogramu daného směru a dané vzájemné vzdálenosti.

Directional influences - vzájemná orientace vyhledávaných bodů (search direction)

Detekce odlehlých hodnot (outliers) – globálních i lokálních



Základní nástroje: histogram
semivariogram/ covariance cloud
Voronoi map

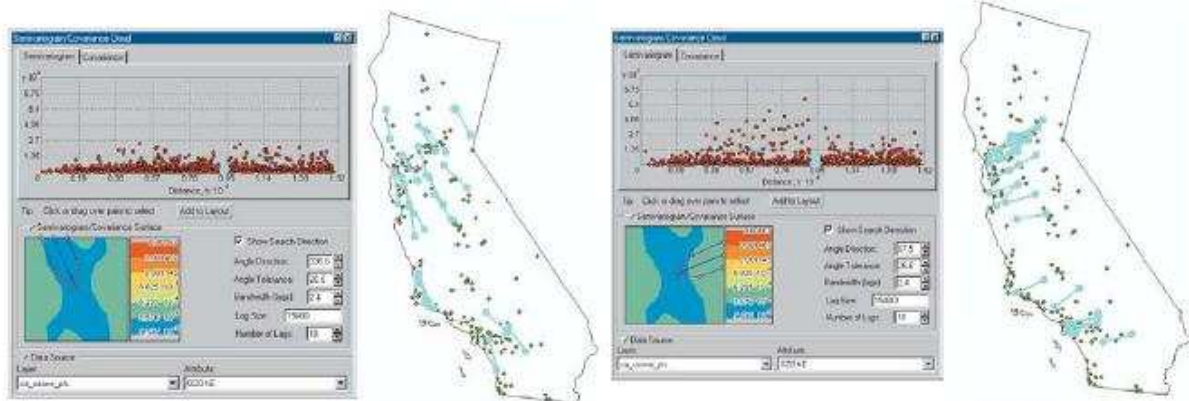


Detekce globální (vlevo) a lokální (vpravo) odlehlé hodnoty. Globální – vysoké hodnoty semivariance bez ohledu na vzdálenost a navíc se budou v mapě separovány s jedním bodem, který právě obsahuje danou odlehlou hodnotu (viz. mapa). Všechny body semivariogramu se rozdělí do dvou shluků (clouds). Globální extrém se projeví i v histogramu.

V případě lokálního extrému budou vysoké hodnoty semivariance vázány pouze na krátkou vzdálenost (v grafu jsou nahoře vlevo).

Voronoi polygons (maps) – mohou prezentovat tzv. **entropii** jako míru nepodobnosti hodnot mezi sousedními polygony. Vysoké hodnoty entropie indikují lokální extrém.

Vyšetřování tvaru okolí – izotropní a anizotropní povrch



Pro daný interval vzdálenosti (na ose x) a pro předem definovaný směr jsou vybrány odpovídající dvojice bodů prezentované hodnotami semivariance. Tyto jsou na obr vlevo všechny velmi podobné a malé. Na obrázku vpravo daleko více rozdílné. To indikuje, že semivariance jako míra podobnosti závisí na směru, kterým je měřena – tzv. izotropní povrch. Okolí bodu bude potřeba definovat jako asymetrické.

Obdobnými metodami a nástroji lze vyšetřovat prostorovou podobnost či nepodobnost dvou proměnných (**crosscovariance** – viz. výše)

2. Rozdělení metod prostorové interpolace

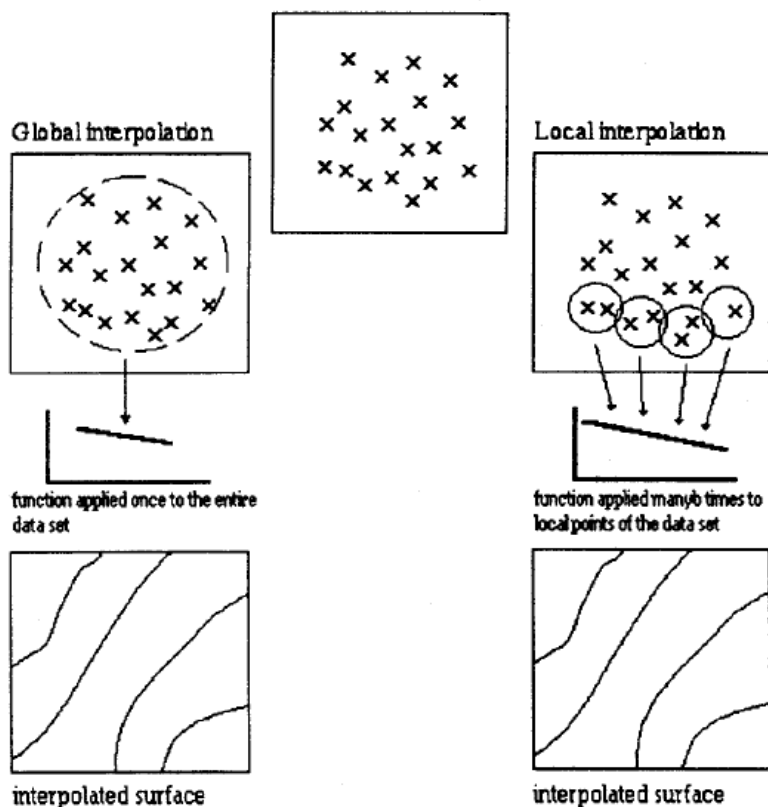
A. Podle prostorových entit, na které jsou aplikovány:

Metody interpolace bodů, linií a ploch. Dále jsou charakterizovány především metody prostorové interpolace bodů

B. Metody lokální a globální

Uvedené hledisko zohledňuje způsob, jakým daná metoda nakládá se vstupními daty (měřenými vzorky). Globální interpolace – aplikují jednu funkci na všechny měřené body ve studované ploše. Využívají princip průměrování, redukují vliv bodů s extrémními hodnotami. Produkují hladké povrchy bez náhlých zlomů.

Globální metody využívají všech měřených bodů. Bývají používány k vystižení obecných tendencí v měřených datech - trendů, jako předstupeň vlastní interpolace lokálními interpolátory, které interpolují rezidua - zbytek po odečtení trendu. Do této skupiny lze zařadit také **klasifikační metody**, které využívají všech dostupných informací k rozdělení studované oblasti do regionů, ve kterých je potom hodnota interpolovaného jevu charakterizována statistickými momenty (průměrem, rozptylem), určenými z měřených bodů v rámci každého regionu.

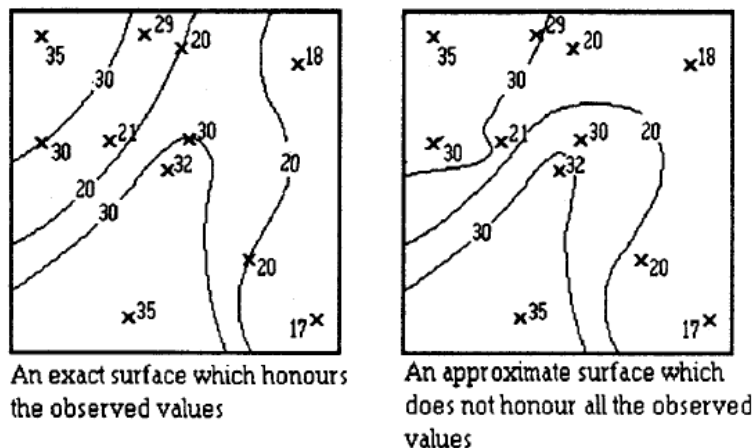


Obr. Globální a lokální metody interpolace

V závislosti na tom, co představuje nezávisle proměnnou lze globální modely interpolace dělit do dvou skupin. První skupinu tvoří modely, u nichž nezávisle proměnnou jsou pouze souřadnice měřených bodů interpolovaného atributu. Tyto metody se označují jako **analýza trendu** (trend surface analysis). Druhou skupinu globálních metod tvoří **regresní modely**. Zkoumají vztahy mezi atributy, které jsou pro dané území známé či dají se snadno změřit a atributem, jehož hodnoty jsou pro danou plochu interpolovány. Sestavený regresní model může mít podobu jednoduché i vícenásobné regrese (např. sestavení pole teplot na základě nadmořských výšek).

Lokální metody interpolace aplikují stejnou interpolační funkci opakovaně na malou část měřených dat. Tato malá část vzorků představuje okolí interpolovaného bodu. Definování okolí bodů (velikosti, tvaru) je podstatným problémem lokálních metod. Přes definici okolí mohou lokální metody přecházet v globální. Lokální techniky lze definovat také jako postupy, u kterých je nutné provést více než jeden běh algoritmu se vstupními daty. Příklady lokálních interpolací: thiessenovy polygony, klouzavé průměry, kriging. Příklady globálních interpolací: analýza trendu, fourierovy analýzy.

C. Metody exaktní a aproximující



Obr. Exaktní a aproximující metody interpolace

Metody exaktní interpolace ve výsledném povrchu zachovávají hodnoty v bodech měření. Jsou vhodné v případech, kdy existuje vysoká pravděpodobnost, že měřené hodnoty jsou správným nestranným odhadem měřené veličiny.

Aproximační metody nahrazují hodnoty v měřených bodech hodnotou vypočtenou, která se více méně liší od hodnoty měřené a je výsledkem použitého algoritmu. Jsou vhodné v případech, kdy existuje jistá míra nejistoty o naší schopnosti naměřit stejnou hodnotu v případě opakovaného měření v tomtéž bodě.

Příklady exaktních interpolací: line threading, thiessenovy polygony, kriging. Příklady aproximujících interpolací: Analýza trendu, klouzavé průměry a všechny druhy založené na filtracích.

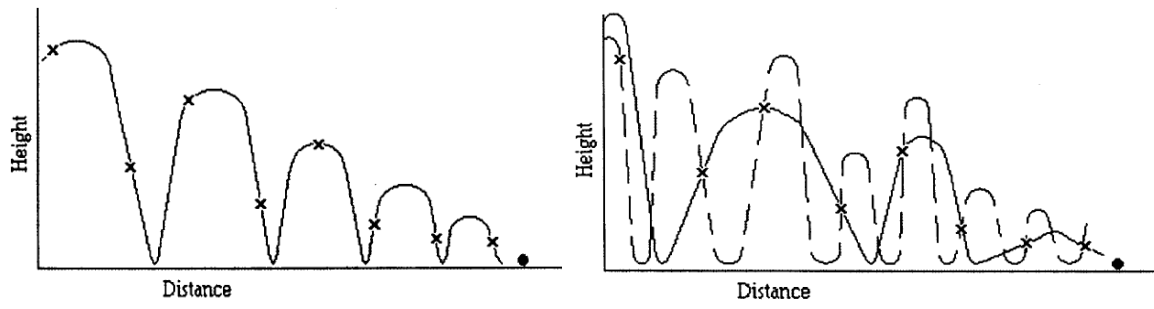
D. Metody spojité a zlomové (abrupt)

Kritériem dělení je spojitost interpolovaných hodnot. Spojité interpolátory produkují hladké povrchy na rozdíl od zlomových (thiessenovy polygony, generování obalových zón – bufferu). I spojitě metody interpolace lze omezit tzv. bariérami (srázy a zlomy při modelování terénu, atmosférické fronty při modelování některých polí met. prvků).

E. Metody deterministické a stochastické

Deterministické metody lze využít v případech, kdy existuje dostatek informací o prostorovém chování studovaného jevu, které dovolují ho popsat matematickou funkcí. Tyto metody umožňují extrapolaci za hranice vymezené měřeními vzorky. Tato extrapolace je však možná pouze za předpokladu, že máme na zřeteli fyzikální podstatu jevu (např. záporné hodnoty atmosférických srážek apod.)

Stochastické modely – zahrnují koncept náhodnosti za předpokladu, že hodnoty interpolovaného povrchu z daného měřeného vzorku jsou jen jednou z nekonečného množství možných variant. Do skupiny těchto metod patří např. metody krigingu či analýza trendu.



Obr. Deterministické a stochastické metody interpolace